

THE 1ST INTERNATIONAL WORKSHOP ON INNOVATIVE SIMULATION FOR HEALTH CARE

SEPTEMBER 19-21 2012

VIENNA, AUSTRIA



EDITED BY
WERNER BACKFRIEDER
AGOSTINO BRUZZONE
FRANCESCO LONGO
VERA NOVAK
JOSEPH ROSEN

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2012

ISBN 978-88-97999-05-8 (Paperback)
ISBN 978-88-97999-13-3 (PDF)

© 2012 DIME UNIVERSITÀ DI GENOVA

RESPONSIBILITY FOR THE ACCURACY OF ALL STATEMENTS IN EACH PAPER RESTS SOLELY WITH THE AUTHOR(S). STATEMENTS ARE NOT NECESSARILY REPRESENTATIVE OF NOR ENDORSED BY THE DIME, UNIVERSITY OF GENOA. PERMISSION IS GRANTED TO PHOTOCOPY PORTIONS OF THE PUBLICATION FOR PERSONAL USE AND FOR THE USE OF STUDENTS PROVIDING CREDIT IS GIVEN TO THE CONFERENCES AND PUBLICATION. PERMISSION DOES NOT EXTEND TO OTHER TYPES OF REPRODUCTION NOR TO COPYING FOR INCORPORATION INTO COMMERCIAL ADVERTISING NOR FOR ANY OTHER PROFIT - MAKING PURPOSE. OTHER PUBLICATIONS ARE ENCOURAGED TO INCLUDE 300 TO 500 WORD ABSTRACTS OR EXCERPTS FROM ANY PAPER CONTAINED IN THIS BOOK, PROVIDED CREDITS ARE GIVEN TO THE AUTHOR(S) AND THE WORKSHOP.

FOR PERMISSION TO PUBLISH A COMPLETE PAPER WRITE TO: DIME UNIVERSITY OF GENOA, DIRECTOR, VIA OPERA PIA 15, 16145 GENOVA, ITALY. ADDITIONAL COPIES OF THE PROCEEDINGS OF THE IWISH ARE AVAILABLE FROM DIME UNIVERSITY OF GENOA, DIRECTOR, VIA OPERA PIA 15, 16145 GENOVA, ITALY.

ISBN 978-88-97999-05-8 (Paperback)
ISBN 978-88-97999-13-3 (PDF)

THE 1ST INTERNATIONAL WORKSHOP ON INNOVATIVE SIMULATION FOR HEALTH CARE

SEPTEMBER 19-21 2012, VIENNA, AUSTRIA

ORGANIZED BY



DIME - UNIVERSITY OF GENOA



LIOPHANT SIMULATION



SIMULATION TEAM



IMCS - INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF SIMULATION



DIMEG, UNIVERSITY OF CALABRIA



MSC-LES, MODELING & SIMULATION CENTER, LABORATORY OF ENTERPRISE SOLUTIONS



MODELING AND SIMULATION CENTER OF EXCELLENCE (MSCOE)



LATVIAN SIMULATION CENTER - RIGA TECHNICAL UNIVERSITY



LOGISIM



LSIS - LABORATOIRE DES SCIENCES DE L'INFORMATION ET DES SYSTEMES



MIMOS - MOVIMENTO ITALIANO MODELLAZIONE E SIMULAZIONE



MITIM PERUGIA CENTER - UNIVERSITY OF PERUGIA



BRASILIAN SIMULATION CENTER, LAMCE-COPPE-UFRJ



MITIM - MCLEOD INSTITUTE OF TECHNOLOGY AND INTEROPERABLE MODELING AND SIMULATION - GENOA CENTER



M&SNET - MCLEOD MODELING AND SIMULATION NETWORK



LATVIAN SIMULATION SOCIETY



ECOLE SUPERIEURE D'INGENIERIE EN SCIENCES APPLIQUEES

FACULTAD DE CIENCIAS EXACTAS. INGENIERIA Y AGRIMENSURA



UNIVERSITY OF LA LAGUNA



CIFASIS: CONICET-UNR-UPCAM



INSTICC - INSTITUTE FOR SYSTEMS AND TECHNOLOGIES OF INFORMATION, CONTROL AND COMMUNICATION



NATIONAL RUSSIAN SIMULATION SOCIETY



CEA - IFAC

TECHNICALLY CO-SPONSORED



IEEE - CENTRAL AND SOUTH ITALY SECTION CHAPTER

I3M 2012 INDUSTRIAL SPONSORS



CAL-TEK SRL



LIOTECH LTD



MAST SRL

I3M 2012 MEDIA PARTNERS



INDERSCIENCE PUBLISHERS - INTERNATIONAL JOURNAL OF SIMULATION AND PROCESS MODELING



INDERSCIENCE PUBLISHERS - INTERNATIONAL JOURNAL OF CRITICAL INFRASTRUCTURES



IGI GLOBAL - INTERNATIONAL JOURNAL OF PRIVACY AND HEALTH INFORMATION MANAGEMENT



HALLDALE MEDIA GROUP: MILITARY SIMULATION AND TRAINING MAGAZINE



HALLDALE MEDIA GROUP: THE JOURNAL FOR HEALTHCARE EDUCATION, SIMULATION AND TRAINING



EUROMERCI

EDITORS

WERNER BACKFRIEDER

UPPER AUSTRIAN UNIVERSITY OF APPLIED SCIENCES, AUSTRIA

Werner.Backfrieder@fh-hagenberg.at

AGOSTINO BRUZZONE

MITIM-DIME, UNIVERSITY OF GENOA, ITALY

agostino@itim.unige.it

FRANCESCO LONGO

MSC-LES, UNIVERSITY OF CALABRIA, ITALY

f.longo@unical.it

VERA NOVAK

BETH ISRAEL DEACONESS MEDICAL CENTER, HARVARD MEDICAL SCHOOL, USA

vnovak@bidmc.harvard.edu

JOSEPH ROSEN

THAYER SCHOOL OF ENGINEERING AT DARTMOUTH, USA

Joseph.Rosen@Dartmouth.edu

**THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION
MULTICONFERENCE, I3M 2012**

GENERAL CO-CHAIRS

AGOSTINO BRUZZONE, MITIM DIME, UNIVERSITY OF GENOA, ITALY

YURI MERKURYEV, RIGA TECHNICAL UNIVERSITY, LATVIA

PROGRAM CHAIR

FRANCESCO LONGO, MSC-LES, MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA, ITALY

**THE 1ST INTERNATIONAL WORKSHOP ON INNOVATIVE SIMULATION FOR HEALTH
CARE, IWISH 2012**

GENERAL CO-CHAIRS

WERNER BACKFRIEDER, UPPER AUSTRIAN UNIVERSITY OF APPLIED SCIENCES, AUSTRIA

VERA NOVAK, BETH ISRAEL DEACONESS MEDICAL CENTER, HARVARD MEDICAL SCHOOL, USA

PROGRAM CHAIR

JOSEPH ROSEN, THAYER SCHOOL OF ENGINEERING AT DARMOUTH, USA

IWISH 2012 INTERNATIONAL PROGRAM COMMITTEE

ROBERT J. ALPINO, *EASTERN VIRGINIA MEDICAL SCHOOL, USA*
MAJA ATANASIJEVIC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
BILAL AWAN, *BAHRIA UNIVERSITY, PAKISTAN*
JERRY BATZEL, *UNIVERSITY OF GRAZ, AUSTRIA*
ALES BELIC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
FELIX BREITENECKER, *TU VIENNA, AUSTRIA*
AGOSTINO BRUZZONE, *UNIVERSITY OF GENOA, ITALY*
ERIK CAMBRIA, *UNIVERSITY OF STIRLING, UK*
TIM DAVID, *UNIVERSITY OF CANTERBURY, NEW ZELAND*
GIANLUCA DE LEO, *OLD DOMINIUM UNIVERSITY, USA*
RAFAEL DIAZ, *OLD DOMINIUM UNIVERSITY, USA*
GOTTFRIED ENDEL, *ASSOCIATION OF AUSTRIAN SOCIAL SECURITY, AUSTRIA*
DAVID FEINSTEIN, *HARVARD MEDICAL SCHOOL, USA*
GIANCARLO FORTINO, *UNIVERSITY OF CALABRIA, ITALY*
GONATA FRAGOMENI, *UNIVERSITY MAGNA GRAECIA, ITALY*
GERHARD FÜLÖP, *GESUNDHEIT ÖSTERREICH GMBH, AUSTRIA*
NANDU GOSWAMI, *MEDICAL UNIVERSITY OF GRAZ, AUSTRIA*
PIERRE GREMAUD, *NORTH CAROLINA STATE UNIVERSITY, USA*
THOMAS HELDT, *MIT, USA*
JERRY HENEGHAN, *HUMANSIM, USA*
KORINA KATSALIAKI, *INTERNATIONAL HELLENIC UNIVERSITY, GREECE*
SAMREEN LAGHARI, *VIRTUAL UNIVERSITY OF PAKISTAN, PAKISTAN*
FRANCESCO LONGO, *UNIVERSITY OF CALABRIA, ITALY*
MARINA MASSEI, *UNIVERSITY OF GENOA, ITALY*
NAVONIL MUSTAFEE, *SWANSEA UNIVERSITY, UK*
MUAZ NIAZI, *UNIVERSITY OF STIRLING, UK*
VERA NOVAK, *HARVARD MEDICAL SCHOOL, USA*
METTE OLUFSEN, *NORTH CAROLINA STATE UNIVERSITY, USA*
JOHNNY OTTESEN, *ROSKILDE UNIVERSITY, DENMARK*
GIUSEPPE PONTRELLI, *CNR, ITALY*
NIKI POPPER, *DWH SIMULATION SERVICES VIENNA, AUSTRIA*
JOSEPH ROSEN, *THAYER SCHOOL OF ENGINEERING AT DARMOUTH, USA*
RICK SEVERINGHAUS, *AEGIS, USA*
ANASTASIYA SHTILYANOVA, *UNIVERSITY OF CLERMONT-FERRAND1, FRANCE*
KULWINDER SINGH, *UNIVERSITY OF CALGARY, CANADA*
HIEN TRAN, *NORTH CAROLINA STATE UNIVERSITY, USA*
ALBERTO TREMORI, *SIMULATION TEAM, ITALY*
HAMIR USSAIN, *UNIVERSITY OF STIRLING, UK*
GEORGE VERGHESE, *MIT, USA*
GERALD ZWETTLER, *UPPER UNIV. OF APPLIED SCIENCE, AUSTRIA*

TRACKS AND WORKSHOP CHAIRS

APPLICATION OF MULTIPLE OPERATIONS RESEARCH TECHNIQUES (MORT) FOR HEALTHCARE

CHAIRS : NAVONIL MUSTAFEE, SWANSEA UNIVERSITY, UK; KORINA KATSALIAKI, INTERNATIONAL HELLENIC UNIVERSITY, GREECE

MODELING AND SIMULATION OF PHYSICAL SYSTEMS

CHAIR: GIANLUCA DE LEO, OLD DOMINIUM UNIVERSITY, USA

HEALTHCARE AND PUBLIC HEALTH M&S

CHAIR: RAFAEL DIAZ, VMASC, OLD DOMINIUM UNIVERSITY, USA

MODELING AND SIMULATION FOR COGNITIVE COMPUTATION

CHAIRS: MUAZ NIAZI, BAHRIA UNIVERSITY, PAKISTAN; AMIR HUSSAIN UNIVERSITY OF STIRLING, UK

STUDYING BIOMECHANICAL PROBLEMS FOR CARDIOTHORACIC AND CARDIOVASCULAR CLINICAL PROBLEMS: MODELS, DESIGNING TOOLS, SIMULATION ENVIRONMENTS AND CRITICAL CONDITION PREDICTION FOR SURGICAL INTERVENTIONS.

CHAIR: GIONATA FRAGOMENI, UNIVERSITY MAGNA GRAECIA, CATANZARO, ITALY

PATIENT SPECIFIC MODELING OF THE CARDIOVASCULAR- RESPIRATORY SYSTEM: INTERDISCIPLINARY APPROACHES TO THEORY AND PRACTICE

CHAIRS: JERRY BATZEL, UNIVERSITY OF GRAZ, AUSTRIA; NANDU GOSWAMI, MEDICAL UNIVERSITY OF GRAZ, AUSTRIA; METTE OLUFSEN, NORTH CAROLINA STATE UNIVERSITY, USA

MATHEMATICAL MODELING AND HEALTH TECHNOLOGY ASSESSMENT

CHAIRS: NIKI POPPER, DWH SIMULATION SERVICES VIENNA, AUSTRIA; FELIX BREITENECKER, VIENNA UNIV. OF TECHNOLOGY, AUSTRIA; GOTTFRIED ENDEL, MAIN ASSOCIATION OF AUSTRIAN SOCIAL SECURITY INSTITUTIONS, AUSTRIA

MODELLING AND SIMULATION IN PHYSIOLOGY AND MEDICINE (COMMON TRACK IWISH-EMSS)

CHAIRS: MAJA ATANASIJEVIC-KUNC, UNIV. LJUBLJANA, SLOVENIA; FELIX BREITENECKER, VIENNA UNIV. OF TECHNOLOGY, AUSTRIA

GENERAL CO-CHAIRS' MESSAGE

WELCOME TO IWISH 2012!

Since the late sixties, in the dawning modern computer age, the evolution of mostly technical disciplines of medicine was strongly influenced by computers. The primary target was the development of novel diagnostic methods, like nuclear medicine and computed tomography, enabled by new computer technology. These new modalities revolutionized medical diagnostics and built the foundations of what we call nowadays the transparent human.

The steady increase of computing power opened a new perspective to the analysis of complex systems. Medical and biological systems are influenced by a great variety of parameters, thus analysis of definite system states by causal deduction to a reduced set of parameters is hardly possible in the context of complex system interaction. Modeling and simulation arose as the novel paradigm to acquire knowledge and became a powerful research tool in bio-cybernetics. The vast issue-area of computer based analysis in biology and medicine is characterized by its interdisciplinary nature, it integrates knowledge from so different scientific fields as mathematics, chemistry, mechanics, economy or demography. Furthermore the multidisciplinary nature of the field stimulates the development of modern research strategies.

The International Workshop on Innovative Simulation for Health Care (IWISH) evolved from the EMSS conference, since the wide range of research topics and the increasing number of contributions demanded a special track within the scope of the I3M multi-conference. In its first year of occurrence the big number of presented papers has required a format of IWISH with seven sessions. They range from classical mathematical to patient specific modeling, health care management issues as well as traditional problems of bioinformatics or medical image sciences are addressed. The interdisciplinary approach to simulation and modeling in healthcare not only focuses on the achievement of new theoretical knowledge, e.g. by the improvement of expert systems, but also provides profit for the patient by improved health services, for example the reduction of waiting times, based on a sophisticated simulation of the clinical workflow.

The scope of IWISH is very broad, and we provide a new platform for scientists, engineers and practitioners of many disciplines to present their results and stimulate discussion on the various problems of healthcare. With this new interdisciplinary forum for simulation we hope to have initiated a fruitful development in health and life sciences being successfully continued during the next years.



Werner Backfrieder
Upper Austrian University of Applied Sciences, Austria



Vera Novak
Harvard Medical School,
USA



Joseph Rosen
Thayer School of Engineering
at Dartmouth, USA

ACKNOWLEDGEMENTS

The IWISH 2012 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The IWISH 2012 IPC would like to thank all the authors as well as the reviewers for their invaluable work.

A special thank goes to all the organizations, institutions and societies that have supported and technically sponsored the event.

LOCAL ORGANIZATION COMMITTEE

AGOSTINO G. BRUZZONE, *MISS-DIPTM, UNIVERSITY OF GENOA, ITALY*

ENRICO BOCCA, *SIMULATION TEAM, ITALY*

ALESSANDRO CHIURCO, *MSC-LES, UNIVERSITY OF CALABRIA, ITALY*

FRANCESCO LONGO, *MSC-LES, UNIVERSITY OF CALABRIA, ITALY*

FRANCESCA MADEO, *UNIVERSITY OF GENOA, ITALY*

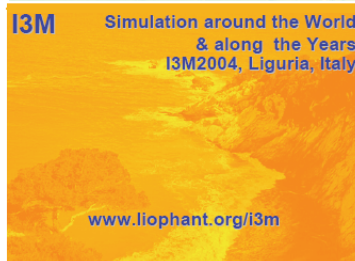
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*

LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*

ALBERTO TREMORI, *SIMULATION TEAM, ITALY*



This International Workshop is part of the I3M Multiconference: the Congress leading Simulation around the World and Along the Years



Index

Simulating the bone-titanium interfacial changes around transfemoral osseointegrated implants using physical models and modal analysis Nicola J. Cairns, Mark J. Percy, James Smeathers, Clayton J. Adam	1
Innovative online training framework supporting immersive scenario-based simulation for clinical decision making and large healthcare community of practice Bruce Joy, Liz Chung, Susan Harrison, Tim Gray, Marco Biagini, Nerida Bardon, George Kantianis	10
Three-dimensional numerical simulations of the aortic flow in presence of a left ventricle assist device Rosario Mazzitelli, Attilio Renzulli, Giuseppe Filiberto Serraino, Gionata Fragomeni	15
RANSAC-based enhancement in drug concentration predictions using support vector machine Wenqi You, Alena Simalatsar, Giovanni De Micheli	21
Simulations of uterine electrical activity using parallel computing Tanguy Hedrich, Jeremy Laforet, Catherine Marque	26
Inner simulation sustaining the deliberative process in a cognitive architecture Othalia Larue, Pierre Poirier, Roger Nkambou	32
From patients' needs to hospital pharmacy management: an holistic approach to the process modelling Raffaella Guida, Raffaele Iannone, Salvatore Miranda, Stefano Riemma, Debora Sarno	42
Improving patient's waiting time at a health screening center Thanon Wongsammacheep, Juta Pichitlamken, Waessara Weerawat	49
Reduction of turnaround time in a hospital's clinical laboratory by simulation modeling Kanyarat Luangmul, Juta Pichitlamken, Waessara Weerawat	54
A new hybrid algorithm based on watershed method, confidence connected thresholding and region merging as preprocessing for statistical classification of general medical images Gerald Zwettler, Werner Backfrieder	59
Adaptive behavior in complex healthcare interventions: assessment using computer simulation Jean-Christophe Chiem, Thérèse Van Durme, Florence Vandendorpe, Olivier Schmitz, Niko Speybroeck, Sophie Cès, Jean Macq	68
Qualitative features of a novel baroreceptor model Adam Mahdi, Johnny Ottesen, Mette Olufsen	75
Global sensitivity and identifiability analysis applied to a model predicting baroreflex regulation during head-up tilt Christian Haargaard Olsen, Jesper Mehlsen, Johnny T. Ottesen, Hien T. Tran,	81

Mette S. Olufsen

Modelling health care utilisation: a method comparison Stephanie Parragh, Patrick Einzinger	87
A comparison of system dynamics and Markov models for cost-effectiveness analysis of chronic diseases Patrick Einzinger, Ruth Leskovar, Claudia Wyrzens	93
Analysis and comparison of different modelling approaches based on an SIS epidemic Andreas Bauer, Carina Pöll, Nina Winterer, Florian Miksch, Felix Breitenecker	101
A kalman filtering based approach for the modeling of the cardiovascular regulation system Brett Matzuka, Jesper Mehlsen, Mette Olufsen, Hien Tran, Nakeya Williams	107
IFEDH - solving health system problems using modelling and simulation Niki Popper, Ingrid Wilbacher, Felix Breitenecker	113
Modular modelling and hybrid combination in health technology assessment models - examples and technology Günther Zauner, Patrick Einzinger, Florian Miksch, I. Zechmeister, Gottfried Endel, Felix Breitenecker	119
Volatile organic compounds in exhaled breath: real-time measurements, modeling, and bio-monitoring applications Julian King, Karl Unterkofler, Susanne Teschl, Anton Amann, Gerald Teschl	125
Investigation of the effect of drugs on solid tumours within a systems-based mathematical modelling framework Cong Liu, J. Krishnan, Xiao Yun Xu	131
Structural correlation method for practical estimation of patient specific parameters in heart rate regulation Johnny Ottesen, Mette Olufsen	136
Pre-tender hospital simulation using naive diagrams as models Gabriel Wurzer, Wolfgang Lorenz, Manfred Pferzinger	143
Towards leaner healthcare facility: application of simulation modelling and value stream mapping Waleed Abo Hamad, John Crowe, Amr Arisha	149
Application of discrete systems simulation to reduce waiting time in the outpatient service of a hospital in the city of São Paulo, Brazil Alexandre A. Massote, Domenico Caruso, João Batista Gonçalves Sousa	156
Uncertainty quantification for cerebral perfusion Rachael Gordon-Wright, Pierre Gremaud, Esther Martens, Vera Novak	162
Experimental and numerical studies of digital arterial elasticity by volume oscillometric analysis Pichitra Uangpairoj, Masahiro Shibata	166
Baroreflex sensitivity during the gravitational stimulus: physiology and pathophysiology	172

Raffaello Furlan, Franca Dipaola, Veronica Pacetti, Carlo Selmi, Francesca Meda, Ilaria Bianchi, Franca Barbic	
Modeling the effects of intra-abdominal hypertension	175
Jerry Batzel, Stefan Fürtinger, Daniel Schneditz	
MORBISIMMOD - morbidity based needs assessment using microsimulation	180
M. Gyimesi, I. Czasny, G. Fülöp, S. Mathis-Edenhofer	
MARIA: an agent driven simulation for a web based serious game devoted to renew education processes in health care	188
Agostino G. Bruzzone, Marco Frascio, Francesco Longo, Marina Massei, Anna Siri, Alberto Tremori	
Agent Based Simulation Model For Obesity Epidemic Analysis	195
Agostino G. Bruzzone, Vera Novak, Francesca Madeo	
Authors' Index	204

SIMULATING THE BONE-TITANIUM INTERFACIAL CHANGES AROUND TRANSFEMORAL OSSEOINTEGRATED IMPLANTS USING PHYSICAL MODELS AND MODAL ANALYSIS

Nicola J. Cairns^(a), Mark J. Pearcy^(b), James Smeathers^(c), Clayton J. Adam^(d)

^(a) Department of Design, Manufacture and Engineering Management, University of Strathclyde, Glasgow, UK

^(b,d) Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

^(c) Institute of Health and Biomedical Innovation and School of Exercise and Nutrition Sciences, Queensland University of Technology, Brisbane, Australia

^(a) nicola.j.cairns@strath.ac.uk, ^(b) m.pearcy@qut.edu.au, ^(c) j.smeathers@qut.edu.au, ^(d) c.adam@qut.edu.au

ABSTRACT

Non-invasive vibration analysis is being considered as a method to monitor the healing progression of femoral implants in transfemoral amputees. Studies to date have successfully detected gross alterations in the physical properties of the interface region of physical bone-implant models using vibration techniques. This paper describes the development of a series of physical models which simulate the incremental bone-implant interfacial changes during progressive osseointegration. The capability of modal analysis to detect the changing interface conditions is investigated. The model resonant frequencies and their mode shapes altered due to the different interface conditions. Higher resonances were shown to be more sensitive to interface changes than the fundamental frequency. The findings demonstrate the potential of modal analysis for this application and the technique warrants further investigation.

Keywords: osseointegration, modal analysis, transfemoral amputation, physical model

1. INTRODUCTION

An alternative to using a prosthetic socket for above knee (transfemoral) amputees is transfemoral osseointegration (TFOI) (Hagberg and Brånemark 2009). A titanium implant is inserted into the medullary canal of the amputated femur; the implant protrudes through the skin and connects directly to the prosthetic limb removing the need for a socket (Ward and Robinson 2005). Studies have reported several advantages of TFOI compared to using a socket (Hagberg 2005; Hagberg, Brånemark et al. 2008; Sullivan, Uden et al. 2003) and it can be particularly appropriate for amputees that experience skin problems due to socket wear, for those with a short residual limb and those with an active lifestyle (Hagberg, Brånemark et al. 2008).

However, it can take twelve to eighteen months for the implant to integrate with the bone and for an amputee to be able to load bear and therefore be fully rehabilitated (Ward and Robinson 2005). The long

rehabilitation time is a significant disadvantage of TFOI and may be impeding the wider adoption of the technique.

Vibration analysis techniques are being investigated as non-invasive methods of assessing the degree of bone-implant integration (known as osseointegration (OI)) (Cairns, Adam et al. 2011; Shao, Xu et al. 2007; Swider, Guérin et al. 2009; Xu, Shao et al. 2005). Physical models have been used to simulate different interface conditions between the femur and implant that may occur during osseointegration. The different physical properties at the femur-implant interface can be identified by measuring the changes in the dynamic properties of the system (Cairns, Adam et al. 2011; Shao, Xu et al. 2007; Xu, Shao et al. 2005). The longer-term aim of the vibration analysis research is to determine when the implant is able to withstand physiological load by monitoring the dynamic properties of osseointegration progression and potentially reduce the overall rehabilitation time.

Despite the reported success of the vibration techniques, the physical models developed in previous studies (Cairns, Adam et al. 2011; Shao, Xu et al. 2007; Xu, Shao et al. 2005) compare interface conditions which represent gross changes at the femur-implant interface *in vivo*. The interface conditions developed in the author's previous work (Cairns, Adam et al. 2011) were intended to "represent extremes of the spectrum of implant integration with the bone" and were used to establish the feasibility of the modal analysis methodology employed. Further work is required to develop more appropriate interface condition models and ascertain if modal analysis is capable of detecting femur-implant interfacial changes that are more representative of the *in vivo* scenario.

This research details a series of composite femur-implant physical models of TFOI developed to represent the known histological and mechanical properties at the femur-implant interface during osseointegration. Modal analysis is then conducted using the models to establish if the technique remains

capable of detecting changes at the interface between the implant and the femur.

2. METHODOLOGY

2.1. Physical Model Development

It is important to develop physical models that represent the real structure as realistically as possible. To this end, literature on the physiology of the bone-titanium interface was reviewed to identify key stages of transfemoral osseointegration. Physical models were then developed to simulate the mechanical characteristics of the femur-implant interface associated with the key stages.

The femur model and implant model were common to all physical models, only the interface condition was altered to represent the different stages of OI. Fourth generation large composite femurs (Sawbones model 3406, Pacific Research Laboratories Inc, WA, USA) were used as the femur model. The femurs were cut to a length of 237mm (distally from the femoral head), replicating the amputated femur (Figure 1(a)). Implants were machined from commercially pure titanium rod with a threaded section 80mm long, 19mm outer diameter and 1.75mm male thread pitch. The profile of the implants changed to a cylindrical section 60mm long, 15mm outer diameter (the dimensions used clinically). Flats were machined on the cylindrical section and a 2.5mm threaded hole was machined in one flat (Figure 1(b)) to allow attachment of the excitation hardware (detailed in section 2.2).

A physiological model found in the literature outlines four stages of bone-titanium osseointegration (Brånemark, Gröndahl et al. 2005). The Brånemark physiological model, other literature supporting the model (including histological and mechanical characteristics of the interface) and the physical model interfaces developed to represent the four stages are detailed in sections 2.1.1 to 2.1.4.

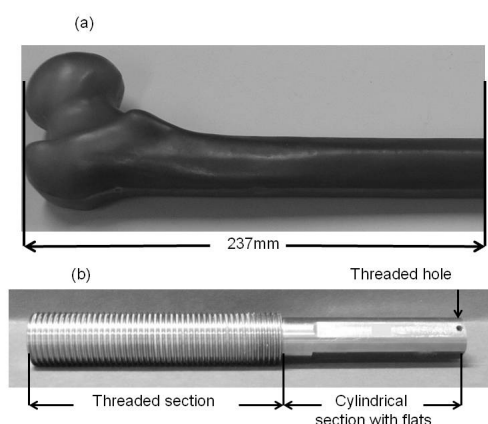


Figure 1: (a) the cut composite femur and (b) the titanium implant

2.1.1. Stage 1 Physical Model

In stage 1 of the Brånemark physiological model, immediately after the implant is inserted, there are areas

of bone-implant contact and also gaps filled with haematoma. The bone surrounding the implant is damaged while the bone further away from the implant is healthy. A combination of bone-implant contact and haematoma filled gaps has also been reported in numerous long bone animal studies immediately after implant insertion (Dhert, Thomsen et al. 1998; Franchi, Fini et al. 2005; Linder, Albrektsson et al. 1983; Sennerby, Thomsen et al. 1993; Uthoff 1973).

To replicate the combination of bone-implant contact and gaps, the medullary canal of the femur was pre-threaded using a CNC machine to achieve an implant insertion torque of 5Nm. The insertion torque value 'at first implant insertion' was chosen to be lower than 12Nm; the torque the implant is tightened to six months after insertion (Ward and Robinson 2005).

Mechanical testing of the bone-implant interface immediately after implant insertion has demonstrated negligible tensile strength (Kitsugi, Nakamura et al. 1996; Steinemann, Eulenberger et al. 1985) and shear strength (Brånemark, Ohnrell et al. 1997; Brånemark, Ohnrell et al. 1998; Ivanoff, Sennerby et al. 1996; Johansson 1987; Rubo de Rezende and Johansson 1993; Sennerby, Thomsen et al. 1993). Therefore no additional interface materials were used to bond the implant to the femur.

To assemble the Stage 1 model the implant was inserted in the threaded canal of the femur using a dial torque wrench which continuously measured the 5Nm insertion torque. The implant was inserted to a depth of 90mm measured from the cut end of the femur. The interface simulation is summarised in Table 1 and the model is illustrated in Figure 2.

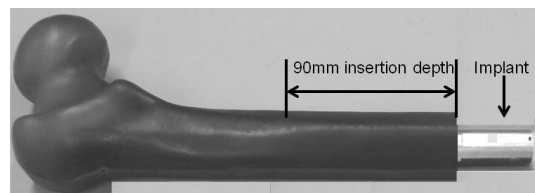


Figure 2: Assembled Stage 1 physical model showing implant inserted in pre-threaded femur

2.1.2. Stage 2 Physical Model

At stage 2 of the physiological model complete contact between the implant and the bone is achieved: the haematoma forms new bone to fill the gaps and the surrounding damaged bone heals. This process of new bone formation and resorption/replacement of damaged bone is supported by the findings of numerous long bone animal studies (Buma, van Loon et al. 1997; Dhert, Thomsen et al. 1998; Franchi, Fini et al. 2005; Linder, Albrektsson et al. 1983; Sennerby, Thomsen et al. 1993; Uthoff 1973; Ysander, Brånemark et al. 2001). This process is estimated to take over four months in the human (Roberts, Turley et al. 1987). It is thought that modal analysis would need to be capable of detecting interfacial changes throughout this period in order to be a useful technique. Therefore two physical models were manufactured to represent different time

points in the formation of mature bone; Stage 2-intermediate (Stage 2-int) and Stage 2-end.

The tensile strength of the interface remains minimal after bone remodelling (Kitsugi, Nakamura et al. 1996; Steinemann, Eulenberger et al. 1985) while the implant removal torque increases (Brånemark, Ohnrell et al. 1997; Brånemark, Ohnrell et al. 1998; Ivanoff, Sennerby et al. 1996; Johansson 1987; Rubo de Rezende and Johansson 1993; Sennerby, Thomsen et al. 1993).

To represent an intermediate point with immature bone in complete contact with the implant, an additional material was inserted at the interface of the Stage 2-intermediate model. The interface material was a liquid to solid resin (Palapress, Heraeus Kulzer GmbH, Germany) with a lower elastic modulus than the femur (16GPa for composite femur; 2.1GPa for resin). The size of the necrotic bone region undergoing remodelling ranges from 0.5mm to 40% of the bone radius (Albrektsson 1985; Buma, van Loon et al. 1997; Roberts, Turley et al. 1987). Therefore the femur was bored out to a diameter 3mm larger than the implant diameter.

To assemble the Stage 2-intermediate model the resin was poured in to the femur canal and then the implant was inserted to a depth of 90mm so that the resin filled the gap around the implant. When cured the resin did not adhere to the implant therefore complete femur-implant contact was achieved with negligible interface tensile strength.

To represent an end point with mature bone in complete contact with the implant, the medullary canal of the femur was pre-threaded using a CNC machine to achieve an implant insertion torque of 20Nm. This value of insertion torque was chosen to be larger than 12Nm required to attach components to the implant (Ward and Robinson 2005). No additional materials were used therefore the interface had negligible tensile strength.

To assemble the Stage 2-end model the implant was inserted in the threaded canal of the femur to a depth of 90mm using a dial torque wrench which continuously measured the 20Nm insertion torque. The interface simulations are summarised in Table 1.

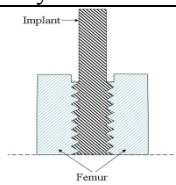
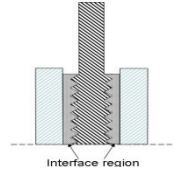
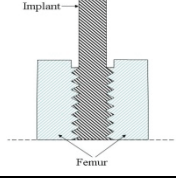
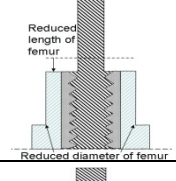
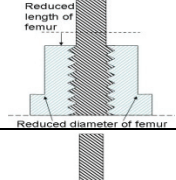
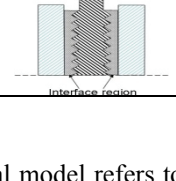
2.1.3. Stage 3 Physical Model

At stage 3 of the physiological model, the healthy revascularised bone can now withstand load and remodels due to the loading stimulus applied. Animal studies that have applied loading protocols to implants (after an initial unloaded healing period) provide further evidence that bone remodelling occurs due to load stimulus (Brunski, Hipp et al. 1989; Duyck, Ronold et al. 2001; Hoshaw, Brunski et al. 1994). Furthermore there is some clinical evidence of bone surface remodelling around the TFOI implant *in vivo* when load is applied to the implant (Xu, Shao et al. 2005). To the author's knowledge there is no information on the tensile and shear strength of the bone-titanium interface after implant loading.

As the process of bone remodelling due to load stimulus would be similar to remodelling of damaged bone in Stage 2, no additional interface models were made to replicate bone remodelling. However the surface modelling resulting in bone thickness variations was thought to be an important clinical observation. Therefore the Stage 2 physical models were modified to simulate external surface resorption and named Stage 3-intermediate (Stage 3-int) and Stage 3-end

To represent the distal bone resorption (Xu, Shao et al. 2005) the femur length was reduced by 10mm on both models. To represent the cortical wall tapering along half the implant length the femur diameter was reduced to 26mm over 40mm length. The interface simulations are summarised in Table 1.

Table 1: Summary of the Mechanical Characteristic of each Physical Model and a Schematic of the Model

Stage	Mechanical Characteristics of Physical Model	Schematic of Physical Model
1	Femur pre-threaded to produce implant insertion torque of 5Nm. No adhesion between implant and femur. Low shear strength at interface	
2-int	Femur bored out to radius 1.5mm larger than implant. Gap filled with resin. No adhesion between resin and implant. Increase in implant removal torque	
2-end	Femur pre-threaded to produce implant insertion torque of 20Nm. No adhesion between implant and femur. Increase in implant removal torque	
3-int	Stage-2-intermediate modified. Femur length reduced by 10mm. Femur diameter reduced over 40mm implant length	
3-end	Stage-2-end modified. Femur length reduced by 10mm. Femur diameter reduced over 40mm implant length	
4	Femur bored out to radius 1.5mm larger than implant. Gap filled with silicone. No adhesion between silicone and implant	

2.1.4. Stage 4 Physical Model

Stage 4 of the Brånemark physiological model refers to the formation of non-mineralised connective tissue between the bone and implant instead of healthy bone; this can occur in unsuccessful cases. Fibrous tissue

encapsulation around titanium implants placed in the femur of two animal studies verify the possibility of Stage 4 in the physiological model (Thomas and Cook 1985; Thomas, Kay et al. 1987). To the author's knowledge there is no information on the tensile and shear strength of the fibrous tissue-titanium interface that is known as unsuccessful osseointegration.

To simulate connective tissue encapsulation of the implant an additional material was inserted at the interface of the Stage 4 model. Fibrous tissue surrounding an implant has been mechanically tested (Hori and Lewis 1982) and a silicone elastomer found to have similar properties (Waide, Cristofolini et al. 2004). Therefore the same grade of liquid-to-solid silicone was used as the interface material (Sylgard 184, Dow Corning Corporation, U.S; 2.6MPa elastic modulus). The thickness of the fibrous tissue region around an implant is greater than 1mm (Waide, Cristofolini et al. 2003). Therefore the femur was bored out to a diameter 3mm larger than the implant diameter (1.5mm thickness of silicone).

To assemble the Stage 4 model the silicone was poured in to the femur canal and then the implant was inserted to a depth of 90mm so that the silicone filled the gap around the implant. When cured the silicone did not adhere to the implant therefore the model had negligible interface tensile strength. The interface simulation is summarised in Table 1.

2.1.5. Femur Boundary Condition model

The boundary condition of the amputated femur *in vivo* is provided by the acetabulum and the connection of the muscles and soft tissue at the femoral head. This was simulated in the physical models by encapsulating the femoral head in a rectangular block of liquid-to-solid resin (Palapress, Heraeus Kulzer GmbH, Germany; 2.1GPa elastic modulus) and clamping the block to create a cantilever. A similar resin block has been used to simulate the boundary condition in the modal analysis of the fractured tibia (Nikiforidis, Bezerianos et al. 1990). Therefore the cantilevered boundary condition (Figure 3) was considered an acceptable first attempt at representing the *in vivo* boundary condition at the femoral head.

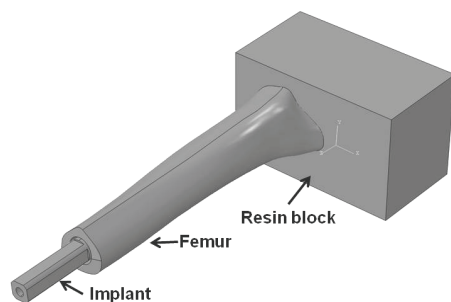


Figure 3: Resin block boundary condition encapsulating the femoral head

To create the resin block (dimensions 120x75x65mm), the femur was fixed in a custom-made mould which provided a minimum resin thickness of 10mm around the extremities of the femoral head. The resin was poured into the mould and allowed to cure. The physical model was then removed from the mould ready for modal analysis.

2.2. Modal Analysis

Forced excitation was applied to the models using an electromagnetic shaker driven by a power amplifier (part numbers 4810 and 2706 Bruel&Kjaer, Naerum, Denmark). The shaker methodology has been previously evaluated for this application using less complex models (Cairns, Adam et al. 2011). A signal generator (33120A, Agilent Technologies, CA, USA) input an excitation signal to the shaker.

The excitation was measured using a dynamic force transducer (0.028kg) and signal conditioner (part numbers 2311-500 and 4416B, Endeveco, CA, USA). A Delrin stinger connected the shaker to the force transducer (Bruel&Kjaer, Naerum, Denmark). The other side of the force transducer was connected to the implant of the physical model using a screw connection in the threaded hole (Figure 1(b)).

The model response was measured using a single axis piezoelectric accelerometer (0.002kg) and a charge conditioning amplifier (part numbers 4393 and 2692-A-0S2, Bruel&Kjaer, Naerum, Denmark). The accelerometer was attached to the model using beeswax to allow the location to be easily changed.

The excitation and response signals were recorded using a 16-bit resolution data logger (USB-6259, National Instruments, NSW, Australia) connected to a personal computer (HP Intel® Core™ 2Duo CPU 3.5GB RAM) using data acquisition software (LabVIEW SignalExpress version 2.5, National Instruments) and a sampling rate of 50kHz.

Figure 4 shows the coordinate system and excitation/response measurement sites identified along the model length. Seventeen response sites were identified for each femur-implant model. Site 17 was chosen as the excitation site.

The resin block was clamped to a steel base (dimensions 500x510x25mm) fixed to the laboratory floor. Sections of 12mm threaded rod were fitted through holes in the steel base and the resin block was fixed between the base and rectangular plates using nuts tightened to 16Nm. The experimental set up is shown in Figure 5. To maintain the correct alignment of the shaker and the model, the shaker was suspended on a spring over the physical model (Figure 5).

With the shaker attached to site 17 (dashed arrow in Figure 4; Figure 5) via the stinger and force transducer, the model was excited in the y-axis direction using a sine sweep signal (100Hz-5kHz frequency range, 500mV peak-to-peak amplitude and 5kHz per second sweep rate). The sine sweep parameters were optimised to obtain multiple resonant frequencies with adequate signal to noise ratio. The sweep was repeated

ten times and averaged in the data processing. The response was measured with the accelerometer attached to response site 1. The test was then repeated using the same excitation site but with the accelerometer attached to site 2-17 in turn. The y-axis testing was conducted on all six femur-implant models which are summarised in Table 1.

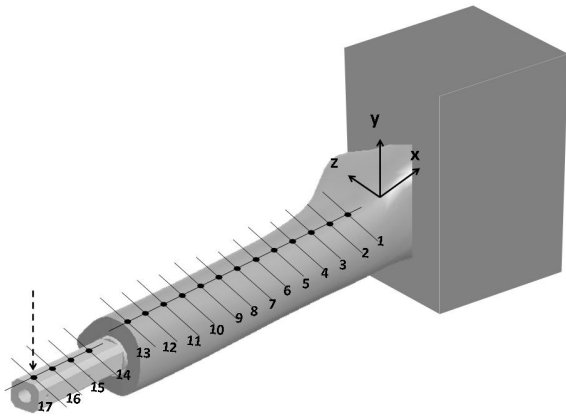


Figure 4: The coordinate system established for the models and the 17 response sites identified along the length of the model. Site 17 was chosen as the excitation site. The dashed arrow represents the y-axis excitation applied at site 17 used in all the modal testing.

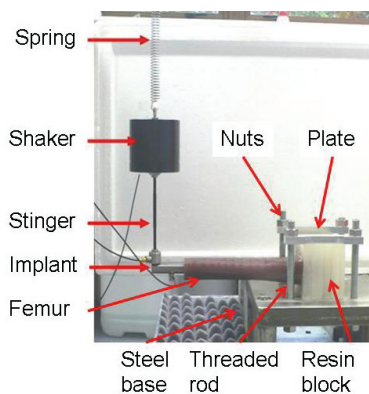


Figure 5: Experimental Set up of Physical Model Cantilevered to Steel Base. The shaker is suspended over the model using a spring and is connected to the implant at site 17 via a stinger and force transducer.

Customized analysis programs were written using MATLAB software (version 2007a, MathWorks Inc, Natick, MA, USA) to process the input and response signals and compute the frequency response function, accelerance, defined as the ratio of acceleration response to excitation force in the frequency domain. The MATLAB programs are detailed in Cairns (2010). Using a Fast Fourier Transform algorithm the accelerance function was computed from the tests performed at each excitation/response site combination (17 accelerance functions in total). These were used to calculate a mean accelerance function which was plotted against frequency to identify the resonant

frequencies; resonances manifest as peaks on this type of plot. The peak picking method is illustrated in the author's previous work (Cairns, Adam et al. 2011).

Plots of the imaginary component of accelerance versus frequency at each excitation/response site combination were used to depict the mode shapes. The amplitude of the imaginary accelerance at a resonant frequency represents the displacement magnitude occurring at that site, while the sign of the amplitude indicates the positive or negative direction of the displacement (Avitabile 1999). Therefore by identifying the amplitude and the sign of the imaginary accelerance at each site (at a resonant frequency), the mode shape of the model can be determined. The resonant frequency values and mode shapes were compared between the physical models to ascertain if the different interface conditions could be detected.

3. RESULTS

3.1. Resonant Frequencies

Four resonant frequencies were identified for each physical model (Table 2). The signal to noise ratio was poor at frequencies over 3.5kHz and no resonances were identifiable above this frequency. It is likely that frequency changes relative to a baseline measurement recorded over time would be relevant *in vivo*. Therefore the percentage change in each resonant frequency from the baseline measurement (Stage 1 model) was calculated (Table 2). There is a maximum change of 5% and 7% in the fundamental and second frequency respectively due to the changing femur-implant interface properties. The third and fourth resonances are more sensitive to the alterations in interface condition; a maximum of 15% and 13% change in frequency respectively.

Table 2: Resonant Frequencies peak picked from Accelerance-Frequency plots for each physical model. Percentage Change in Resonant Frequency from the baseline Stage 1 model is shown in parenthesis [% change = ((Stage 1- Stage N)/Stage 1)*100, where N is physical model in table row]

Model (Stage)	Resonant Frequency (Hz)			
	1 st mode	2 nd mode	3 rd mode	4 th mode
1	221	662	988	2258
2-int	220 (0)	662(0)	1133(-15)	2435(-8)
2-end	223(-1)	669(-1)	1071(-8)	2394(-6)
3-int	230(-4)	671(-1)	1123(-14)	2450(-9)
3-end	231(-5)	669(-1)	1078(-9)	2541(-13)
4	214(3)	614(7)	986(0)	2240(-1)

3.2. Mode Shapes

The mode shape of the fundamental frequency of each physical model is illustrated in Figure 6. Typically mode shapes are normalised for comparison. The mode shapes have not been normalised so that the relative magnitude of the mode for each model can be compared (given that the excitation force was constant throughout

the modal testing). Data at site 13 is missing from the Stage 3 model because the femur length was reduced to simulate bone resorption. The fundamental mode shapes are similar to that of a classic cantilevered beam – approximately zero displacement at cantilevered end (site 1) increasing to maximum displacement at the free end (site 17). The modes have similar magnitude (imaginary accelerance value) as well as shape for all the physical models with the largest discrepancy evident in the Stage 2-int model.

The mode shape of the third frequency of each physical model is illustrated in Figure 7. It is evident that there are differences in the shape and magnitude of this mode across the models. The modes of the Stage 2 models are similar in shape and magnitude to the modes of the Stage 3 models. The Stage 1 model has a similar deformation pattern to Stage 2 and 3 at sites 1-8 and 14-17, but behaves differently at sites 9-13. The Stage 4 model is isolated on the plot with low magnitude displacement along the model length.

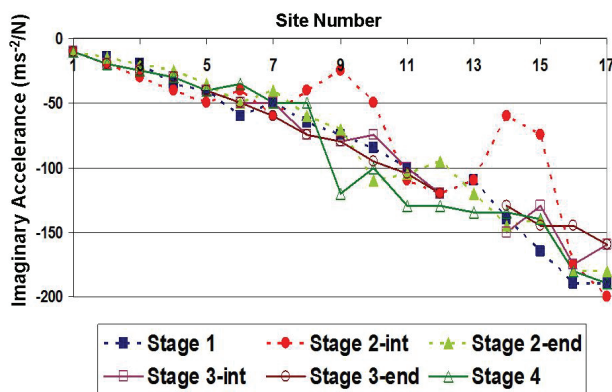


Figure 6: Fundamental Frequency Mode Shape of the Physical Models

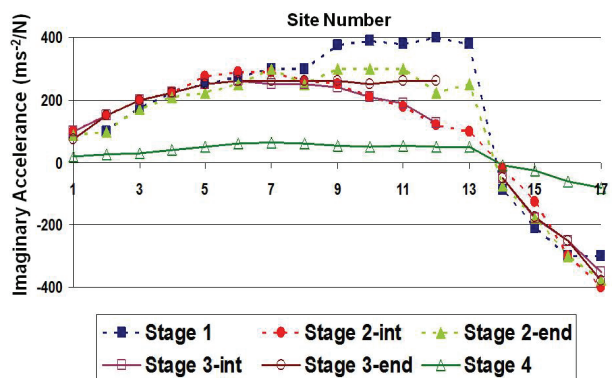


Figure 7: Third Frequency Mode Shape of the Physical Models

4. DISCUSSION

The first four resonant frequencies of the physical models changed due to the change in the physical properties at the femur-implant interface. In particular the third and fourth frequencies were more sensitive to the changing interfacial condition than the fundamental and second frequencies (greater percentage change in frequency; Table 2). This suggests that the higher

frequencies may be more useful in the detection of osseointegration progression than the fundamental frequency. Therefore the modal analysis methodology reported here, which uses a broad frequency range excitation and detects multiple resonances, may offer greater functionality than impact excitation like that used in Shao, Xu et al. (2007) and Xu, Shao et al. (2005) where only the fundamental frequency of the femur-implant model was reported. This finding reinforces the author's previous work where sine sweep excitation delivered using an electromagnetic shaker was determined to be superior to impact excitation for this application (Cairns, Adam et al. 2011).

The small change in fundamental frequency (0-5%; Table 2) appears to contradict the author's previous work (Cairns, Adam et al. 2011) where the change in frequency between two femur-implant models with different interface conditions was 47% (estimated from the accelerance plot). The earlier investigation used different interface conditions and femur boundary conditions than those developed in the current study. Therefore, the results are not directly comparable. Nevertheless this finding suggests that the femur boundary condition has an effect on the magnitude of the frequency changes detected and this requires further investigation.

Arguably frequency changes compared to a baseline value would be useful for longitudinal measurements of osseointegration progression *in vivo*. In consideration of this, the changes in the third frequency are particularly interesting. The third frequency of the Stage 1, Stage 2-int and Stage 2-end models are quantifiably different. However the frequency change between the Stage 2 models and their Stage 3 counterparts (femur mass alterations simulating bone resorption but no interface change) is small (1%). This indicates that the third frequency may be capable of detecting different characteristics at the femur-implant interface but is not sensitive to bone mass changes. This finding may be important as the modal analysis technique *in vivo* would need to detect bone remodeling at the femur-implant interface due to load stimulus and not erroneously detect femur surface modeling which occurs at the same time. Furthermore there is no change in frequency when the Stage 4 model is compared to the baseline. This suggests that unsuccessful fibrous tissue formation could be distinguished from the progression of OI using the zero frequency change.

The fundamental frequency mode shape is similar for all six physical models (Figure 6). There is some digression from this shape in the Stage 2-int model (site 9,14,15). This is possibly due to inadequate fixation of the response accelerometer using the beeswax at these sites which led to lower than expected response magnitudes. Nevertheless the maximum and minimum of the fundamental mode shape are consistent for all six models. The similarity in the mode shape across models despite their different interface conditions supports the

suggestion that this may not be the optimum mode to investigate the detection of OI progression.

By contrast, the third frequency mode shape reveals differences between the models. In particular the Stage 2 and Stage 3 models have similar mode shapes along most of the model length while the Stage 1 and Stage 4 models exhibit differences in both shape and magnitude. The results suggest that monitoring the mode shapes could be a complementary or alternative method to monitoring the change in the resonant frequency values and could be used to identify changes at the femur-implant interface. Previously, mode shape analysis was not conducted by Xu and Shao (Shao, Xu et al. 2007; Xu, Shao et al. 2005) because their methodology did not allow for it. Therefore the modal analysis technique used in the current study has enabled further investigation of the feasibility of vibration analysis applied to the TFOI system.

The physical models with their different femur-implant interface conditions were developed to simulate key stages of OI progression based on histological observations of titanium implants in long bones and the mechanical properties of the bone-titanium interface. However the physical models cannot replicate the complex mechano-biological *in vivo* system in full. This is a limitation of using physical models to represent a biological system. Nevertheless the current study provides an iterative improvement in the physical modeling of the TFOI system. Furthermore using physical models enables all other variables of the system to be controlled. This allows the vibration methodology to be evaluated with respect to detecting femur-implant interfacial changes only. This is not possible using *in vitro* or *in vivo* models.

The cantilevered resin block was used as the femoral head boundary condition. The authors acknowledge that the *in vivo* femoral boundary condition is more complex than this. However the resin block was considered an appropriate first estimation of the boundary condition *in vivo*. Furthermore a similar boundary condition has been used successfully in the modal analysis of the tibia (Nikiforidis, Bezerianos et al. 1990).

5. CONCLUSION

A series of physical models were developed to simulate the mechanical characteristics of key stages in OI progression. Changes in the resonant frequencies and mode shapes as a result of physical property changes at the femur-implant interface were demonstrated, showing that the modal analysis technique is capable of detecting the incremental interfacial changes.

The findings indicate that higher resonances and their mode shapes may be more appropriate for the detection of OI progression than the fundamental resonance. The model boundary conditions may affect the success of the modal analysis technique and further investigation of boundary condition influence is required.

ACKNOWLEDGMENTS

This work was funded by an Australian Government International Postgraduate Research Scholarship.

REFERENCES

- Albrektsson, T., 1985. Bone Tissue Response. In: Albrektsson, T., eds. *Tissue-integrated prostheses: Osseointegration in clinical dentistry*. Quintessence Publishing, 129-143.
- Avitabile, P., 1999. Modal space: back to basics. *Experimental Techniques* 23:17-18.
- Brånemark, P.I., Gröndahl, K., Brånemark, B.K., 2005. Why Osseointegration Would Work and How It Did in the First Patients Treated. Basic Facts and Philosophical Thoughts. In: Brånemark, P.-I., S. Chien, H.-G. Gröndahl, K. Robinson, eds. *The Osseointegration Book: From Calvarium to Calcaneus*. Quintessenz, 35-114.
- Brånemark, R., Ohnells, L.O., Nilsson, P., Thomsen, P., 1997. Biomechanical characterization of osseointegration during healing: an experimental *in vivo* study in the rat. *Biomaterials* 18:969-978.
- Brånemark, R., Ohnells, L.O., Skalak, R., Carlsson, L., Brånemark, P.-I., 1998. Biomechanical characterization of osseointegration: an experimental *in vivo* investigation in the beagle dog. *Journal of Orthopaedic Research* 16:61-69.
- Brunski, J.B., Hipp, J.A., Cochran, G.V.B., 1989. The influence of biomechanical factors at the tissue-biomaterial interface. *Material Research Society Symposium* 110:505-515.
- Buma, P., van Loon, P.J.M., Versleyen, H., Weinans, H., Slooff, T.J.J.H., de Groot, K., Huiskes, R., 1997. Histological and biomechanical analysis of bone and interface reactions around hydroxyapatite-coated intramedullary implants of different stiffness: a pilot study on the goat. *Biomaterials* 18:1251-1260.
- Cairns, N.J., 2010. *The feasibility of vibration analysis as a technique to detect osseointegration of transfemoral implants*. PhD. Queensland University of Technology.
- Cairns, N.J., Adam, C.J., Percy, M.J., Smeathers, J., 2011. Evaluation of modal analysis techniques using physical models to detect osseointegration of implants in transfemoral amputees. *Engineering in Medicine and Biology Society (EMBC), 2011 Annual International Conference of the IEEE*, 1600-1603, Boston, USA.
- Dhert, W.J.A., Thomsen, P., Blomgren, A.K., Esposito, M., Ericson, L.E., Verbout, A.J., 1998. Integration of press-fit implants in cortical bone: A study on interface kinetics. *Journal of Biomedical Materials Research* 41:574-583.
- Duyck, J., Ronold, H.J., Van Oosterwyck, H., Naert, I., Vander Sloten, J., Ellingsen, J.E., 2001. The influence of static and dynamic loading on marginal bone reactions around osseointegrated

- implants: an animal experimental study. *Clinical Oral Implants Research* 12:207-218.
- Franchi, M., Fini, M., Martini, D., Orsini, E., Leonardi, L., Ruggeri, A., Giavaresi, G., Ottani, V., 2005. Biological fixation of endosseous implants. *Micron* 36:665-671.
- Hagberg, K., 2005. Physiotherapy for patients having a trans-femoral amputation. In: Chien, S., H.G. Grondahl, K. Robinson, eds. *The Osseointegration Book From Calvarium to Calcaneus*. Berlin:Quintessenz Verlags-GmbH, 477-487.
- Hagberg, K., Brånemark, R., 2009. One hundred patients treated with osseointegrated transfemoral amputation prostheses - rehabilitation perspective. *Journal of Rehabilitation Research and Development* 46:331-344.
- Hagberg, K., Brånemark, R., Gunterberg, B., Rydevik, B., 2008. Osseointegrated trans-femoral amputation prostheses: Prospective results of general and condition-specific quality of life in 18 patients at 2-year follow up. *Prosthetics and Orthotics International* 32:29-41.
- Hori, R.Y., Lewis, J.L., 1982. Mechanical properties of the fibrous tissue found at the bone-cement interface following total joint replacement. *Journal of Biomedical Materials Research* 16:911-927.
- Hoshaw, S.J., Brunski, J.B., Cochran, G.V.B., 1994. Mechanical loading of branemak implants affects interfacial bone modeling and remodeling. *International Journal of Oral and Maxillofacial Implants* 9:345-360.
- Ivanoff, C.-J., Sennerby, L., Lekholm, U., 1996. Influence of mono- and bicortical anchorage on the integration of titanium implants: A study in the rabbit tibia. *International Journal of Oral and Maxillofacial Surgery* 25:229-235.
- Johansson, C., 1987. Integration of screw implants in the rabbit: a 1-year follow-up of removal torque of titanium implants *The International Journal of Oral and Maxillofacial Implants* 2:69-75
- Kitsugi, T., Nakamura, O., Yan, W.-Q., Goto, T., Shibuya, T., Kokubo, T., Miyaji, S., 1996. Bone bonding behaviour of titanium and its alloys when coated with titanium oxide(TiO₂) and titanium silicate(Ti₅Si₃). *Journal of Biomedical Materials Research* 32:149-156.
- Linder, L., Albrektsson, T., Branemark, P.-I., Hansson, H.-A., Ivarsson, B., Jonsson, U., Lundstrom, I., 1983. Electron microscopic analysis of the bone-titanium interface. *Acta Orthopaedica Scandinavica* 54:45-52.
- Nikiforidis, G., Bezerianos, A., Dimarogonas, A., Sutherland, C., 1990. Monitoring of fracture healing by lateral and axial vibration analysis. *Journal of Biomechanics* 23:323-330.
- Roberts, W.E., Turley, P.K., Brezniak, N., Feilder, P.J., 1987. Bone physiology and metabolism. *Journal of the California Dental Association* 54-61.
- Rubo de Rezende, M.L., Johansson, C.B., 1993. Quantitative bone tissue response to commercially pure titanium implants. *Journal of Materials Science in Medicine* 4:233-239.
- Sennerby, L., Thomsen, P., Ericson, L.E., 1993. Early tissue response to titanium implants inserted in rabbit cortical bone . Part I Light microscopic observations. *Journal of Materials Science: Materials in Medicine* 4:240-250.
- Shao, F., Xu, W., Crocombe, A., Ewins, D., 2007. Natural frequency analysis of osseointegration for trans-femoral implant. *Annals of Biomedical Engineering* 35:817-824.
- Steinemann, S.G., Eulenberger, J., Maeusli, P.-A., Schroeder, A., 1985. Adhesion of bone to titanium *Biological and Biomechanical Performance of Biomaterials : Proceedings of the Fifth European Conference on Biomaterials*, September 4-6, Paris, France.
- Sullivan, J., Uden, M., Robinson, K.P., Sooriakumaran, S., 2003. Rehabilitation of the trans-femoral amputee with an osseointegrated prosthesis: The United Kingdom experience. *Prosthetics and Orthotics International* 27:114-120.
- Swider, P., Guérin, G., Baas, J., Søballe, K., Bechtold, J.E., 2009. Characterization of bone-implant fixation using modal analysis: Application to a press-fit implant model. *Journal of Biomechanics* 12:1643-1649.
- Thomas, K.A., Cook, S.D., 1985. An evaluation of variables influencing implant fixation by direct bone apposition. *Journal of Biomedical Materials Research* 19:875-901.
- Thomas, K.A., Kay, J.F., Cook, S.D., Jarcho, M., 1987. The effect of surface macrotecture and hydroxylapatite coating on the mechanical strengths and histologic profiles of titanium implant materials. *Journal of Biomedical Materials Research* 21:1395-1414.
- Uthoff, H.K., 1973. Mechanical factors influencing the holding power of screws in compact bone. *Journal of Bone and Joint Surgery* 55B:633-639.
- Waide, V., Cristofolini, L., Stolk, J., Verdonshot, N., Toni, A., 2003. Experimental investigation of bone remodelling using composite femurs. *Clinical Biomechanics* 18:523-536.
- Waide, V., Cristofolini, L., Toni, A., 2004. An experimental analogue to model the fibrous tissue layer in cemented hip replacements. *Journal of Biomedical Materials Research part B* 69B:232-240.
- Ward, D.A., Robinson, K.P., 2005. Osseointegration for the skeletal fixation of limb prostheses in amputations at the trans-femoral level. In: Chien, S., H.G. Grondahl, K. Robinson, eds. *The Osseointegration Book From Calvarium to Calcaneus*. Berlin:Quintessenz Verlags-GmbH, 463-476.
- Xu, W., Shao, F., Ewins, D., 2005. A resonant frequency measurement system for

osseointegration trans-femoral implant. *Key Engineering Materials* 295:139-144.

Ysander, M., Brånemark, R., Olmarker, K., Myers, R.R., 2001. Intramedullary osseointegration: Development of a rodent model and study of histology and neuropeptide changes around titanium implants. *Journal of Rehabilitation Research and Development* 38:183-190.

AUTHORS BIOGRAPHY

Dr Nicola Cairns is a postdoctoral research fellow at the University of Strathclyde, U.K. She received a B.Eng in Mechanical Engineering from the University of Edinburgh and an MSc in Bioengineering from the University of Strathclyde before working as a Clinical Bioengineer with the West of Scotland Mobility and Rehabilitation Centre, Glasgow, U.K. Her 2010 PhD research at Queensland University of Technology, Australia investigated using vibration analysis to monitor bone healing and growth around femoral orthopaedic implants and the implementation of the technique as a clinical tool. She combines her clinical and technical engineering knowledge to pursue her current research interests in the design of lower limb prosthetic devices and the rehabilitation of lower limb amputees using non-invasive technologies. She has published 10 journal and conference papers on prosthetic device design, assessment and use in amputee rehabilitation.

Professor Percy has conducted research in Biomedical Engineering for more than 30 years. His experience ranges from the measurement of three-dimensional movements and mechanics of the human spine, in Glasgow, Oxford and Durham, UK, to the mechanics and biology of bone fracture healing, in Adelaide and Brisbane. He has published over 200 refereed papers and conference proceedings in international publications.

Dr James Smeathers is a senior lecturer in Clinical Biomechanics at Queensland University of Technology. He graduated from Reading University with a B. App. Sci. (hons) degree in Applied Zoology with Agriculture and PhD thesis topic on a mechanical analysis of the mammalian lumbar spine. His current research interests include studies on acute changes in post exercise tissue mechanics of the Achilles tendon, tendinopathy, the spinal column and non-invasive diagnostic techniques for endoprosthetic loosening. Previously, he worked at the Bioengineering Unit within the Rheumatism and Rehabilitation Research Unit, University of Leeds researching the viscoelastic properties of the intervertebral disc, musculoskeletal stiffness of the spinal column and associations between low back pain and whole body vibration in professional drivers.

A/Prof Adam has led the research activities of the QUT/Mater Hospitals Paediatric Spine Research Group

in Brisbane, Australia since 2002. His work focuses on the development of validated computational biomechanics models of the human musculoskeletal system, with the aim of better understanding the onset of disease and improving treatment outcomes using patient-specific surgery simulation. He has expertise in non-linear constitutive modelling and finite element analysis of biological tissues, as well as in vitro biomechanical testing of tissues and joints. He is currently undertaking a Marie Curie fellowship with Laboratoire de Biomécanique, Arts et Métiers ParisTech on multiscale modelling and characterisation of the intervertebral disc.

INNOVATIVE ONLINE TRAINING FRAMEWORK SUPPORTING IMMERSIVE SCENARIO-BASED SIMULATION FOR CLINICAL DECISION MAKING AND LARGE HEALTHCARE COMMUNITY OF PRACTICE

Bruce Joy^(a) Liz Chung^(b) Susan Harrison^(c) Tim Gray^(d) Marco Biagini^(e)
Nerida Bardon^(f) George Kantianis^(g)

^(a) VastPark Services Inc., USA

^(b) VastPark Pty Ltd, Australia

^{(c)(d)(f)(g)} Australian Centre for Health Innovation, Australia

^(e) Mathematics Engineering and Simulation PhD program, Genoa University, Italy

^(a) bruce.joy@vastpark.com

^(c) s.harrison@healthinnovation.com.au

^(e) m.biagini@liophant.org

ABSTRACT

There is reasonable doubt that current arrangements for clinical training will sustain projected increased requirements for health education. It is evident that Simulation Education provides a solution to this sustainability problem, and the increasing sophistication of Virtual world technologies create opportunities to migrate elements from traditional to online models, thus addressing the issues of cost of access to traditional Simulation Centres. However, creating exciting online environments in isolation is insufficient in addressing the complex requirements of undergraduate health care training.

The authors propose that by balancing the focus between cognitive reasoning structure and the need to establish an emotional connection, they have delivered a high quality, learner-centric online world that increases the “real world readiness” for undergraduates. Furthermore, by delivering the tool through a supportive online community of practice, the learner has a sustainable, continuous opportunity to practice essential skills in a safe environment before entering the clinical environment.

Keywords: healthcare simulation education, virtual simulation, community of practice

1. INTRODUCTION

The traditional approach to healthcare education relies on patients as the primary vehicle for clinician training. Aside from the ethical issues of patient safety and well-being, practice on live patients is opportunistic and does not allow repeated practice to achieve competence.

Furthermore the rapid growth in the number of trainees combined with a focus on increased efficiency of the healthcare system mean that the current system will not be able to sustain the projected clinical placement requirements for health education.

Simulation Based Education (SBE), using mannequins and part-task trainers, provides a solution partial solution to this sustainability problem, but access to SBE is limited by geographical isolation and cost.

Online ‘push’ based tools may augment traditional and SBE learning platforms, however the growing sophistication of virtual world technologies create further opportunities to enhance the learning provided by online models.

There is no resource base that allows professionals at different levels of learning, from different geographic locations and different craft groups to collaborate openly. Importantly, there is no forum for these groups of professionals to shape their own learning and teaching resources and link their day-to-day world with their world of learning, specifically the multi-disciplinary nature of their world.

These issues were highlighted by the Productivity Commissions in their research report [2005] which noted the poor coordination between education and training and the health delivery aspects of the system and the low responsiveness to changing needs.

To address this situation, the Australian Federal Government has established the Health Workforce Australia (HWA) to help understand the challenges and to implement a robust system of providing a skilled, flexible and innovative health workforce that meets the needs of the Australian community. The use of simulated learning environments is one of the strategies acknowledged by HWA as assisting to meet the training requirements.

This focus from Federal government has enabled a project to focus on communities of practice in the Simulation Education community. Additional support from the Victorian state government enabled the creation of a Virtual environment for nurse training. The main project collaborators have been able to blend aspects of both these projects to create an Innovative online training framework supporting immersive scenario-based simulation for clinical decision making as well as large healthcare communities of practice.

2. SUSTAINABLE CLINICAL PLACEMENTS

The Productivity Commission report of 2005 projected the need for an additional 1.3million clinical placement days by the 2013 academic year. The present system of

health workforce training and deployment cannot meet the growing need for health services either locally or globally. Chronic disease and an aging population are placing escalating demands on resources; additionally there is a compounding challenge of the health workforce itself aging.

In order to cope with this additional demand for training, more innovative education methods are needed to supplement the traditional classroom and hospital learning environments. With significant increases in interest in virtual learning environments there is huge potential for online immersive simulation education.

3. OPPORTUNITIES CREATED BY ADVANCES IN TECHNOLOGY

A number of developments over the last decade have allowed physical simulation education to be moved to a completely virtual world. The exponential growth of computer processing power over time, famously predicted by Moore in 1965, has been consistently occurring for over 40 years. This dramatic increase in processing power underpins the specific developments required to provide an effective virtual medical simulation. Graphics rendering has improved immensely over the last decade, allowing for interaction with very high fidelity 3D visual representations of real geographical locations. Although 'ground up' software development costs have been, and can still be, prohibitively expensive, the use of already existent graphical engines has made this technology more economical and transferrable to an almost limitless array of educational requirements.

The concurrent growth of data transmission infrastructure allows much of this software and processing power to be centralised. The rollout of virtual simulation education packages is therefore simplified both in cost and complexity; learners simply participate online from any modern personal computer, without the need to install complex software on their computer.

The computer gaming industry has been the largest innovator and producer of these technologies and much can be learnt from what has been done before. Massively multiplayer online games (MMOG), in which many players, often numbering in the thousands, can interact in a permanent online virtual world, have existed for over a decade. The purpose of these games can vary. Some are traditional role playing games with player success being determined by objective goals to be accomplished, whereas others have no predefined goals, but merely serve as a medium where participants can interact with each other via avatars. In this social respect, they overlap extensively with traditional communities of people, whether physical or online, and in the process blur the boundary between game and reality. Perhaps the most well recognised example of this is the online game 'Second Life' by Linden Labs. 'Second Life' has already been used as a platform for medical education.

4. THE REQUIREMENTS

Simulation based training is a growing part of the training process for health professionals. By providing an online immersive training simulator for clinical decision making (a first for Australia), health professionals will be able to access training 24/7. The online immersive environment offers a very significant cost benefit over physical simulation training. It is not expected to replace the critical value of true hands-on training under expert supervision that the physical simulation facilities provide, its value is in augmenting physical facilities and resources and enabling a 24/7 solution to busy health professionals needing training and practice on-demand wherever they are.

Long-established modalities of learning where students are taught by specialist individuals within educational institutions have gradually moved to include the use of simulation education, but in recent times the focus has moved to communities of practice and of learning. This means a more experiential and collaborative basis that incorporates social media and communication rather than an internalized individual learning method.

Lev Vygotsky the early 20th century Russian psychologist suggested the concept of a zone of proximal development (ZPD) where a learner can progress the solution of a problem in collaboration with adult guidance or more capable peers. In the same vein, Lave and Wegner assert that mastery resides not in the teacher but in the organization of the community of practice.

Research in online learning has shown that students favour an environment that promotes discussion. Engaging online creates an opportunity to engage with peers in relation to the problem at hand which then allows for better critical thinking and thus more accurate decision making as reflected in real life work experience. This in turn has positive flow-on effects that can minimize problems with serendipitous clinical placement opportunities. This level of engagement builds the ability of students to understand clinical roles and responsibilities which has been shown to be a key area for miscommunication and misunderstanding.

In creating such a portal or online environment it is also vital to build appropriate support for the students' learning experience. It is not enough to merely focus on curriculum at the cost of a realistic experience or vice versa but a combination of both that are symbiotic within the greater structure of a community of practice. This concept is observed often in virtual simulation research that concerns itself more with the technological possibilities of the tool itself rather than the community it is based in. Often simulation based training seems dominated by technology therefore losing affinity with its original purpose in health and education.

For a student to feel comfortable with what they are learning and how they are progressing within set learning outcomes, their emotional needs must be met. Access to appropriate debriefing after virtual

simulations is vital as is the continued support of a mentor or facilitator who can both guide and encourage the student and intervene if an issue arises. Nevertheless a hidden problem can exist where there is a subjective vision of success and that a clinical need is being met purely because of the perception that a simulation centre or an online virtual experience by its very nature must meet these needs. This is where the strength of an online community lies in that not only does it provide a support network, engage students and promote participation and conversation, it can shape the user experience into skills and confidence that therefore apply in reality. Effectively, simulations should map real-life clinical experiences gained within communities of actual practice. These environments need to be motivational and learner centered which is essentially conducive to learning. Similarly, simply providing access to online content and simulations is not enough to guarantee effective learning. Regular reinforcement is of prime importance so that recently acquired skills are not lost.

5. PROJECT SCOPE

Although the rapid growth in information technology outlined earlier has opened up many possibilities for health education, it also poses significant new challenges. All too often, shiny innovations in technology per se have taken the focus away from the educational goals themselves. A high degree of psychological fidelity is required to adequately train cognitive and decision-making skills, the level of physical fidelity required to achieve this will vary with each situation. Too much physical detail in the wrong sections of a virtual environment (hyperfidelity) can actually impair rather than improve psychological fidelity. Conversely, too little or unrealistic animation will decrease the level of psychological immersion attained by students. Packages already in existence such as 'Second Life' have been used for health education as described before, particularly for their powerful community and collaboration aspects. By designing our own package (Nursim), we were able to exert more control over the details of the simulation, and in doing so achieve an appropriate but not distracting level of physical fidelity.

6. THE SOLUTION



Nursim combines the technological opportunity and the system-wide drive for solutions that are both economically valid and educationally relevant. We considered three main themes as most relevant to develop the best solution:

- Cognitive Structure
- Emotive immersion
- Supportive framework

The most constant component of simulation education is the development of appropriate clinical scenarios to best meet the learning objectives of the students. Whether presented at the bedside, in physical simulation centres, or in completely virtual environments, the learning material must be relevant and challenging but not overwhelming. Essential elements include case based learning, making decisions with regard to real-time clinical information presented to participants, and receiving real-time feedback with regard to consequences of their previous decisions. This narrative style of consequences immediately following actions is critical to learning, as it contextualises the role the learner plays in the scenario, and promotes reflection. Fortunately, extensive use of case based learning scenarios throughout health education already exists, and these have been applied to our virtual simulation scenarios.

As mentioned earlier, high physical fidelity can certainly improve immersion within virtual scenarios, but only if it is applied to appropriate sections of the simulation. Detailed and realistic graphical representation of real world objects and procedures should be saved for those elements of the simulation that demand more attention, focus and manipulation by the learners. In fact, attention to detail at these critical moments can drive psychological fidelity, enriching the learning experience for the learner. Ideally, analysis of the real world performance from a human factors perspective will crystalize the most important physical components of the work environment that need to be replicated with high physical fidelity in the virtual world. Unfortunately, there is a paucity of data in this area. Care must also be taken to ensure that learners, educators, and software developers appreciate that ultimately, an educational program's success is judged by achieving learning outcomes, and not by how photorealistic the virtual world was.



7. NURSIM – TYPICAL SCENARIOS

The use case presented focuses on delivering an immersive clinical decision training solution where trainees can log on and they will enter into a virtual hospital environment where they are greeted by a virtual

instructor, meet a patient and go through an assessment of the patient including observation and clinical decision making process where the patient gives common clues and the trainee must effectively question the patient and determine an effective course of action. Then the virtual instructor will give feedback. The instructor can also provide clues if needed by the trainee.

Standard clinical observation processes that will be supported include checking pulse, blood pressure, breathing, considering a patient's gestures that may reveal pain points and asking questions (based on a branching conversation model). The branching conversation model will be initially created by experienced simulation education staff and users will then be able to input new scenarios so that healthcare trainers, staff, students can all potentially develop and experiment with new scenarios and make them available to others to utilize.



This environment may also be used as a “free play” virtual role playing environment where multiple trainees and instructors can enter into the scene together and, using voice (VoIP) can talk together to run through a training exercise without requiring travel.

The architectural framework will provide:

- Private simulation training for healthcare students on-demand 24/7 from their homes. There are a number of aspects of this which are likely to be world-firsts
- Lower cost of providing a course of simulation-based training by having several sessions occurring online in a low cost online simulator and at least one session occurring in the physical simulators at CHI
- Prototype secure online immersive learning so that if successful, this can be more widely used to increase training capacity at a low cost of delivery.

The level of success of the environment as a clinical decision training solution hinges on the level of choice that the trainee has whilst in the simulation. The aim of the project is to create a responsive and as true-to-life

rendition of a set of circumstances as possible based on training for common errors.

Given the critical level of quality needed in applications such as training in health sector, the immersive quality of the experience is very important. This necessitates a larger set of high quality 3D models, textures and other data including streaming videos be transferred from the virtual world file server to participants in the immersive training session. Therefore good broadband networks are fundamental to delivering these learning outcomes.

CONCLUSION

This paper presents innovative technologies in social networking and immersive simulation which can provide significant benefits to Healthcare through a virtual world and community of practice solution which overcomes traditional barriers of firewalls and software downloads to enable rapid universal access to online services.

The principal outcome of this research is the design of an architectural framework to enable traditional ‘face to face’ services including complex team-based simulation training to be moved completely online, thus providing more automation as well as allowing greater accessibility on-demand to those who require it.

REFERENCES

- Biagini M., Joy B. (2011) "an open virtual worlds platform" Proceedings of MindSh@re, SET 2, Rome (Italy) July.
- Bruzzone A.G., (2007) "Challenges and Opportunities for Human Behaviour Modelling in Applied Simulation", Keynote Speech at Applied Simulation and Modelling, Palma de Mallorca, August.
- Eldabi, T., Young T. (2007) "Towards a framework for healthcare simulation", Proceedings of the 2007 Winter Simulation Conference.
- Eppich, W., Howard, V., Vozenilek, J., Curran, I. (2011) "Simulation-Based Team Training in Healthcare", Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare: August 2011 - Volume 6 - Issue 7 - pp S14-S19.
- Productivity Commission (2005), 'Australia's Health workforce', research report, Canberra.
- Bergero, C., Hargreaves, L., Nichols, A, (2012) "Collaborative Healthcare Immersive Learning Dynamic: Transitioning to Simulation-Based Learning", Clinical Nurse Specialist: January/February - Volume 26 - Issue 1 - p 42-4.
- Stone RJ. The (human) science of medical virtual learning environments. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 2011 Jan. 27;366(1562):276-85.
- Stott D. Student BMJ: Attending medical school in virtual reality. Student BMJ. 2007
- Moore GE. Cramming More Components Onto Integrated Circuits, Electronics, April 19, 1965.

- Guise V, Chambers M, Välimäki M. What can virtual patient simulation offer mental health nursing education? *J Psychiatr Ment Health Nurs*. 2012 Jun.;19(5):410–8.
- Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. *Med Teach*. 2008;30(2):170–4
- Wenger, E. (2000). *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press
- Lave, J., Wenger, E. (2012) *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press
- Guhde, J. (2010). Using online exercises and patient simulation to improve students' clinical decision-making. *Nursing Education Perspectives*, 31(3), 387-389.
- Edelbring, S., Dastmalchi, M., Hakan, H., Lundberg, I., Owe Dahlgren, L. (2011). Experiencing virtual patients in clinical learning: a phenomenological study. *Advances in Health Science Education*, 16:331-345.
- Bond, F., Lammers, R., Spillane, L., Smith-Coggins, R., Fernandez, R., Reznek, M., Vozenilek, J., Gordon, J. (2007). The use of simulation in emergency medicine: a research agenda. *Academic Emergency Medicine*, 14(4), 353-363
- Kneebone, R. (2005). Evaluating Clinical Simulations for Learning Procedural Skills: A Theory-Based Approach. *Academic Medicine*, 80(6), 549-553.
- Redfern, S., Naughton, N. (2002) Collaborative Virtual Environments to Support Communication and Community in Internet-Based Distance Education. *Journal of Information Technology Education*, 1(3), 201 – 211.
- Alverson, D., Saiki, S., Kalishman, S., Lindberg, M., Mennin, S., Mines, J., Serna, L., Caudell, T. (2008). Medical Students Learn Over Distance Using Virtual Reality Simulation. *Simulation in Healthcare*, 3(1), 10 – 15.

AUTHORS BIOGRAPHY

Bruce Joy: is an innovator and entrepreneur and currently serves as Chairman of VastPark. A veteran of film, media & web, Bruce Joy has run businesses for almost 20 years and worked with many the biggest brands in the world. In 2003, he set up an R&D team to create an immersive multiuser platform for rapid visual prototyping and collaboration. He founded VastPark in 2007. Bruce studied VC finance at HAAS, UC Berkeley and his interests include technology, trends and a smarter, healthier world. He has spoken at conferences around the world including FCVW, Virtual Worlds, SXSW, Military Training and Simulation Asia, ITEC, and Defense GameTech.

Susan Harrison: is General Manager at the Australian Centre for Health Innovation. Susan has considerable experience in private health in operations, logistics and process modelling. She has been involved in IT

implementation and its use in improving productivity, prosthesis management, facilities management and supply chain logistics. Susan's interest is in the practical use and leveraging of IT resource to help address health care demand in preventative health, communication, and patient management.

Tim Gray: is Associate Director of Simulation Education at the Australian Centre for Health Innovation. Tim is the retrieval coordinator for Adult Retrieval Victoria and has 12 years of experience in the development and running of interdisciplinary simulation-based healthcare education programs. He has a particular interest in critical care education and rural/regional trauma, remote clinician critical care telemedicine support and interdisciplinary trauma team training.

Marco Biagini: is a PHD Program researcher in Mathematics Engineering and Simulation at University of Genoa. His research is Modeling & Simulation (M&S) applied to Augmented Reality, Virtual Worlds and Command and Control in crowdsourced interoperable environments for military and civilian applications. He worked at the Italian Army Simulation and Validation Center as M&S Proponent Officer (OF 3) in the M&S Branch for ten years. As section chief he was responsible for many projects regarding Constructive Simulation and Command and Control systems to support Commanders and Staff training.

THREE-DIMENSIONAL NUMERICAL SIMULATIONS OF THE AORTIC FLOW IN PRESENCE OF A LEFT VENTRICLE ASSIST DEVICE

R. Mazzitelli^(a), A. Renzulli^(b), G. F. Serraino^(c), G. Fragomeni^(d)

^{(a)(d)}School of Computer and Biomedical Engineering, "Magna Graecia" University, Catanzaro – ITALY

^{(b)(c)}Cardiac Surgery Unit, "Magna Graecia" University, Catanzaro – ITALY

^(a)mazzitelli@unicz.it, ^(b)renzulli@unicz.it, ^(c)filiberto@live.it, ^(d)fragomeni@unicz.it

ABSTRACT

Cardiovascular diseases are considered to be the main cause of death in developed countries. Due to limitations resulting *in-vivo* measurements of velocity, the analysis and evaluation of hemodynamic parameters by means of computational simulations become the only efficient solution, especially in pathological or assisted condition cases.

The aim of this study is to analyze, through a CFD model, the hemodynamic effects of a continuous flow Left Ventricle Assist Device (LVAD) evaluated in three different working conditions.

The proposed model provides a computational fluid dynamics analysis of LVAD inserted into the descending aorta in order to evaluate flow distribution at the most interesting aorta branches.

The results show the development of the flow in three different cases, showing that the LVAD investigated is able to regulate flow in the aorta.

Keywords: Left Ventricular Assist Device (LVAD), Aortic flow, CFD, Cardiovascular modeling.

1. INTRODUCTION

Cardiac failure has been defined as the epidemy of third millenium. The high number of patients requiring heart transplant for end stage cardiac failure shows that replacement therapy is not a real option for all patients. Therefore mechanical assistance is becoming a valid option to bridge patients to cardiac transplantation and to a destination therapy. In 2002, Dr Westaby and colleagues (Westaby *et al.* 2002) implanted axial flow pump from the apex of the left ventricle to the descending aorta (Jarvik Heart Inc., New York). Low incidence of postoperative complications and totally implanted system allowed an increasing number of patients treated with Jarvik Heart. Some of them, reached follow-up above five years from the implantation. However, the main question in patients of Jarvik Heart is represented by the flow pattern to the brain with a source of retrograde flow.

Computational flow study of hemodynamics problems is extremely challenging, and the potential

advantage of computational tools for improving health care is huge. More Computational Fluid Dynamics (CFD) studies about mechanical hearts, heart valves or assist devices have been performed in the last years. CFD is a valid alternative to experimental methods because it can produce flow field data in great detail. Computational analysis can also give helpful information to optimize mechanical devices design. Costs and times are lower than those required by an empirical approach (Kiris *et al.* 1998; Bazilevs *et al.* 2009).

The aim of this study is the evaluation of the pattern of flow in the sovra-aortic vessels and descending aorta in different condition of flow generated by LVAD implanted with an outflow in the descending aorta.

2. MATERIALS AND METHODS

LVADs are now part of routine devices at surgeons disposal for mechanical support in case of failing heart (Selzman *et al.* 2007). An LVAD is a life saving tool used when the natural heart is unable to provide sufficient blood flow. This device is a compact pump that is helpful for both bridge to transplantation and permanent mechanical circulatory support.

An implantable LVAD has to be small and efficient, generating 5 l min^{-1} blood flow rate. Since the instrumentation for flow measurements is extremely difficult, it was helpful to analyze hemodynamic behavior by using computational techniques (Kiris *et al.* 1998).

LVAD produces two different hemodynamics effects:

- It causes an end-diastolic pressure decreasing in the left atrium;
- The systemic flow rate increases and, as a consequence, the systemic venous return and the right ventricle preload increase as well.

The device considered in this work is Jarvik 2000 Heart (Jarvik Heart Inc, New York, NY), a compact axial flow impeller pump implanted within the apex of

the failing left ventricle that is positioned inside the descending thoracic aorta (Westaby *et al.* 2002).

The effect of an LVAD on hemodynamics is complex and demands a three-dimensional model of the aorta. For this study the geometrical model of the aortic arch has been constructed from Magnetic Resonance Images (MRI) and numerical simulations have been performed by means of commercial CFD software.

2.1. Geometric model

The geometry of the aorta is based on NURBS mesh reconstruction by using a commercial CAD software.

The geometric aortic model is generated as a smooth circular tube surface based on a given parametric geometry (Fung *et al.* 2008).

Blood vessels are tubular objects. Therefore, to reconstruct the aorta model the sweeping method has been used, as shown in Figure 1. This method is based on creating a path (Figure 1a), which constitutes the skeleton of the aorta. On the path, a template circle is translated and rotated to each cross-section in order to obtain the three-dimensional geometry (Figure 1b) (Zhang *et al.* 2007).

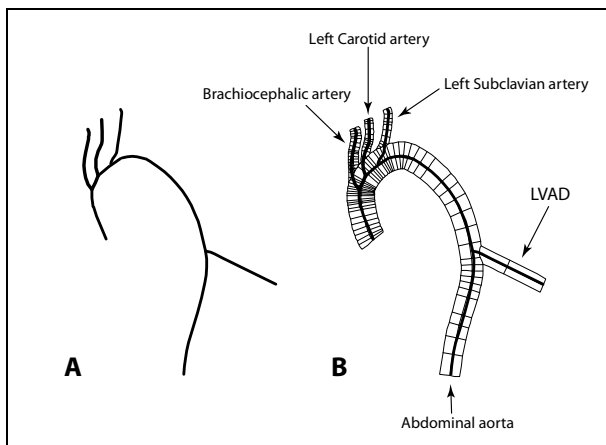


Figure 1: The aorta model reconstructed with skeleton-based sweeping method. Path (a); the sweeping method (b).

An LVAD branch was added to the model in the aorta descending part (Bazilevs *et al.* 2009).

The CAD geometry of the 3D model was used as input for the CFD simulations.

The simulations, preprocessing, meshing, solving, and post-processing were performed by using a commercial finite element analysis software (COMSOL 4.2a, COMSOL Inc, Stockholm, Sweden).

2.2. Governing equations

Blood flow in large arteries is typically well approximated as a Newtonian fluid (Fung *et al.* 2008). The fluid is assumed to be incompressible and homogeneous, with a Newtonian behavior (Shipkowitz *et al.* 2000).

The blood flow in this model was predicted using the three-dimensional Navier-Stokes equations for incompressible flow (Equation (1), (2)).

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u} \cdot \nabla) \mathbf{u} = \nabla \cdot \left[-p \mathbf{I} + \mu(\nabla \mathbf{u} + (\nabla \mathbf{u})^T) \right] + \mathbf{F} \quad (1)$$

$$\rho \nabla \cdot \mathbf{u} = 0 \quad (2)$$

where ρ is the fluid density, $\mathbf{u} = (u, v, w)$ is the fluid velocity, p is the pressure, \mathbf{I} is the unit diagonal matrix, μ is the kinematic viscosity, and \mathbf{F} is the volume force.

Aortic wall was assumed to be rigid and therefore no-slip condition was applied at the aortic wall (Tse *et al.* 2011).

2.3. The mesh grid

On the basis of the geometry described above, the model is discretized with tetrahedral elements.

The final mesh, used in the whole flow simulation, consisted of 180,181 elements and 135,822 degrees of freedom.

The system of linear algebraic equations were solved using the in-built direct sparse solver, PARDISO. A desktop computer with 2.53 GHz Intel Xeon 64, Quad Processor and 8.0 GB of RAM was used in all computations.

2.4. Blood flow parameters and boundary conditions

Time-varying flow rates boundary conditions are applied at the inlet branch in the fluid domain. Three different cases have been considered where the aortic inlet flow decreases with increasing flow rate in LVAD (Figure 2):

- **CASE 1** - LVAD has a flow rate of $16.66 \text{ cm}^3/\text{s}$ (1 l min^{-1}) and flow occurs through the aortic root.
- **CASE 2** - LVAD is operating in the regime where over one half of the blood supplied to the aorta comes from the pump, with a flow rate of $33.33 \text{ cm}^3/\text{s}$ (2 l min^{-1}).
- **CASE 3** - LVAD is operating in the regime where nearly all the flow comes from the LVAD with a flow rate $66.66 \text{ cm}^3/\text{s}$ (4 l min^{-1}), while the aortic valve is closed and the flow is null.

The fluid is a blood substance with a density $\rho = 1060 \text{ kg} \cdot \text{m}^{-3}$ and dynamic viscosity $\mu = 0.0035 \text{ Pa} \cdot \text{s}$ (Park *et al.* 2007).

The flow is limited to the flow model without considering the heat transfer and the temperature is constant at $38 \text{ }^\circ\text{C}$.

The pulsatile flow rates waveform applied to the inlet aorta have been derived, analytically, from literature data (Kim *et al.* 2009, Olufsen *et al.* 2000).

The waveforms of flow rate have been obtained according to the equation (3):

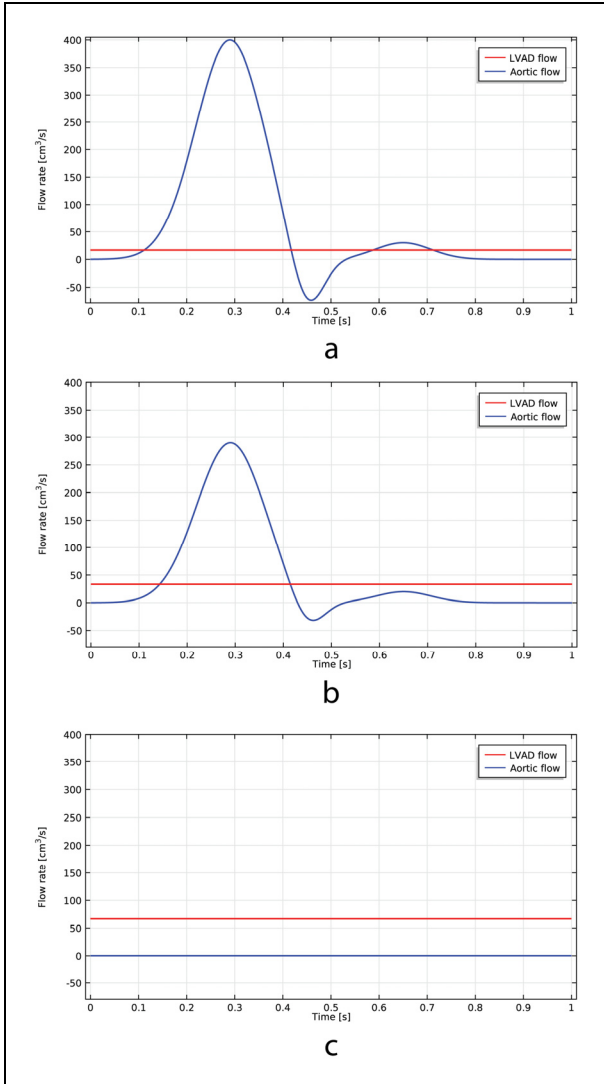


Figure 2: Inlet flow waveforms of the aorta (blue line) and of the LVAD (red line) in presence of an LVAD flow rate of $16.66 \text{ cm}^3/\text{s}$ (a) $33.33 \text{ cm}^3/\text{s}$ (b) and $66.66 \text{ cm}^3/\text{s}$ (c).

$$\sum_{i=1}^n a_i \cdot e^{-b_i(c_i-t)^2}, \quad n = 3 \quad (3)$$

where a_i , b_i and c_i are constant, and t is the time variable. The course of the function is shown in figure 2.

For the fluid domain, the boundary conditions at the outlet branches are set as shown in Figure 3 (Bazilevs *et al.* 2010). The boundary conditions are described in detail in the following:

$$P_n = C_r \cdot q + p_0 \quad (4)$$

where C_r is a constant resistance, q is the volumetric flow rate, and p_0 is the physiological pressure level.

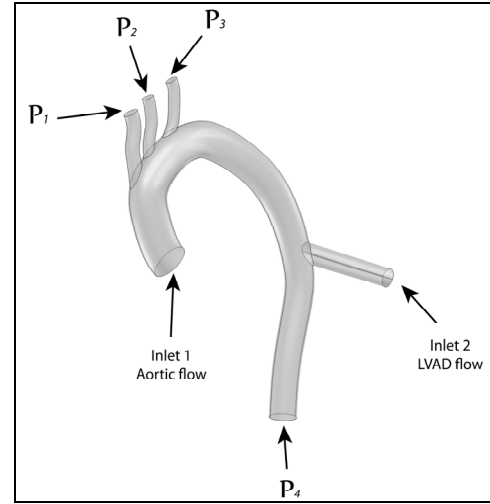


Figure 3: CAD Model and boundary conditions for the fluid domain.

In the performed simulations, only two cardiac cycles have been considered. The output was taken at the second cycle in order to damp the unstable effect of initial cycle (Tse *et al.* 2011).

Based on the steady-state second cycle, in the subsequent paragraph numerical analysis will be discussed.

3. NUMERICAL SIMULATIONS AND RESULTS

3.1. Computational results

Numerical results obtained for the descending aortic distal anastomosis are in agreement with clinical observations and findings for this configuration.

This study was undertaken to assess the effect of blood flow in the outlets branches of the aorta in patients with LVAD. Referring to figure 3, the following outlets branches of the aorta have been considered: (P1) brachiocephalic artery, (P2) left carotid artery, (P3) left subclavian artery. Also, the outflow of the abdominal aorta (P4) was considered.

The flow field in these arteries is more difficult to study and depends on vessel geometry.

The results are a numerical comparison of the hemodynamic response in the cardiovascular system that describes the values of the flow-rate output of the arteries in the three configurations considered.

In figure 4, the values relating to the outlet flows are shown.

In figure 4a, blood flow through the aortic branches during the first work conditions has been illustrated. In correspondence of the systolic peak, an increasing flow in the abdominal aorta and in the brachiocephalic artery can be seen. Inside the carotid and subclavian arteries the same peak is shifted in time.

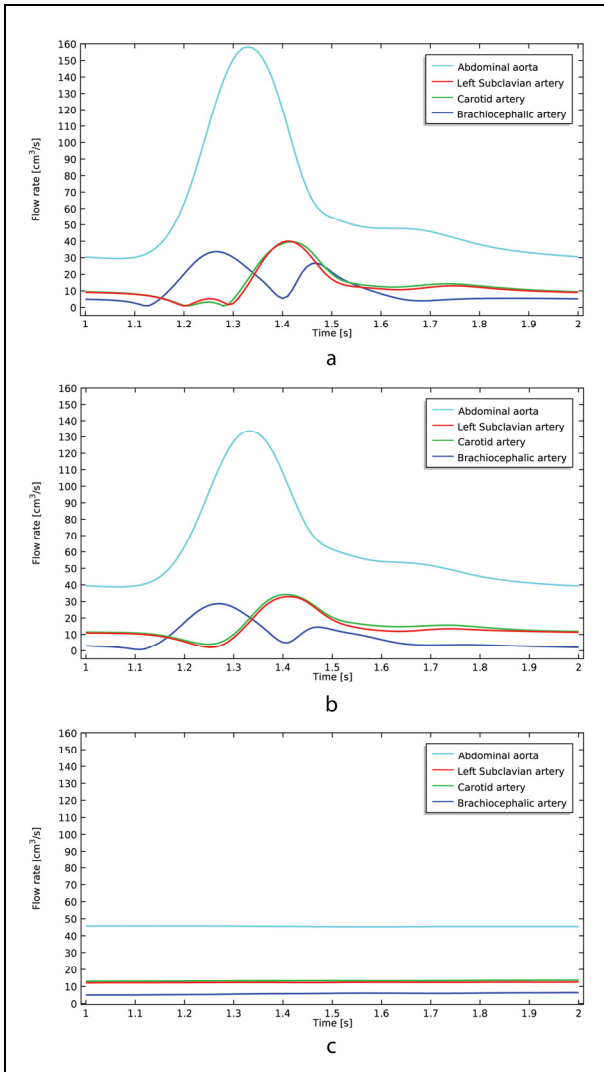


Figure 4: Flow rate distribution among outflow branches during a heart cycle in presence of an LVAD flow rate of $16.66 \text{ cm}^3/\text{s}$ (a) $33.33 \text{ cm}^3/\text{s}$ (b) and $66.66 \text{ cm}^3/\text{s}$ (c).

Figure 4b describes the outlet values of the flows assuming the second condition. These values are slightly different from those reported in figure 4a. It is possible to note a decrease of the systolic peak in all the four ducts; during the diastolic phase, only a little increase of the flow in the subclavian artery, carotid and thoracic aorta appears.

Figure 4c describes the outlet values of the flows considering the third condition, in which the flow comes almost entirely from LVAD. In this case, the output values keep constant values.

In Table 1 the mean outlet flow rate values of a cardiac cycle are reported, by considering the various aortic districts. Values are expressed in liters/minute, for each of the three cases analyzed. These values are referred to the second cardiac cycle.

Table 1: Flow rate through the peripheral vessels. The values are expressed with l/min dimensions.

Flow rate [l/min]			
	Case 1	Case 2	Case 3
Brachiocephalic artery	0.57	0.42	0.30
Carotid artery	0.69	0.75	0.67
Subclavian artery	0.64	0.68	0.61
Abdominal aorta	3.08	3.13	2.40

Figure 5 shows the velocity vector plot of the flow lines for all the cases considered during the maximum value of the systolic phase ($t=1.3 \text{ s}$). Figures 5a and 5b illustrate how the flow in the aorta is controlled by the pulsed input signal and maintains a laminar profile,

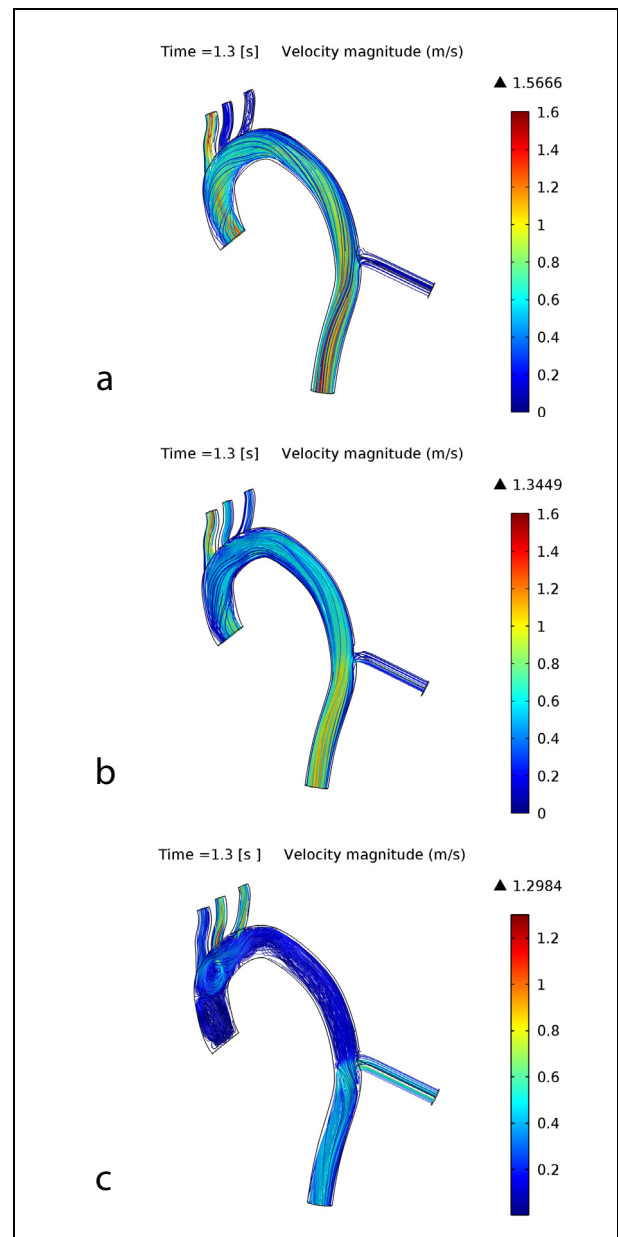


Figure 5: Velocity magnitude streamlines at systolic peak in presence of an LVAD flow rate of $16.66 \text{ cm}^3/\text{s}$ (a) $33.33 \text{ cm}^3/\text{s}$ (b) and $66.66 \text{ cm}^3/\text{s}$ (c).

while in Figure 5c it can be seen that the blood, reaching entirely from the LVAD, creates vortices in the aortic arch and in the closeness of the brachiocephalic artery; these vortices are due to the closure of the aortic valve.

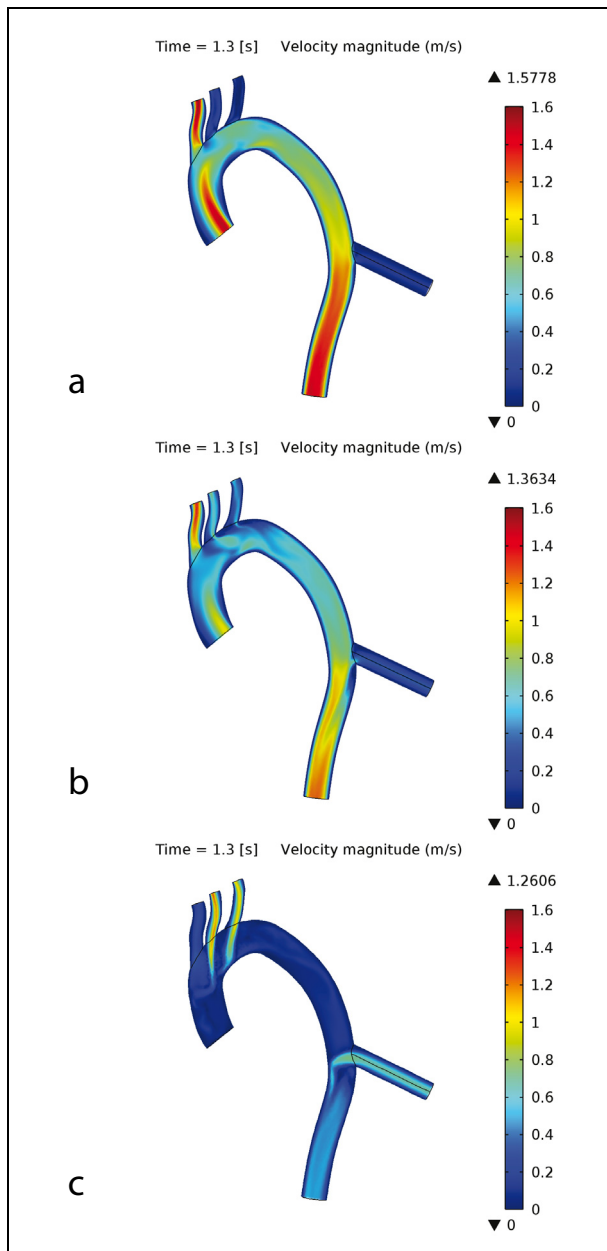


Figure 6: Instantaneous distributions of velocity in the cross-sections, during the systolic peak in presence of an LVAD flow rate of $16.66 \text{ cm}^3/\text{s}$ (a) $33.33 \text{ cm}^3/\text{s}$ (b) and $66.66 \text{ cm}^3/\text{s}$ (c).

Figure 6 shows instantaneous distributions of velocity in the cross-sections within the aorta during the maximum systolic peak.

In Figure 6a it is possible to observe a parabolic evolution of the flow into the aorta, with velocity value greater than in the other two cases (figure 6b-c). Therefore, in the aorta velocity decreases with the increasing of the flow rate in the LVAD.

The corresponding values of average velocity during a cardiac cycle for the three cases are listed in Table 2.

Table 2: Average velocity in output through the peripheral vessels during a cardiac cycle. The values are reported in m/s .

Velocity Magnitude [m/s]			
	Case 1	Case 2	Case 3
Brachiocephalic artery	0.28	0.21	0.13
Carotid artery	0.50	0.49	0.45
Subclavian artery	0.46	0.46	0.41
Abdominal aorta	0.38	0.36	0.27

DISCUSSION

Nowadays, it is plausible that the onset of disease is also closely connected to the alteration of the fluid structures of the cardiovascular system. In order to study and interpret quantitatively these relationships, a certain degree of control over the fluid-dynamic variables, both versus time and versus space is essential. Unfortunately, conventional methods of investigation, that are based on theoretical or *in vitro* models, have proved inadequate or, only partially, effective in supporting this type of analysis. In this context, an alternative approach is provided by models of CFD.

In recent years, many researches have been oriented to the mathematical representation of vascular LVAD, to the description of the model of blood rheology, to the complex multi-layer structure of the vascular tissue and to possible diseases, with the chemical and mechanical interactions among blood and vessel walls.

The employ of computational models can accurately reproduce the conditions of circulatory and ventricular origin and evaluate the related effects of ventricular assistance device under different operating conditions.

These applications are a useful test bench for numerical optimization methods available for studying in order to ensure cardiologists and cardiac surgeons the necessary support for the study and prediction of the arterial system behavior in physiological and pathological conditions in the presence of LVAD.

The goal of this paper has been to characterize the dynamics of the blood flow, in the presence of LVAD implanted by a simple suture, in the section of the descending aorta. In this way, it was possible to demonstrate the effectiveness of this device in the regulation of the flow inside the conduit.

By means of transient simulations, it has been possible to verify three different operating conditions, each characterized by a particular continuous flow input to the LVAD. In spite of this variation, a certain equivalence of velocity profiles and outlet flow rate quite balanced in the various branches was noted.

Based on these considerations, it has been possible to describe this phenomenon by means of a quantitative model able to predict the behavior of the aortic blood

flow in the section, in the presence of such a device, in order to provide a useful tool in the clinical field.

Appropriate simulations, indeed, allow pre-assessments from both the diagnosis and the achievable therapies points of view, since they guarantee correct pre-surgery estimates in relation to the choice of the VAD to be implanted and a valuable observation about the artery reactions when exposed to any continuous and pulsatile flows.

CONCLUSION AND FUTURE WORK

A future interesting development could include the analysis of the blood flow based on the placement of the cannula of the VAD, by also considering other surgical techniques, such as the apical positioning.

Furthermore, the model will be considered also in conditions of wall deformation and elasticity in order to characterize, in a suitable way, the behavior of such systems.

ACKNOWLEDGMENTS

The present work is supported by the European Committee (F.S.E.) and by Regione Calabria. This article is co-funded with support from the European Commission, European Social Fund and "Regione Calabria".

This publication reflects only the views of the author, so the European Commission and "Regione Calabria" cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- Bazilevs, Y., Gohean, J.R., Hughes, T.J.R., Moser, R.D., Zhang, Y., 2009. Patient-specific isogeometric fluid–structure interaction analysis of thoracic aortic blood flow due to implantation of the Jarvik 2000 left ventricular assist device. *Comput. Methods Appl. Mech. Engrg.* 198, 3534–3550.
- Bazilevs, Y., Hsu, M.C., Zhang, Y., Wang, W., Kvamsdal, T., Hentschel, S., Isaksen, J.G., 2010. Computational vascular fluid–structure interaction: methodology and application to cerebral aneurysms. *Biomech Model Mechanobiol*, (9), 481–498.
- Fung, G.S.K., Lam, S.K., Cheng, S.W.K., Chow, K.W., 2008. On stent-graft models in thoracic aortic endovascular repair: A computational investigation of the hemodynamic factors. *Computers in Biology and Medicine* (38), 484–489.
- Kim, H. J., Figueroa, C.A., Huges, T.J., Jansen, K. E., Taylor, C.A., 2009. Augmented Lagrangian method for constraining the shape of velocity profiles at outlet boundaries for three-dimensional finite element simulations of blood flow. *Comput. Methods Appl. Mech. Engrg. (Elsevier)* (198) , 3551–3566.
- Kiris, C., Kwak, D., Benkowski, R., 1998. Incompressible Navier-Stokes calculations for the development of a ventricular assist device. *Computers & Fluids* Vol. 27, Nos 5±6, pp. 709±719.
- Olufsen, M.S., Peskin, C.S., Kim, Y., Pedersen, E.M., Nadim, A., Larsen, J., 2000. Numerical Simulation and Experimental Validation of Blood Flow in Arteries with Structured-Tree Outflow Conditions. *Annals of Biomedical Engineering (Biomedical Engineering Society)* (28), 1281–1299.
- Park, J.Y., Park, C.Y., Hwangb, C.M., Sun, K., Min, B.G., 2007. Pseudo-organ boundary conditions applied to a computational fluid dynamics model of the human aorta. *Computers in Biology and Medicine* (37) 1063 - 1072.
- Selzman, C. H., Chang, P. P., Vernon-Platt, T., Bowen, A., Kowalczyk, S., Sheridan, B.C., 2007. Use of the Jarvik 2000 Continuous Flow Left Ventricular Assist Device for Acute Myocardial Infarction and Cardiogenic Shock. *The Journal of Heart and Lung Transplantation* 26, (7), pp.756-758.
- Shipkowitz, T., Rodgers, V.G.J., Frazin, L.J., Chandran, K.B., 2000. Numerical study on the effect of secondary flow in the human aorta on local shear stress in abdominal aortic branches. *Journal of Biomechanics* (33) ,717-728.
- Tse, K.M., Chiu, P., Lee, H.P., Ho, P., 2011. Investigation of hemodynamics in the development of dissecting aneurysm within patient-specific dissecting aneurysmal aortas using computational fluid dynamics (CFD) simulations. *Journal of Biomechanics*, 827-836.
- Westaby, S., Frazier, O.H., Pigotta, D.W., Saito, S., and Jarvik, R.K., 2000. Implant technique for the Jarvik 2000 heart. *The Annals of Thoracic Surgery*, 73, (4) 1337–1340.
- Zhang, Y., Bazilevs, Y., Goswami, S., Bajaj, C.L., Hughes, T.J.R., 2007. Patient-specific vascular NURBS modeling for isogeometric analysis of blood flow. *Comput. Methods Appl. Mech. Engrg.* (196), 2943–2959.

RANSAC-BASED ENHANCEMENT IN DRUG CONCENTRATION PREDICTIONS USING SUPPORT VECTOR MACHINE

Wenqi You^(a), Alena Simalatsar^(b), Giovanni De Micheli^(c)

^{(a)(b)(c)}Ecole Polytechnique Federale de Lausanne

^(a)wenqi.you@epfl.ch, ^(b)alena.simalatsar@epfl.ch, ^(c)giovanni.demicheli@epfl.ch

ABSTRACT

Training Support Vector Machines (SVMs) to predict drugs concentrations is often difficult because of the high level of noise in the training data, due to various kinds of measurement errors. We apply *RANdom SAmple Consensus* (RANSAC) algorithm in this paper to solve this problem, enhancing the prediction accuracy by more than 40% in our particular case study. A personalized sample selection method is proposed to further improve the prediction result in most cases.

Keywords: RANSAC, SVM, drug concentration predictions

1. INTRODUCTION

The decision-making regarding the drug dosage has been one of the main challenges of the pharmacological studies for decades. Population Analysis is a classical method to decide a dosage. It looks at a small number of data points per patient over many subjects [Bourne1995]. The models built by this method are applied to any new patient in clinical practice. However, due to both intra- and inter-differences of patients' characteristics, these models are not always accurate, therefore not applicable to some drugs whose therapeutic ranges are narrow. Furthermore, Population Analysis methods suffer from other limitations, such as not considering specific features, e.g. binary values, and moreover the number of features is limited. The Support Vector Machine (SVM) algorithm has been applied to address these problems [You2011]. Nevertheless, the performance of this algorithm highly relies on the quality of the training dataset. When building the predictor, SVM minimizes a cost function where a mean-squared error is very sensitive to noise in input data. Clinical measurements are particularly faced with the risk of measurement errors.

RANdom SAmple Consensus (RANSAC) is a general parameter estimation method proposed to filter out the outliers (errors) from input data [Fischler1981]. It resamples the input data and generates candidate solutions with respect to a minimum number of observations (data points) required to estimate the underlying model parameters. Depending on a threshold value, the input data are classified with different proportions into inliers (good data) and outliers. Only

the inliers are considered to be useful to build the SVM model for drug concentration predictions. Unlike other sampling techniques that use as many data as possible to obtain an initial solution and then prune the outliers, RANSAC uses the smallest set possible and then enlarges this set with consistent data points [Fischler1981]. It has been applied to various domains such as sensor networks [Buttyan2006, Furukawa2006, Shafique2008], Integrated Chip (IC)'s three-dimensional information recognition [Liang2011] etc.

In this paper, we use the RANSAC algorithm to filter the datasets before running a Support Vector Machine (SVM) algorithm. Compared to an SVM-based algorithm in [You2011] and the Pharmacokinetic (PK) method [Widmer2006], it enhances the prediction by more than 40% in our experiments. Two scenarios for personalized predictions have also been tested to further improve the prediction accuracy.

The paper is organized as follows: Section 2 presents the methodology used in this paper. Section 3 shows the experimental results and comparisons with previous works. Finally, Section 4 draws a brief conclusion.

2. RELATED WORK

In the literature, predictions of drug concentrations are usually carried out using analytical models. These models are built based on some assumptions that the system of human body is one-, two- or three-compartment [Bourne1995, Bailey1991, Hahn2011]. This kind of assumption is widely applied to clinical practices nowadays, but suffers from some drawbacks such as it cannot take into account binary numbers and such as it is difficult to modify (add or remove) a parameter in the model. Thus, in [You2011], a Support Vector Machine (SVM)-based approach was proposed to overcome these drawbacks and tried to enhance the prediction accuracy.

SVM [Boser1992] was introduced by Boser, Guyon and Vapnik and became rather popular in several domains e.g. pattern recognition, computer vision, etc. It is a supervised learning model with associated learning algorithms that analyze data and recognize patterns. It has been successfully applied to human detection [Dalal2005], object recognition [Pontil1998],

Algorithm 1 RANSAC algorithm, where $data$ is a set of observations, $model$ is a model that can be fitted to data, K is the minimum number of data points required to fit the model parameters, N is the number of trials performed by the algorithm, T is a threshold determining if a data point fits a model, and $bestmodel$ is the model fitting the highest number of data points.

Input: $data, model, K, N, T$
Output: $bestmodel$

```

bestinliers  $\leftarrow \emptyset$ 
for  $i = 1 \rightarrow N$  do
  possibleinliers  $\leftarrow$  SampleUniformly( $data, K$ )
  possiblemodel  $\leftarrow$  Fit( $model, possibleinliers$ )
  inliers  $\leftarrow \emptyset$ 
  for all  $point \in data$  do
    if Distance( $point, model$ )  $< T$  then
      inliers  $\leftarrow inliers \cup \{point\}$ 
    end if
  end for
  if  $|inliers| > |bestinliers|$  then
    bestinliers  $\leftarrow inliers$ 
  end if
end for
return  $bestmodel \leftarrow$  Fit( $model, bestinliers$ )

```

image classification [Chapelle1999], etc. It is simple in computation but also robust in data classification and regression, compared with other common machine learning methods, e.g. decision trees, neural networks, etc. However, according to our survey, SVM has not yet been applied to estimating drug concentrations.

To further enhance the performance of SVM, several previous works applied the *RANdom Sample Consensus* (RANSAC) algorithm. RANSAC, proposed by Fischler and Bolles [Fischler1981], is a general parameter estimation method used to deal with a large proportion of outliers in the input data. It was developed within the area of computer vision and applied to many other domains for data analysis. In [Nishida2008], the author claims a reduction of the computation requirement to about 1/170 compared with SVM libraries, and in [Kuo2007], RANSAC algorithm was used in a fine-selection stage for face recognition and achieved a lowest mean error rate.

3. METHODOLOGY

In this section, we will first introduce RANSAC algorithm that is used to improve drug concentration predictions together with Support Vector Machine (RANSAC-SVM). We will then propose a personalized drug concentration prediction scheme based on the RANSAC algorithm with two clinical scenarios.

3.1. RANSAC Algorithm

The RANSAC [Fischler1981] algorithm works as described in Algorithm 1. The number of trials N is set to be big enough to guarantee that at least one of the

Algorithm 2 RANSAC-based personalized algorithm, where $training$ is a set of M training samples, $newpatients$ is an ordered set containing one sample per new patient, Y is the index of a particular feature, $model, K, N, T$ are parameters of the RANSAC algorithm, and $bestmodels$ is an ordered set containing the SVM model fitting the best each new patient.

Input: $training, newpatients, F, model, K, N, T$
Output: $bestmodels$

```

bestmodels  $\leftarrow \emptyset$ 
for all  $patient \in newpatients$  do
   $data \leftarrow \{patient, \dots, patient\}$   $\{|data| = M\}$ 
   $data \leftarrow data \cup training$ 
  inliers  $\leftarrow$  RANSAC( $data, model, K, N, T$ )  $\{\text{predict } Y\}$ 
  inliers  $\leftarrow inliers \setminus \{patient\}$ 
  model  $\leftarrow$  SVM( $inliers$ )
  bestmodels  $\leftarrow bestmodels \cup model$ 
end for
return  $bestmodels$ 

```

sets of possible inliers does not include any outlier with a high probability p . Usually p is set to 0.99. Let us assume, that u is the probability that any selected data point is an inlier, then $v=1-u$ is the probability of selecting an outlier. N trials of sampling each K data points are required, where $1-p = (1-u^K)^N$. This way:

$$N = \frac{\log(1-p)}{\log(1-(1-u)^K)} \quad (1)$$

The model of the RANSAC algorithm is a linear combination of several basis functions. The number of basis functions corresponds directly to the minimum number of points K required to fit the model. The parameters of the model are the weights of each basis function. In this paper, the drug concentration prediction method enhanced with filtering of the training dataset using RANSAC algorithm is called RANSAC-SVM method.

3.2. RANSAC-based Personalization

In Algorithm 1, inliers and outliers are separated without considering the information of a new patient. The SVM predictor for any new patient is estimated with the same inliers chosen as the training data. However, we believe that it might happen that the set of inliers for one patient is actually a set of outliers for another. Therefore, a predictor built out of the same set of inliers for a number of new patients might not be applicable for some others. Hence, it is important to find an individual set of inliers for each patient. So we propose to use RANSAC-based personalization method to solve this task. Previously [You2011], a 'closest point' strategy has been used which, despite using a much fewer number of training points (up to 30% of the total number), retains the initial performance of the original SVM (<3% degradation). However, that strategy needs a set of predefined weights for each feature in order to select the 'closest point'.

Assume that we already have M samples from previous patients (our training dataset), for each new

patient we want to find a best subset of samples of M to train the SVM. To do so we treat them as if they were noisy samples of a new patient, and use a RANSAC to remove the outliers. The whole procedure is detailed in Algorithm 2. The new sample from each patient is first replicated M times to make sure that it will always be considered as an inlier. Then RANSAC is applied to those replicated samples plus to the original M training samples in order to predict the feature Y as a linear combination of basis functions of the remaining features X . The new patient sample is then removed from the inliers and finally an SVM is trained on the remaining original training samples to predict the drug concentration of this new patient.

Two clinical scenarios can be applied here: the target feature Y is set as:

1. Any feature other than the concentration value.
2. The measured drug concentration.

For scenario 1, no invasive blood test is required, while in 2 the drug concentration value should be measured after the first dosing.

4. EXPERIMENTS AND COMPARISONS

The experiments are conducted on a set of data collected during patients' treatment with *imatinib*, a drug designed to treat chronic myeloid leukemia and gastrointestinal stromal tumors [Widmer2006]. The training dataset consists of 54 patients and 252 samples, while the validation (testing) dataset contains the data of 65 patients and 209 samples.

To apply RANSAC, we first preset the basis using some typical functions: $\{x^{-2}, x^{-1}, x, x^2, x^3, \log(x), \cos(x), (1-e^{-x}), e^x\}$. This requires at least $K=9$ data points to estimate the parameters. However, not all the listed basis functions are useful to get the final model of drug concentration. Table 1 shows the experimental results on each basis function with respect to different thresholds (tolerable difference between the measured concentration and the predicted one). In practice, we set the threshold to be as small as possible to minimize the difference between the measured concentration values and the predicted ones. Hence, we combine the first two rows of the chosen basis functions (scored '1') in Table 1: $f(x) = \{x^{-2}, x, x^3, \log(x), \cos(x), (1-e^{-x}), e^x\}$. Figure 1 shows the AUC (Area Under the drug Concentration) curve estimated using RANSAC, the green points denote inliers and the blue represent outliers.

After determining the basis functions, the drug concentrations over the validated dataset are predicted via SVM algorithm. We evaluate the drug concentration prediction results of three algorithms (the traditional Pharmacokinetic (PK) [Widmer2006], SVM-based [You2011], and the proposed RANSAC-SVM) by computing an Absolute Difference between the Predicted concentration values and the Measured ones (ADPM). In practice, we expect ADPM values to be small. In our experiments, the RANSAC-SVM algorithm enhances the prediction performance by about 44.7% over the PK method and 42.6% over SVM-based

Table 1: RANSAC Basis Function Analysis With Respect To Different Thresholds. (T: Threshold with unit [mg/L]. '0' stands for 'unused' and '1' for 'in use'.)

T	x^{-2}	x^{-1}	x	x^2	x^3	$\log(x)$	$\cos(x)$	$1 - \exp(-x)$	$\exp(x)$
250	1	0	1	0	1	1	1	1	0
500	0	0	1	0	0	1	1	0	1
1000	0	1	0	0	0	0	1	1	0
1500	0	1	0	0	0	0	0	0	0

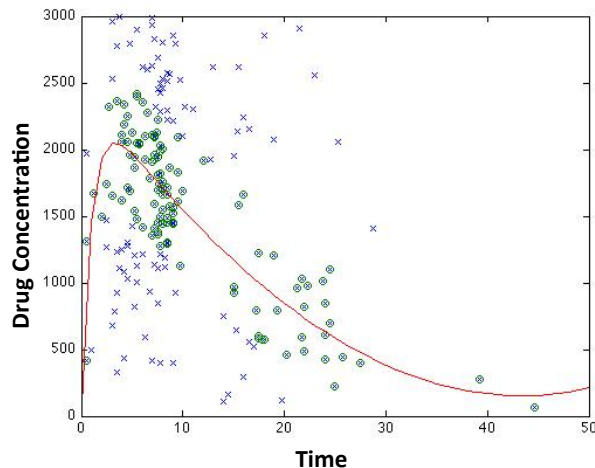


Figure 1: AUC (Area Under the plasma concentration time Curve) Using RANSAC Analysis. Green points are inliers and blue points are outliers.

method, respectively. Around 71% of mean ADPM values of RANSAC-SVM results are smaller than 500mg/L, while this number decreased to around 50% for the PK and SVM-based methods.

For further prediction improvement, we apply two personalization scenarios with RANSAC algorithm (see Section 2). By choosing different features as X and Y to select the individual set of inliers for each new patient, we obtain the results shown in Table 2. In *imatinib* case study, the following features are available: {Measured Drug Concentration (MDC), Measuring Time (MT), Drug Dosage (DD), Age (A), Gender (G), and Body Weight (BW)}. Scenario 1 uses any feature other than MDC values while scenario 2 uses only MDC to be Y . We also compute the enhancement percentages with SVM [You2011] (shown as 'v.s. SVM' in the table) and Bayesian algorithm [Widmer2006] (shown as 'v.s. BAYE' in the table) with the Mean and STD results. The number of prediction samples whose predicted results are greater than 500mg/L from the measured values is denoted as '>500' in the table.

The Table 2 shows that the RANSAC-based personalization performs slightly better than RANSAC-SVM algorithm in many cases (1, 2, 3, 4, 5, 6), which results from a reduced number of predictions whose ADPM values are larger than 500mg/L. Both algorithms improve the prediction accuracy compared with SVM by around 40%. In scenario 2, BAYE outperforms the other two in the average prediction values in cases 7, 8, 9. However, in most cases, BAYE gives a larger STD value in that the predictions by BAYE deviate more from the measured values, while

Table 2: Comparisons Of The Drug Concentration Predictions Using RANSAC-based Personalization (RPER), RANSAC-SVM (RSVM), SVM [You2011], And Bayesian Estimation (BAYE) [Widmer2006]. (>500': Number of prediction samples that are more than 500mg/L different from the measured values.)

Scenario 1: without blood measurement after first-dosing							
case	Features	Method	Mean	v.s. SVM	STD	v.s. SVM	> 500
1	Y =BW	RPER	258.60	42.10%	239.97	39.62%	6
	X =MT	RSVM	261.00	41.56%	240.64	39.45%	8
2	Y =A	RPER	222.39	50.21%	168.56	57.59%	3
	X =MT	RSVM	224.89	49.65%	168.40	57.63%	4
3	Y =G	RPER	282.21	36.81%	258.92	34.85%	7
	X =MT	RSVM	283.90	36.44%	260.56	34.44%	9
4	Y =DD	RPER	212.72	52.37%	170.93	56.99%	1
	X =MT	RSVM	213.35	52.46%	170.99	56.97%	1
5	Y =A	RPER	235.68	47.23%	181.60	54.31%	6
	X =BW	RSVM	243.79	45.42%	184.47	53.58%	6
Scenario 2: with blood measurement after first-dosing							
case	Features	Method	Mean	v.s. BAYE	STD	v.s. BAYE	> 500
6	Y =MDC	RPER	229.59	8.02%	211.34	43.65%	2
	X =MT	RSVM	239.58	4.01%	202.48	46.01%	3
7	Y =MDC	RPER	244.67	-17.74%	168.92	20.09%	10
	X =DD	RSVM	232.77	-12.02%	168.79	20.16%	7
8	Y =MDC	RPER	401.31	-103.42%	363.09	-89.80%	21
	X =G	RSVM	243.98	-23.67%	184.74	3.43%	7
9	Y =MDC	RPER	279.63	-15.30%	196.42	46.58%	8
	X =A	RSVM	247.46	-2.04%	199.13	45.84%	5
10	Y =MDC	RPER	219.33	8.18%	173.5	53.53%	3
	X =BW	RSVM	212.23	11.15%	155.22	58.43%	2

the other two algorithms estimate the concentration values without a large deviation. Hence, we can see that the proposed algorithms are robust to predict the concentrations more accurately for any individual patient, while Bayesian algorithm only predicts well for some patients and less accurate for the others.

5. CONCLUSIONS

This paper presents a RANSAC-based SVM algorithm to estimate drugs concentrations. RANSAC filters out the outliers in the input datasets to reduce the number of measurement errors. It exceeds the traditional Pharmacokinetic and SVM-based prediction methods by more than 40% in accuracy. The paper also introduces a way to personalize drug concentration prediction with the RANSAC algorithm. Experiments show that it enhances the proposed initial RANSAC-based SVM algorithm in many cases.

ACKNOWLEDGMENTS

Authors thank Carlotta Guiducci from EPFL for the help in manuscript revision and T. Buclin, N. Widmer and V. Gotta from CHUV Hospital of Lausanne for the precious suggestions on clinical data modeling and provision with sufficient data.

This work was supported by ISyPeM Project, with a grant from Swiss Nano-Tera.ch initiative, evaluated by Swiss National Science Foundation.

REFERENCES

Bailey, J. M. and Shafer, S. L., 1991. A Simple Analytical Solution to the Three-Compartment Pharmacokinetic Model Suitable for Computer-Controlled Infusion Pumps. *IEEE Transactions on*

Biomedical Engineering, vol. 38, no. 6, pp. 522-525.

Bourne, D. W. A., 1995. *Mathematical Modeling of Pharmacokinetic Data*. 2nd ed. Technomic Publishing Company, Inc.

Buttyan, L., Schaffer, P. and Vajda, I., 2006. Ransac-based Resilient Aggregation in Sensor Networks. *Proceedings of the fourth ACM workshop on Security of ad hoc and sensor networks*. Oct. 30.

Chapelle, O., Haffner, P. and Vapnik, V. N., 1999. Support Vector Machines for Histogram-Based Image Classification. *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055-1064.

Dalal, N. and Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. *Proceeding of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886-893.

Fischler, M. and Bolles, R., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, vol. 24, no. 6, pp. 381-395.

Furukawa, Y., Sethi, A., Ponce, J. and Kriegman, D. J., 2006. Robust Structure and Motion From Outlines of Smooth Curved Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 302-315.

Hahn, J. O., Dumont, G. A. and Ansermino, J. M., 2011. Closed-Loop Anesthetic Drug Concentration Estimation Using Clinical-Effect Feedback. *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 3-6.

Kuo, C. H. and Lee, J. D., 2007. A Two-Stage Classifier Using SVM and RANSAC for Face Recognition. *TENCON 2007*, pp. 1-4.

Liang, Z., Ye, B. and Xiao, Z., 2011. Vector Optimization of Integrated Chip Micro-image Chromatic Characteristic. *International Journal of Advancements in Computing Technology*, vol. 3, no. 11, pp. 170-177.

Nishida, K. and Kurita, T., 2008. RANSAC-SVM for Large-Scale Datasets. *Proceeding of the 19th International Conference on Pattern Recognition*, pp. 1-4.

Pontil, M. and Verri, A., 1998. Support Vector Machines for 3D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 637-646.

Shafique, K., Hakeem, A., Javed, O. and Haering, N., 2008. Self Calibrating Visual Sensor Networks. *IEEE Workshop on Applications of Computer Vision*, pp. 1-6.

Widmer, N., Decosterd, L., Csajka, C., Leyvraz, S., Duchosal, M. A., Rosselet, A., Rochat, B., Eap, C. B., Henry, H., Biollaz, J. and Buclin, T., 2006. Population Pharmacokinetics of Imatinib and the Role of α_1 -acid Glycoprotein. *British Journal of Clinical Pharmacology*, vol. 62, no. 1, pp. 97-112.

You, W., Widmer, N. and De Micheli, G., 2011. Example-based Support Vector Machine for Drug Concentration analysis. *Engineering in Medicine and Biology Society, EMBC*, pp. 153-157. Aug. 30 – Sept. 3, Boston, USA.

AUTHORS BIOGRAPHY

Ms. Wenqi You is currently a Ph.D. candidate in School of Computer and Communication Sciences at Ecole Polytechnique Federale de Lausanne, Switzerland. Her thesis is Mathematical Modeling for Decision Support System of Personalized Medicine. She holds a Bachelor degree from School of Electronic, Information and Electrical Engineering at Shanghai Jiao Tong University, China and a Master degree from Graduate School of Information, Production and Systems at Waseda University, Japan.

Dr. Alena Simalatsar is a postdoctoral researcher at Ecole Polytechnique Federale de Lausanne, Switzerland since February 2011. She holds a Bachelor and a Master degrees from the Faculty of Radiophysics and Computer Technologies of Belarusian State University in 2005. She received the Ph.D. degree (2009) in Computer Science and Telecommunication Technologies from University of Trento, during which she also spent six months as a Visiting Scholar at Electrical Engineering and Computer Science Department in University of California at Berkeley (2007).

Dr. Giovanni De Micheli is Professor and Director of the Institute of Electrical Engineering and of the Integrated Systems Centre at Ecole Polytechnique Federale de Lausanne, Switzerland. He is program leader of the Nano-Tera.ch program. Previously, he was Professor of Electrical Engineering at Stanford University. He holds a Nuclear Engineer degree (Politecnico di Milano, 1979), a M.S. and a Ph.D. degree in Electrical Engineering and Computer Science (University of California at Berkeley, 1980 and 1983). Prof. De Micheli is also a Fellow of ACM and IEEE and a member of the Academia Europaea.

SIMULATIONS OF UTERINE ELECTRICAL ACTIVITY USING PARALLEL COMPUTING

Tanguy Hedrich^(a), Jeremy Laforet^(b), Catherine Marque^(c)

^{(a)(b)(c)}CNRS UMR 7338
Biomecanique et Bioingenierie
Universite de Technologie de Compiegne
60200 Compiegne, France

^(a)thedrich@etu.utc.fr, ^(b)jlaforet@utc.fr, ^(c)cmarque@utc.fr

ABSTRACT

The uterine electrical activity can be realistically modeled by representing the principal ionic dynamics at the cell level, the propagation of electric activity at the tissue level.

A simplified model based on the physiology was already presented (Laforet 2011). We were able to simulate 0-dimension to 2.5-dimension grids of muscle cells.

In this study, we implemented this model as an easy-to-use software using oriented-object python, under an open-source license. We developed a parallel integration based on shared memory by using python standard libraries to improve the computational effectiveness of the simulations.

The execution times of parallel computed simulations are up to 25 times less than the serial integration. We observe an effect of the size and the dimension of the grid on the computation time per cell.

Thanks to this improved computation time, further studies will provide bigger and more realistic simulations in a reason-able time.

Keywords: Simulation, Electrophysiological signals, uterus, Python

1. INTRODUCTION

The sequence of contraction and relaxation of the myometrium results from the electrical activity associated to the generation and propagation of cellular action potential bursts. The uterine electrical activity can be measured noninvasively on the abdomen by standard electrodes and it is referred to as electrohysterogram (EHG) [2].

At the myometrial level, the cellular action potential generation and propagation have been previously modeled as function of a large number of electro-physiological parameters related to ionic dynamic (Rihana 2009). This model is physiology-based and has been demonstrated to be representative of the main ionic mechanisms responsible for the genesis and evolution of the uterine electrical activity (Rihana

2009). However, this model is computationally demanding and therefore unsuitable for a direct integration in a multi-scale approach.

In view of this challenging objective, we have defined a new simplified physiology-based model of the electrical activity generation at the cell level and of the electrical propagation at tissue level, based on the full model described in (Laforet 2011).

Unfortunately, even the simplified model is still computationally demanding for large amount of cells. To step closer to obtain full organ simulations in a decent time, we present a parallel implementation of the models integration. We show how this can help us reduce the computation time by a factor up to 24.

2. MODEL

In view of this challenging objective, we define a global multi-scale model, from the myometrial cell to the skin surface (Laforet 2011). This model is physiology based and takes into account the following levels:

- myometrial cell: generation of the electrical activity.
- uterine tissue: propagation of this activity from cell to cell.
- organ level: propagation of the electrical field through the tissue layers from the uterus to the skin surface (Rabotti 2010).
- abdominal level: recording on the skin surface by an electrode grid.

In this paper, we focus on the improvements of the computational effectiveness of the uterine muscle model, as it is the most time consuming step within the global model. We previously presented a simplification of the complete model in (Laforet 2011). Starting from the 6 variables uterine cell reduced model (referred as Red6 model) (Rihana 2009), we obtain a 3 variables model (referred as Red3) defined as follow:

$$\frac{dV_m}{dt} = \frac{1}{C_m} (I_{stim} - I_{Ca} - I_K - I_{KCa} - I_{leak}) \quad (1)$$

$$\frac{dn_K}{dt} = \frac{h_{K_\infty} - n_K}{\tau_{n_K}} \quad (2)$$

$$\frac{d[Ca^{2+}]}{dt} = f_c(-\alpha I_{Ca} - K_{Ca}[Ca^{2+}]), \quad (3)$$

with V_m the transmembrane potential, n_K the potassium activation variable, and $[Ca^{2+}]$ the intracellular calcium concentration.

The Red3 model is generally 50% faster than the Red6 model and will be used as default model in this work. However, the Red6 will be kept for studies on small number of cells when requiring a higher accuracy.

At the tissue scale, the communication between the my-ometrial cells through gap-junctions is modeled by the spatial diffusion of the electrical potential over the cells.

3. IMPLEMENTATION

We implemented both Red6 and Red3 models using Python (version 2.7) in an object oriented approach. The code will be published under an open-source license and its required dependences are only common Python modules beside the standard library (Numpy and Scipy). It can be executed under Mac-OS X , Windows, and GNU/Linux systems.

3.1. Models

First, we defined a generic cell model class from which Red6 and Red3 classes inherit. It enables us to define all common elements in the models only once, in the parent class. This structure is illustrated in figure 1.

At the tissue level, the cells are arranged into a Cartesian grid. This tissue model can be 0D (a single cell), 1D (a cable-like line of cells), 2D (a flat surface), or 2.5D (a flat surface with non-null thickness). The grid is represented directly by a N-dimensional state array where each element represent a cell which is electrically coupled with its direct neighbors (2, 4 or 6 depending on the dimensions).

For 2D and 2.5D tissue model, in order to simulate a simple model of cylindrical geometry, the border in one of the dimensions of the model may be considered as the neighbor of the opposite border.

The dimension of a cell, the membrane resistance, and different physiologic features can be modified for each cell individually along each spatial dimension. This way makes possible to add anisotropy to the model and also to take into account local variations of the parameters. It is also possible to represent non-muscular cells (ie connective tissue). These

cells would be affected by the spatial diffusion as the other cells but the model defined in Eq. 1 would not be applied to generate their own response.

To simulate the pacemaker cells, an external current is applied to the cells that will act as pacemaker. This external current is modeled by a half sine wave to avoid discontinuities. It can be applied to two different areas of the model during simulation.

Finally, we added a two-layer padding on the borders to avoid side effects of the spatial filtering. These 'ghost cells' have a coefficient of diffusion 10^4 times lower than the other cells to efficiently attenuate the signals at the borders.

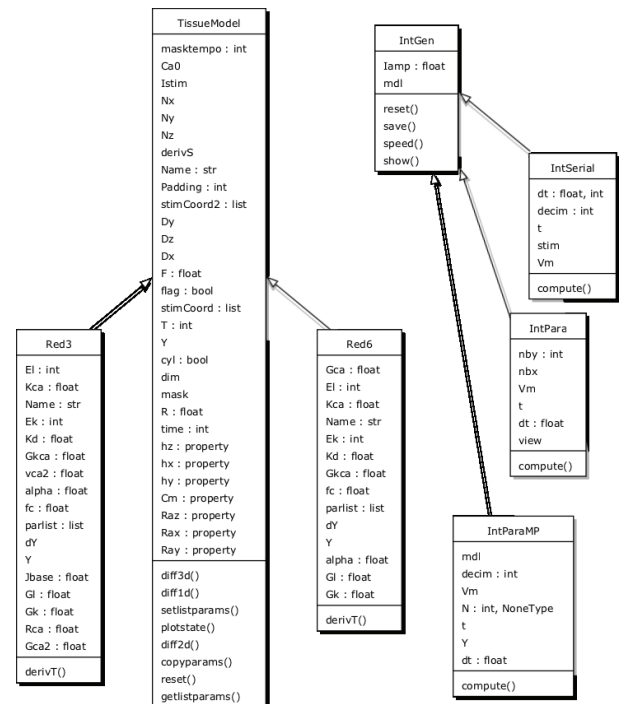


Figure 1: Main classes of uEMG.

3.2. Integration

We implemented two methods to compute the models. The first method is the serial integration, which computes the model by using only one process. This is the usual way to integrate such models, but it is ineffective on modern multi-core CPUs as only one of the cores is used. For a huge number of cells or for a long simulation duration, the computation time will rise to an unacceptable level. To improve the performance of the model (in term of computational time), parallel computing can be used. The idea is to divide the size of the tissue model into several processes which are running in parallel so that the global execution decreases.

A different process computes each part of the tissue model and for each time step, the processes share their results. For this study, we developed an integration model based on shared memory.

1) SERIAL: The serial integration computes the model using the Euler finite difference method. For each time sample, a discrete version of the cell model described in Eq. 1 is applied for every muscle cell of the tissue model.

After the muscle model, for each time step as well, the spatial diffusion model is applied to the cells to represent the electrical communication through gap-junctions. On a Cartesian grid the computation of this diffusion is a spatial Laplacian filtering which global weights are the diffusion coefficients. Those coefficients are dependent on the geometry and the membrane conductance of the cells.

We implemented an adaptive temporal step size to improve the efficiency of this simple scheme in terms of computation time. After each time step, the step size is refined according to the previously computed derivative value. The new temporal step Δt at the n th iteration is computed as follows:

$$\Delta t_n = \Delta t_{min} \frac{\Delta V_{max}}{\max(\Delta(V_m)_{n-1}) \cdot Ft}$$

where Δt_{min} (0.05ms) is the minimum step size accepted, ΔV_{max} (1V) is the estimate of the maximum of change in V_m and Ft is an adjustment constant. The minimum step has been determined from previous study, to assure the model stability.

2) PARALLEL: We used the multiprocessing module of Python standard library for the parallel implementation. It allows several processes to use a shared memory.

To deal with Numpy arrays (more convenient than the standard type of array) using shared memory, we used the module `shmarray`, written by David Baddeley (under BSD license).

At the beginning of the stimulation, a master process does the initializations and divides the grid of muscle cells and then generates as many processes as required (the default value being the total number of cores available) and then waits until all the processes are done.

The spatial domain is divided according to the number of working processes so that all the processes receive a sub-domain of roughly the same size (Minkoff 2002). Empirically, we found that the 1D-division presented better results than 2D- or 3D-parcellization. For that reason, only the rows of the state array are fairly divided for all the processes used since this method gives better results than dividing both the rows and the columns of the grid. Each process is dependent of its neighbors because it needs border information for the spatial diffusion process.

Each process has two rows of the cell grid in common with its neighbors so that the spatial diffusion can be computed accurately along the whole grid. In other words, a process can only modify the part of the grid that has been affected to him and can only “see”

this domain as well as the two rows above and below. All the models of each sub-domain are computed asynchronously (both the cell and the spatial diffusion models). However all processes need to stay synchronized to process the same time step. This synchronization has been implemented as a barrier for all the processes, each of them storing its own results in the shared memory. A barrier basically blocks all participating threads until the slowest participating thread reaches the barrier call.

3.3. Tools

We also implemented several tools which could be useful for further studies as methods of the generic integration class.

1) Show(): A first need to arise was the ability to display the results of the simulation. To do so the generic integration class provides the `show()` method. Concerning 0D and 1D models, we used the `matplotlib` library (<http://matplotlib.sourceforge.net/>) to plot matlab-like graphics. Figure 2 shows the example of a 1D simulation, each column expresses the state of the whole model at a given time (100 cells computed here) and respectively each lines is the time course of one cell.

For higher dimensions model, we use the 3D-visualization library `mayavi` (<http://code.enthought.com/projects/mayavi/>). It allows us to show 2D and 3D animated simulations. Figures 3 and 4 show examples of simulation results displayed `mayavi`.

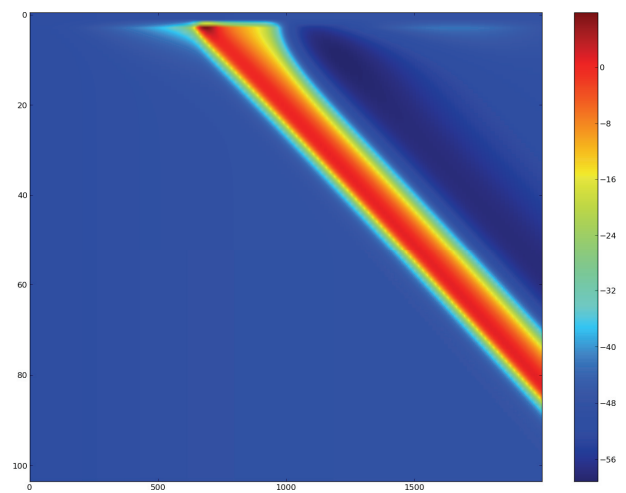


Figure 2: Representation of a 1D simulation using `matplotlib` library. The x-axis represent the time, and the y-axis represents the cells. In this simulation, the cell number 3 the pacemaker.

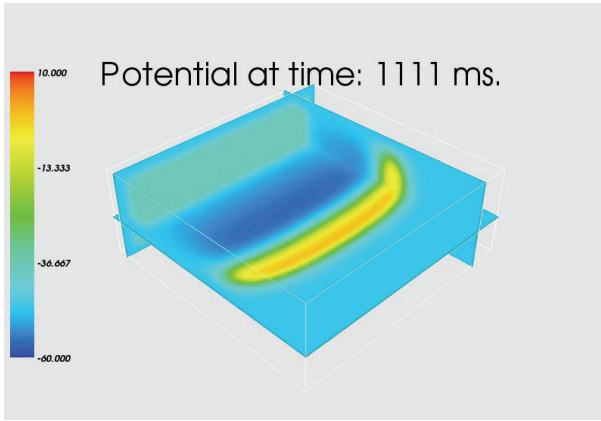


Figure 3: Representation of a 2.5D simulation using mayavi library. Electrical potential amplitude is color-coded.

2) Save(): To store the simulation results, we implemented the method save() using the file management abilities of the Numpy library. It saves a time array describing the sampling and the N-dimensional array describing the electrical potential for each saved time step. It is possible to part the results into several lighter files if the weight of the data to store is too important. A text file containing a description of the model used is saved as well.

3) Speed(): Finally, we implemented an easy way to measure the conduction velocity between two cells on the tissue model. To do so, we simply compute the delay between the time course of the two chosen cells by using the maximum of the cross-correlation. We then calculate the Euclidean distance to get the value of the conduction velocity. The existence of an actual propagation is assessed by a threshold applied on the value of the maximum of the cross-correlation.

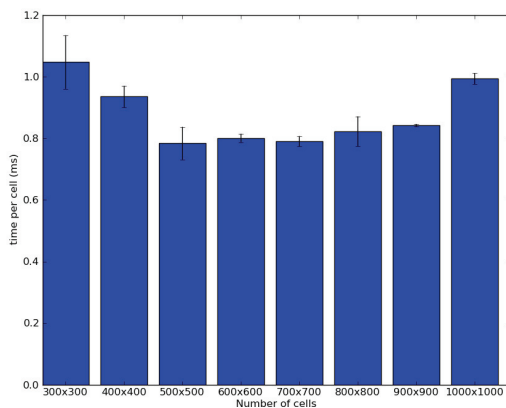


Figure 4: Time per cell (execution time divided by the number of cells) for different number of cells and 12 cores.

4. RESULTS

To test the efficiency of the parallelization of the model, the model was first run on a 8 core MacBookPro (Intel Core i7 2.2Ghz, 8GB RAM, OSX 10.6 64bits). Once

our code became reliable enough for deployment, the simulations were run on a 24-core calculator (Intel Xeon X7542 2.67GHz, 256GB RAM, RedHat Enterprise 64bits). We first present the computation time per cell for simulations of a 2D models (going from 300×300 to 1000×1000 cells) with a given number of cores, to test the grid size effect. Then, we present the effect of parallel computing on a 500×500 -cell model (250000 cells, referred later as 2D model) and on a $150 \times 150 \times 20$ -cell model (450000 cells, referred later as 2.5D model) by using 1 to 24 cores.

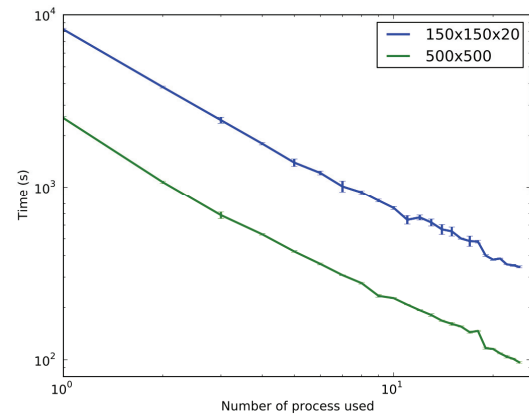


Figure 5: Execution time for 500x500 and 150x150x20 models. Curves are in Log-Log scales to simplify their reading.

Figure 5 shows the time per cell for a 1-second simulation for different grid dimensions and by using 12 processors. To assess repeatability, each simulation was run 5 times. For small dimensions, the parallelization is less efficient. Symmetrically, when the dimension of the model increases, the time of execution per cell increases as well. For both cases, it can be explained by the fact that the cost of the access to the shared memory stops being negligible compared to the gain of time due to the parallelization. For higher dimensions, one can also think that the memory hierarchy effects are less visible for these cases. The figure shows that the dimension of the grid does affect the performance of the parallel computing. Indeed, for too small or too big sizes of the grid, the benefit of the parallel computation is limited by issues of accessing the data.

Figures 6 and 7 show the mean of computation times over 5 simulations respectively for the 2D and 2.5D models, by using 1 to 24 cores. As expected, the 2.5D model takes more time than the 2D one. The time ratio between the times per cell of the two models is 1.88 (standard deviation: 0.07) which means that even with the same number of cells, a 2.5D model would be more time-consuming than a 2D model. The main reason of the time gap is attributed to the spatial diffusion model, where spatial diffusion is applied in 3 directions for the 2.5D model instead of 2 for the 2D model.

In both cases we notice a dramatic reduction of the computation time, from several hours to a few minutes.

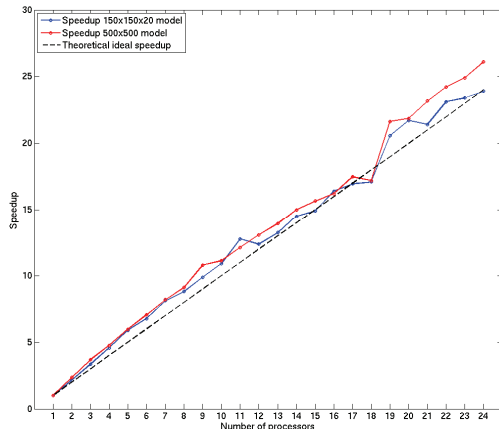


Figure 6: Speedup values for both configurations, for 1 to 24 cores.

To analyze more precisely the gain in the computation time given by the parallel computing, we calculated the speedup, which is defined as:

$$speedup(n) = \frac{T(1)}{T(n)} \quad (4)$$

where $T(n)$ is the execution time on the parallel system with n processors, and $T(1)$ is the execution time with only one process. In other words, speedup is defined by how much faster a simulation will run on n processors compared to a serial execution. In theory, the ideal speedup is linear, so for n processors it would be n . However in practical cases, supra-linear speedup has been noticed. Indeed, in some cases the computation time is reduced by a factor higher than the number of processors used for the parallelization (Camargos 2009). the memory hierarchy effects can explained this result, and mostly the use of memory cache that reduces the access time to the data. Figure 7 presents the results of the two models. This effect can be noticed in our data, since speedup values for both models follow a slightly supra-linear curve (results over the ideal curve).

5. CONCLUSIONS

Modeling the electric activity of the uterine muscle by realistic simulation is a challenging issue that raises the problem of computational time. We developed a complete tool to be able to parallelize time-consuming simulation. Our model is able to simulate uterine tissue model from 0D to 2.5D.

By using shared-memory parallel computing, we were able to reduce the execution time up to 25 times. In other terms, we could turn a simulation that lasted more than 2 hours into a 5-minute run. Indeed, our

speedup curves follow a slightly supra-linear progression. However, for large dimensions model, the efficiency of the parallelization decreases. This will permit us to increase the dimension of the model, thus inducing an increase in the simulation duration, but keeping this increase reasonable enough. It will allow realistic simulations that could be validated by physiological experiments.

Our model still presents some limits. For instance, it is only able to simulate tissue areas of simple shapes (rect- angular, cylindrical). We will need better meshing abilities to be able to represent the complex uterine tissue structure. Another limit will be memory usage as the size of the simulated model increases. A work-around will be to support distributed memory calculators. We are developing an integrator class relying on message based parallelization (using MPI library) in this aim. Once deployed on a similar scale calculator we'll be able to confront the two approaches.

Finally, a user-friendly graphical interface will be added to the software in order to make its exploitation possible by non-specialists.

ACKNOWLEDGMENTS

This work was funded by ANR, partner of the ERASys-Bio+ initiative supported under the EU ERA-NET Plus scheme in FP7.

REFERENCES

- Camargos, A., Batalha Martins Silva Soares, R. C. E. G., 2009, Super-linear speedup in a 3-d parallel conjugate gradient solver, *IEEE Transactions on Magnetics*, 45:1602–1605
- Devedeux, D., Marque Mansour Germain Duchene, C. S. G. J., 1993, Uterine electromyography: a critical review, *American Journal of Obstetrics and Gynecology*, 169:1636–1653
- Laforet, J., Rabotti Terrien Mischi Marque, C. J. M. C., 2011, Toward a multiscale model of the uterine electrical activity, *IEEE Transactions on Bio-Medical Engineering*, 58:3487–3490
- Minkoff, S.E. 2002, Spatial parallelism of a 3D finite difference Velocity-Stress elastic wave propagation code, *SIAM Journal on Scientific Computing*, 24:1 [5]
- Rabotti C., Mischi Beulen Oei Bergmans, M. L. S. J., 2010, Modeling and identification of the electrohysterographic volume conductor by high-density electrodes, *IEEE Transactions on Biomedical Engineering*, 57:519–527
- Rihana, S., Terrien Germain Marque, J. G. C., 2009, Mathematical modeling of electrical activity of uterine muscle cells, *Medical & Biological Engineering & Computing*, 47:665–675

AUTHORS BIOGRAPHY

Tanguy Hedrich was born in Poissy (France) in 1989. In 2011 he received both the Engineering degree in computer engineering and the Master degree in biomedical engineering from Compiègne University. He is now a Master student in biomedical engineering at McGill university, Canada. His research interests include statistical signal processing and localization of epileptic spikes using distributed sources modeling.

Jeremy Laforet was born in France on February 19, 1983. He received both the Master and Ph.D. Degrees from Montpellier 2 University, Montpellier, France, in 2006 and 2009. He's now post-doctoral researcher at Compiègne University in the biological engineering department. His interests include neuro-muscular models and their identification and validation.

Catherine Marque was born in France on January 15, 1958. She received the degree of Engineer from the Ecole Nationale Supérieure des Arts et Métiers, Paris, France, in 1980 and the Ph.D. degree from Compiègne University, Compiègne, France, in 1987. She is presently Professor at Compiègne University in the biological engineering department where she manages the biomedical engineering formation. Her interests include signal processing and instrumentation applied to the biomedical field.

INNER SIMULATION SUSTAINING THE DELIBERATIVE PROCESS IN A COGNITIVE ARCHITECTURE

Othalia Larue^(a), Pierre Poirier^(b), Roger Nkambou^(a)

^(a) Computer Science department, GDAC Research Laboratory

^(b) Philosophy Department

Université du Québec à Montréal

P.O. Box 8888 Centre Ville Station, Montréal, Québec, Canada

^(a) larue.othalia@courrier.uqam.ca, ^(b) poirier.pierre@uqam.ca, ^(c) nkambou.roger@uqam.ca

ABSTRACT

We propose a three level cognitive architecture for the simulation of cognitive phenomena. This architecture is based on Stanovich's tripartite framework (2010), a unified model of cognition, which provides an explanation of how deliberative (characterized by sequentiality) and adaptive (characterized by reactivity) human behaviour emerges from the interaction of three distinct cognitive levels (autonomous/reactive, algorithmic/cognitive control, and reflective). In previous work (Larue et al, 2012) we focused on the interaction of algorithmic and reactive level on a task evaluating cognitive control. In this paper, we focus on the interaction between reflective and algorithmic level. More precisely, thanks to a Wisconsin card Sorting task, a task that evaluates cognitive flexibility (the ability to change strategies), we study how cognitive decoupling (or inner simulation) supports the deliberative behaviour and hypothesis testing.

Keywords: Wisconsin Card sorting task, cognitive simulations, cognitive architecture, dual process theories

1. INTRODUCTION

Many cognitive architectures are designed with the intent of reproducing the human cognitive architecture, the human mind, either by replicating known fine-grained structures of the brain (Eliasmith, 2005) or by building systems with capacities that are functionally equivalent to those of humans. Designers of such architecture all face what might be called "the duality challenge." Evidence from many fields in cognitive psychology (psychology of reasoning, moral psychology, social psychology, etc.; see Evans 2008 for a review) and from cognitive neuroscience (see e.g., Goel, 2009) suggests that evolution may have built functionally incompatible (Sherry & Schacter, 1987) features into the human mind. On the one hand the mind is dynamical and reactive, simultaneously responding to many features in the environment in a seamless

dynamical agent environment loop, while, on the other hand, it is sequential and rule following, applying explicitly learned rules one by one to plan its long-term behaviour or solve other complex problems.

One important source of evidence that helped create and sustain the duality challenge (at least in its contemporary version) is to be found in the various tests used by cognitive psychologists and neuropsychologists to assess cognitive performance: the Stroop task (attention), the Wisconsin card sorting test (WCST) (cognitive flexibility), the Wason selection task (logical reasoning), the Iowa Gambling task (decision making), and others. One of the early proposals was Evans' (1984) attempt to explain the bias observed in the Wason selection task (a task frequently used in the psychology of reasoning) by positing a competition between opposed heuristic (fast and automatic) and analytic (slow) processes. Explanations of the Stroop effect (Stroop, 1935) likewise often posit competing automatic processes, where attention, an opposed voluntary process, has to favor the less automatic of the two competing automatic processes. Similarly, many explanations of the perseverative errors in the Wisconsin Card Sorting Test rest on interactions between automatic and attentive processes. To address the duality challenge, the current trend in cognitive science and neuroscience is to posit dual system (Kahneman, 2011) or dual-process (Evans, 2008; Stanovich, 2010; Stanovich, 2011) theories, a "two-minds mind" (Frankish & Evans 2009), often called "System 1" and "System 2," where processes with features taken from a list of duals (implicit vs. explicit, automatic vs. voluntary, etc.) compete or collaborate to explain observed human behaviour. Designers of cognitive architectures in general have met the duality challenge by building hybrid architectures, interfacing dynamical and parallel systems such as neural networks with sequential and rule following systems, such as production systems. Dual-process theories (Evans, 2006; Sloman, 1996; Kahneman, 2011), however, have attracted much criticism in the literature (e.g., Keren &

Schul, 2009) where they are (rightly) said to be oversimplifications. One pressing problem is to account for the interaction between the opposing processes: how, for instance, can a voluntary process interrupt an automatic one; or again how can the output of a domain-specific process affect domain-general decision making. Call this the “interface problem” (of current-trend solutions to the duality challenge). Answers to the interface problem range from mere hand-waving (simply declaring that processes (somehow) interact) to nihilism (simply declaring that one set of processes do not exist – e.g., the mind is all (massively) modular) to despair (Fodor, 2000). Designers of cognitive architectures in general have dealt with the interface problem by converting the cognitive dichotomy into a paradigm (connectionist-symbolic) dichotomy (see for instance CLARION (Connectionist Learning with Adaptive Rule Induction ON line, Sun, 2004).

A few philosophers (e.g., Carruthers, 2006), computer scientists (e.g. Sloman & Chrisley, 2005) and psychologists, however, have addressed the interface problem directly, attempting to provide positive accounts of the means of interaction between mind’s two minds. One such attempt is Stanovich’s Tripartite Framework, which, paradoxically, begins by positing a three-minds mind: an Autonomous Mind, an Algorithmic Mind and a Reflective Mind. We use this model as the basis for the design of our architecture.

Our aim is to motivate the resulting architecture as a simulation tool. In previous work, we illustrated and validated the performance of the system on the two lower levels (autonomous and algorithmic minds), thereby demonstrating the adaptive behaviour of our system, by means of two variants of the Stroop task (classical and semantic) (Larue et al., 2012). In this paper, we focus on the interaction between the reflective mind and the algorithmic mind of the system. To do so, we added to our architecture a function that has been deemed, in Stanovich’s tripartite framework, a key aspect in the production of deliberative behaviour: cognitive decoupling. We use a Wisconsin card sorting, a task which requires cognitive decoupling and cognitive flexibility to be performed, as a way to illustrate the system’s performance.

2. RELATED WORK

2.1. Stanovich’s tripartite model

2.1.1. Autonomous, algorithmic and reflective minds

Our cognitive architecture is based on Stanovich’s tripartite framework (Stanovich, 2010). Letters in this section refers to letters on Figure 1. Stanovich’s tripartite model is a unified model of cognition that gives an account of how automatic (implicit) processes and explicit processes (control - attention and executive functions and more abstract planning and reasoning functions) are able to coexist. Stanovich’s tripartite framework belongs to the “dual-process theories” we

previously described System 1 (called “Autonomous Mind” in Stanovich’s tripartite framework), is the locus of fast and automatic reasoning where instinctive behaviour, over-learned process, domain-specific knowledge, emotional regulation and implicit learning are found. System 2 is the locus of abstract and hypothetical reasoning. The division of human cognition into three sets of processes, instead of the traditional two of dual-process theories, provides a better account of individual cognitive differences. System 2 in Stanovich’s tripartite framework is divided in two classes of processes, respectively called the “Algorithmic Mind,” responsible for cognitive control, and the “Reflective Mind,” responsible for deliberative processes. Owing to this subdivision, the framework can capture the distinction between on the one hand cognitive ability and fluid intelligence achieved by the algorithmic mind, and on the other hand thinking dispositions and critical thinking skills achieved by the Reflective mind.

The Algorithmic mind sustains three distinct processes: (1) override of Autonomous Mind processes (A), (2) cognitive decoupling (see section 2.1.2 for a complete description) (C), and (3) serial associative cognition.

Algorithmic Mind is linked to cognitive functions such as cognitive control and working memory and has access to information from the Autonomous Mind via two sets of pre-attentive processes (G) (1) perceptual processes and (2) beliefs and memory retrieval processes. Operations supported by the Reflective Mind define the subject’s cognitive style. The Reflective Mind initiates: the override of Autonomous Mind (B) processes by the Algorithmic Mind and the cognitive decoupling operation (D) (the cognitive decoupling capacity is sustained by the Algorithmic Mind). The Reflective mind can also act upon the Algorithmic mind by interrupting (F) serial cognition to send a new action plan for execution or start cognitive decoupling.

2.1.2. Cognitive decoupling

Cognitive decoupling is a key mechanism that supports human rationality. Individual differences in the operation of this mechanism lead to differences in rational thinking (Stanovich 2010).

Decoupling has been largely studied in the dual process literature. It has been referred to as decoupling, cognitive simulation (Stanovich 2010) or hypothetical thinking (Evans 2008). It consists in the creation of temporary models (D) of the world upon which alternative scenarios can be experimented. Nichols and Stich (2000) dubs it “possible words box”, a separate box in which simulation are carried out. The particularity of these temporary models is to be independent of the current mental representation of the world (primary representation). This prevents the real world representation to be confused with imaginary situations (secondary representation), since their manipulation doesn’t affect the current representation of the world.

While the Algorithmic mind carries out cognitive decoupling, it has a hard time performing other processes. Decoupling is a cognitively expensive operation, (its cognitive load is higher than that of serial associative cognition) therefore it is not systematically performed, and when performed, it can be carried out incompletely. As a result, suboptimal responses that are cognitively easier solution provided by serial association (i.e. simple and incomplete models that appear appropriate for the situation) are often applied. Serial associative cognition (E) supports the implementation of these simple models.

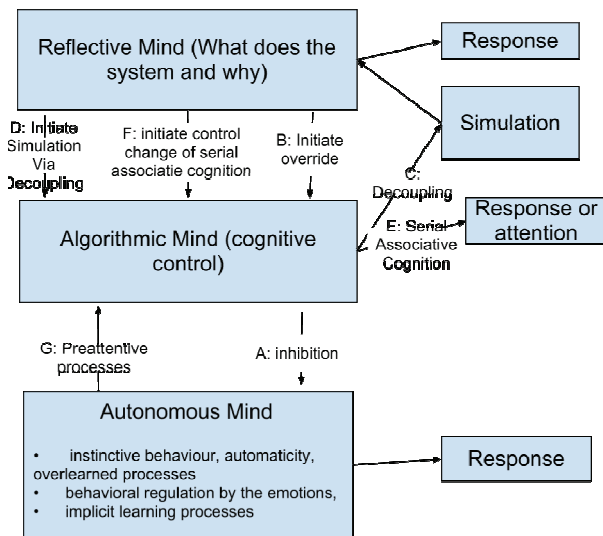


Figure: Stanovich's tripartite framework

2.2. Wisconsin Card Sorting Task

The Wisconsin card sorting test (WCST) (Grant & Berg, 1948) is a task widely used to test executive functioning, especially cognitive flexibility and abstract reasoning. The subject is shown a set of target cards. The figures on the cards vary in shape, number and color. Stimulus cards are shown to the subject, one by one, and the subject is asked to match them to one of the target cards. The subject is not told what the sorting rule is; he has to discover it. However, he receives negative feedback when his matching is wrong. After a number of successful matching, the rule changes without the subject being warned. After discovering the new rule, subjects experience difficulties inhibiting the previous rule: they cannot shift attention from one complex visual stimulus to another (set shifting). The perseverative error is the tendency to use the previous rule after a switch (Nyhus & Barcelo, 2009). Perseverative errors are more common in patients with lesions to the frontal lobe and increase with the age of subjects. The selection of the response is achieved through two mechanisms: the attentional set shift and the reversal shift, which are located in two distinct neuroanatomical structures of the PFC (Nagahama et al., 2005). The reversal shift, achieved through the posterior region of the PFC, is in charge of the update

of associations between stimulus and response modules. The attentional set shift, achieved through the rostradorsal PFC, is a higher level cognitive mechanism allowing adoption of the new rule by cognitive control (Anterior Cingular Cortex).

2.3. Computer simulations of the WCST

Computer simulations of this task have been achieved, mainly with connectionist approaches (Dehaene & Changeux, 1991; Kaplan et al., 2006), but also with symbolic and hybrid approaches (Kimberg & Farrah, 1993).

Dehaene and Changeux's simulation focuses on the functional aspect of the task by including in their neural network model the three cognitive components they deem critical to the accomplishment of the task: the ability to change rules when negative reward occurs, the capacity to memorize previously tested rules and the possibility of rejecting rules because of a priori reasoning. These three components are achieved by means of a hierarchical structure: a sensorimotor loop, a higher level assembly of rule-coding cluster codes – the current rule – shifted in case of negative reward (episodic memory of the system), an endogenous auto-evaluation loop allowing the internal testing of rule. The structure is compatible with the organization and specialization of cortical areas.

To simulate the distinction between the hypothesis generator and action, Kaplan et al. (2006) connected two distinct subsystems (respectively a hamming block and a Hopfield network), which allowed them to study perseverance and distractibility and reproduce lesions in the prefrontal cortex. The Hopfield network acts as the system's Working Memory, and the hamming block as its hypothesis generator.

Kimberg and Farrah (1993) explicitly represent the sorting behaviour of their system, to study the involvement of Working memory in the task, in an ACT-R architecture (Anderson & Lebiere, 1998). ACT-R is a production system and as such, uses a set of productions (its procedural knowledge and a working memory representation). The Working memory associations are weakened in order to model the effects of Frontal Lobe Damage in humans.

Being an hybrid approach, our architecture presents some similarities with these various architectures, if only because they all use a model of the human brain in the performance of the task. The endogenous auto-evaluation loop of Dehaene and Changeux's architecture, allowing the internal testing of rule, can be compared functionally to our decoupling ability. Their hierarchical structure can be compared to that one of our own architecture; however in our system, behaviour emerges through the interaction of the different levels, rather than being strictly controlled by a higher level structure. We share aspects with the different connectionist approaches since memory of the previous correct and incorrect rules is kept by the degree of activation of knowledge (at the reactive level) and objectives (at the reflective level). We share aspects

with the symbolic approaches since knowledge and objectives in our system are represented by symbols. All of the previous simulations study the involvement of one specific cognitive component in the performance of the task. In this paper, however, we chose the WCST as a way of studying the interaction between the components (or minds) in Stanovich's Tripartite Framework, this task being known to tap reflective and algorithmic processes (inhibition and decoupling at the rule discovery stage).

3. OUR PROPOSAL

3.1. Architecture

We implemented our cognitive architecture in the multi-agent system (MAS) platform Madkit. The MAS is organized into groups (In Madkit, a "group" is a set of agents that share common characteristics), each corresponding to a cognitive level in our architecture. As there are three levels in our cognitive architecture (reactive, algorithmic and reflective), the MAS is composed of three groups. To keep matters simple, groups are given the name of the level they correspond to. Each agent has one or more roles (in Madkit, a "role" is an abstract representation of an agent's functionality) and belongs to one or more groups. All agents work in parallel, sending messages to each other. Each message received by an agent will activate the agent a little; the activation of the agent is thus a function of the number of messages it receives. The activation of an agent determines the number of messages it sends by unit of time.

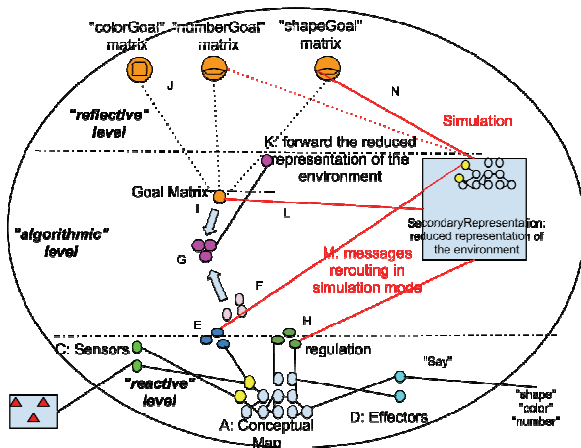


Figure 2 : System's architecture

3.1.1. Reactive level

Corresponding to Stanovich's Autonomous Mind, the Reactive level contains agents assigned with one of three roles: the "sensor" role (C - Letters in the section refer to letters in Figure 2), the "effector" role (D) and "knowledge" role (A), which we thus respectively name: Sensor Agents, Effector Agents and Knowledge

Agents. Knowledge Agents form a network isomorphic to the conceptual map the system is initialized with. Each Knowledge Agent receives, as an additional role, a word from the conceptual map (e.g., "Red") and is linked to other Knowledge Agents according to links between the words in the map (e.g., the Knowledge Agent "Red" will be linked to the Knowledge Agent "Tomato", and so on). The links between Knowledge Agents are weighted according to the semantic distance between them in the conceptual map (e.g. the link between "Red" and "Tomato" will receive a higher weight than that between "Red" and "Meat"). The network of knowledge Agents makes up the system's declarative knowledge (semantic memory). In the experiments reported in this paper, the network of Knowledge Agents MAS was initialized with the common sense knowledge database Conceptnet (Havasi, Speer & Alonso, 2007). The number of messages exchanged between the agents, and therefore their activation, is at first determined by the distance between them in the conceptual map (it will also be determined by activation signals from higher levels – see below). Activation thus spreads through the network of Knowledge Agents (a process similar to semantic memories; Anderson, 1983).

Sensor Agents are sensitive to conditions in the environment (colors, words, numbers, shapes, etc.) and send messages to selected Knowledge Agents. For instance, a Sensor Agent that can detect colors (here: the recognizeColor agent) will be linked to color Knowledge Agents; it will send a message to the "Red" Knowledge Agent if it detects red in the environment. The system's environment is similar to (portions of) human environments. In the WCST simulation described below, the system is presented with cards identical to those human subjects see in real WCST experiments. Effectors Agents act on the environment. They are linked to Knowledge Agents whose role is relevant to the action they can perform. For instance, the NameColor Effector Agent will be linked to every color Knowledge Agent (thus implementing the capacity to name the colors we see). Taken together, the Sensor and Knowledge Agents make up the system's perceptual system. This means that the system's perceptual abilities are always a function of the Sensor Agents' information extracting capacities and of the system's knowledge about the environment, as implemented in the network of Knowledge Agents. Similarly, the Effector and Knowledge Agents form the system's action system. Both perception and action are thus fully situated and contextual. The system's long-term memory is made up of the network of Knowledge Agents in the Reactive Group, and the system's working memory (WM) at a given time is made up of the Knowledge Agents that are activated at that time. This implementation of working memory is consistent with the work of Engle (2010), in which WM is seen as a set of temporarily activated representations in long-term memory.

3.1.2. Algorithmic level

Corresponding to Stanovich's Algorithmic Mind, the Algorithmic Group is responsible for the control of the system. Based on a general idea developed by Cardon (2005), control is achieved by means of assessing and affecting the "morphology" of the system. RequestStatus Agents (E), which belong to both the Reactive and Algorithmic Groups, regularly query the status of Knowledge Agents; that is, number of messages sent to other agents during the last interval. Status Agents (F) represent the activity of the Reflective Group at a given time in the form of a *distance matrix* that records the message passing activity at that time (Status Agents will also send a reduced representation of the activity in the Reactive Group to the Reflective Group; see below). The distance between two concepts in the conceptual map is measured by the number of messages sent between the Knowledge Agents that bear the corresponding words as their role. Globally, this matrix thus represents a form or shape. The Algorithmic Level also contains the short-term goals of the system in the form of a graph of Goal Agents (which is sent by the Reflective level; see below). Each Goal Agent (I) contains a distance matrix that specifies the distance between each Knowledge Agents that is necessary if the system is to reach this goal.

Delta Agents (G) compute the difference between the matrix provided by the Status Agents and that provided by the Goal Agents. The resulting difference (another matrix) is provided to Control Agents (H) that send regulation messages to Agents in the Reactive Group, telling them to modify (i.e., increase or decrease) their activation so that their global activity more closely matches the shape of the current short-term goal. By activating elements of the system's long-term memory in relation to its current goal, thereby determining the current content of working memory, agents in the Algorithmic Group constitute the system's attentional system.

3.1.3. Reflective level

Corresponding to Stanovich's Reflective Mind, the Reflective Group is responsible for the logical and analytical skills of the system. Each agent in the Reflective Group has a shape (a distance matrix) as its role, which, as explained above, indicates the shape the Reflective Level must be in for the system to achieve a simple goal. Goal Agents (I) are organized in a directed graph. Every path in this graph represents a plan the system can apply to achieve a complex behaviour. Goal Agents are organized into a graph where each path represents a complex plan or strategy decomposed into a sequence of simple objectives (steps in the plan). A path (a sequence of simple goal path – in the WCST the path consist in a single objective but it could be longer for other tasks) (J) will be sent to Goal Agents of the Algorithmic Group, which will take care of its

execution. Following Stanovich's Tripartite Framework, agents in the Reflective Group have access to a reduced representation of the environment, which, as explained, is provided as a matrix by the Status Agents of the Algorithmic Group (K). The similarity between these two matrices, computed by the Goal Agents, determines the activation of the Goal Agents, which propagates from the Agent most matching the reduced representation to those that follow in its path. The last agent in the path sends the parsed path to the Algorithmic Group. The shortest path (the simplest model) or the one the most activated (the model used more recently or more often) thus prevails over the other paths. Goal Agents of the Algorithmic Group will execute this path step by step (this corresponds to Stanovich's serial associative cognition).

The path executed by serial associative cognition provides the system with the sequentiality necessary to achieve complex goals. However, the system does not thereby lose its dynamicity. Reduced Representation of the environment are sent on a regular basis to the Reflective Group that can, based on the current state of the environment, interrupt serial cognitive association either by setting a new starting point in the path or by taking a new branch in the graph. Decision-making at the Reflective level is therefore dynamically influenced by the current strategy (the decided path) of the system and the state of the environment.

3.1.4. Decoupling

Cognitive decoupling is an operation that is initiated by Agents of the Reflective Group and achieved by agents of the Algorithmic Group when multiple strategies (meaning two or more GoalSet Agents) are selected at the algorithmic level. The Goal Agent, which usually carries the unique goal matrix selected at the Reflective level triggers cognitive decoupling.

When the Goal agent hesitates between two strategies (when the activation levels of two GoalSet agents are close – closeness being defined by a sensitivity degree which we will further explain in section 4.4.1), it sends (L) a message to agents of the reactive group informing them that the system is now in simulation mode. It also triggers the creation of a possible world. A limited number of agents (currently 20) mirrors the activity of agents at the reactive level according to the reduced representation previously sent by Delta agents. Agents in this possible world are assigned dynamically the same roles and links as those agents from the Reactive Group they are replicating. This possible world corresponds to the separate secondary representation we presented earlier in section 2.1.2. Accordingly, SecondaryRepresentation agents are used instead of Knowledge agents and a distinct group (Algorithmic instead of Reactive) is used, to ensure that action on this secondary representation doesn't impact the current representation of the world (i.e., Knowledge Agents from the Reactive Group).

When agents of the algorithmic group are informed that they are in simulation mode, they reroute (M) their

messages to the SecondaryRepresentation Agents of the Algorithmic Group instead of the Knowledge Agents of the Reactive Group. Once the simulation is completed, the activation of Goal Agents is regulated accordingly at the Reflective level (N), therefore dictating the future course of action.

Cognitive operations (goal inhibition and selection) are carried out by the Control Agents, Delta Agents, and agents from the Reflective Group during cognitive decoupling, making it considerably difficult to sustain other activities in the system at the reactive level (cognitive cost of decoupling).

For the WCST, the decoupling has been set to perform opposite style of thinking: a chosen categorization rule (the one with the highest activation or a random one if activations are equal) is applied in the possible world and is thereafter negated (negative feedback from the environment), thus a new categorization rule (the alternative) emerges. Goal agent therefore sends activation messages to the GoalSet agent bearing the first activated rule, and half less activation messages to the one bearing the alternative rule: if the first rule is negated in the “real” world, the alternative rule will therefore be the second one to be activated.

3.2. Neurological plausibility

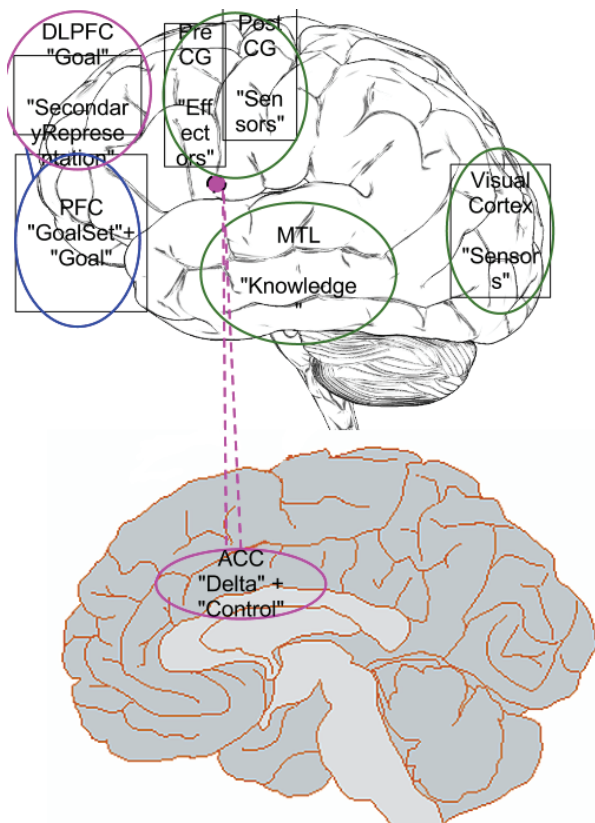


Figure 3: Functional mapping. □

We cannot claim fine-grained neurological plausibility for this system; however, parallels can be drawn at the higher level of gross neurological structure and mesoscopic dynamical activity, allowing us to claim that a measure of neurological plausibility for the architecture. The neurological plausibility of dual process theories has been extensively studied (Goel, 2009; Lieberman, 2009). Stanovich’s tripartite framework, individually, is also supported by neurological data (Stanovich, 2010). Since the design of our architecture is based on this model, it naturally inherits its neural plausibility. Cognitive tasks supported by the Algorithmic Mind lead to an activation of the Anterior Cingulate cortex (ACC). Performance of Algorithmic Mind processes leads to an activation of the ACC (Stanovich, 2010). There is furthermore evidence that decoupling (supported in our system by SecondaryRepresentation Agents) is achieved by the Dorsolateral PreFrontal Cortex (DLPFC) (Stanovich, 2010).

Furthermore, the different roles ascribed to the agents in the architecture correspond to functional roles that has been mapped to specific anatomical structures – see Figure 3 (Fuster, 2008; Botvinick et al., 2001). It must be noted that the ACC has been identified as the response conflicting monitoring system (Botvinick et al. 2001) in the human brain, regulating control’s engagement. Conflict monitoring is achieved in our system by the collaboration between the Delta Agents and the Control Agents. The DLPFC, which provides support for goal directed behaviour is implemented in our system by Goal Agents. Regulation of posterior brain regions is implemented by the regulating messages sent by the Control Agents to agents of the Reactive Group. Knowledge Agents are linked to Sensor Agents as the Medial Temporal Lobe (MTL in figure 3) is known to mediate sensory memory. The medial temporal lobe is also identified as the functional locus of semantic memory. “Effector” and “Sensor” agents are associated with distinct roles in the Reactive Group since their functional role is achieved by distinct anatomical structures (PreCentral Gyrus, PostCentral Gyrus and Visual Cortex in figure 3). Through nested sensorimotor and goal-directing loops, we are therefore able to implement the cognitive dynamics of a goal-sensitive sensory-motor architecture.

Finally, we intend future developments of the architecture to further increase its neurological plausibility. The majority of the existing circuits in the system (its message passing channels) mirrors neurochemical pathways of the human brain (for example, dopamine pathway will in our system be represented by the message passing activity between GoalSet Agents and Goal Agents in the Reflective Group, and between Knowledge Agents and Effector Agents, to respectively represent the reward effect of dopamine, and its impact on motor function (Fuster, 2008)). This property will allow us to further explore the dynamics of cognitive processing in the system in

future works by varying the message passing activity in distinct pathways (neuromodulation).

4. EXPERIMENT AND RESULTS

4.1. Implementation

We chose a “generic” multi-agent Platform (Madkit) that allows the creation of agents with different ranges of complexity and in which large numbers of agents can operate in parallel. Madkit implements the AGR (Agent/group/role) model, which we found particularly suitable to implement the various groups of agents (levels), and the diversity of agents in each group, as described previously. In AGR (Ferber et al., 2003):

- An agent is an active entity communicating and playing a role within one or several groups. There are no design constraints regarding the complexity of the agents (reactive/cognitive).
- A group is a set of agents sharing common characteristics. Groups define the organizational structure of the system.
- A role is an abstract representation of agent’s functionality within a group. We chose to assign a thread to each agent in order to preserve its dynamicity.

4.2. Experimental setting

The system was given as parameters three goals matrices corresponding to the color, shape and number categorization. To implement the context effect of the four reference cards that the subject must select to place each of his response cards (material adapted from Heaton et al., 1993), we added the following links in the conceptual map of the system: Red – triangle – one, Green – star – two, Yellow – square – three, Blue – circle – four. We also linked the shape, color, and number knowledge (already present in ConceptNet) to the sensors. A script provided the system with the series of cards it had to categorize and evaluated the system’s answer. The categorization rule was changed after 6 consecutive successes. The experiment consisted of 128 cards with figures varying in shape, number and color. The script attempted to test 6 categorization rules (“shape”, “color”, “number” x 2) on the 128 cards. No warning was sent to the system before a rule change.

4.3. Description of the interactions

Although our system is not sequential in nature (organizations work in parallel and the global behaviour of the system emerges from the interactions of all agents in the system), we describe below the WCST task processing sequentially to ease its understanding. However, it should be borne in mind, that this sequentiality emerges from the system’s parallel processing.

Table 1: Variation of decoupling’s degree of sensitivity.

Degree of sensitivity	1/6	1/8	1/10
Trials	89.1	120.41	126.6

Table 2: Excerpt of one simulation’s logs

Serie	Trial	Response	Simulation
Color	1	correct	
	2	correct	
	3	correct	color shape
	4	Incorrect : shape	
	5	correct	

Step 1 - When a new card appears in the environment, sensors (for color, shape, and number) extract its relevant properties and forward the information to Knowledge Agents in the Reactive Group where activation spread according the links in the conceptual map.

Step 2 - Status Agents forward this information (activation) to the algorithmic level.

Step 3 - A reduced representation of the environment is produced by Delta Agents and forwarded to the reflective level

Step 4 - The received reduced representation leads to the activation of competing rules (by pattern matching between the goal they are bearing and information from the reduced representation).

(Please note that Step 5 to Step 7 – included – are conditional to the fact that there is more than one winning GoalSet Agent, otherwise, step 8 is directly applied.)

Step 5 - When there is more than one winning GoalSet Agent, meaning more than one winning goal, cognitive decoupling occurs. A mini world (secondary representation made after the – primary - representation of the world carried by the system at the reactive level) is created where agents are initialized after the reduced representation of the world sent at step 4.

Step 6 - Messages from the Control agents (regulation) and from RequestStatus agents (status of the concept agents) are rerouted from the reactive level to the secondary representation that has been initialized at step 5.

Step 7: Rules that emerged in the secondary representation during cognitive decoupling are sent to the reflective level (with different activations).

Step 8: The corresponding matrix (the one that a GoalSet Agent was carrying) is sent to the algorithmic level.

Step 9 - Rerouting is stopped. Agents of the Algorithmic group are back in charge of the regulation the Reactive level. They regulate agent’s activity

according to the activity matrix associated with the system's current goal.

4.4. Results

4.4.1. Cognitive decoupling

In the system, decoupling occurs when two competing answers are identified (when two answers have close activation levels). We ascribed a degree of sensitivity to the Goal agent which starts decoupling operations: the degree of sensitivity determines how close the two competing answers need to be to start cognitive decoupling.

In Table 1, we show the performance of the system in regards to this degree of sensitivity. We can see that the higher the degree of sensitivity is (meaning the more decoupling operation occur), the lower is the number of trials required to perform the task. Please note that, due to the experimental procedure (section 4.2), the maximum number of trials allowed to complete the task can't be higher than 128. For the "1/6" degree, system's performance was better than for human subjects, this results shows interesting premises for the usability of this architecture for the simulation of cognitive tasks with different cognitive profiles (replicating individual differences of subjects with different cognitive styles).

Table 2 presents the log for the simulation of a WCST task with the 1/8 degree. For the "color" serie, we can observe the decoupling being launched in the following situations:

- Trial 3: a simulation/decoupling is launched due to two competing answers ("color" and "shape"). The system creates via decoupling a possible world (separate box) where the color categorization rule is applied and ruled wrong. The second emerging rule is the shape categorization rule.
- Trial 1 and 2: "color" (a wrong answer) is selected. In the possible world created, "shape" activated first. The possible world being a copy of the system's representation of the world, "color" has already been marked as a wrong answer, therefore the selected second answer is number.

4.4.2. Factors analysis

We focused our analysis on 4 factors:

- Failure to maintain set: failure to carry out a complete categorization after a number of consecutive correct trials;
- Perseverative errors: Mean number of trials in which the subject persists in categorizing items with the preceding rule after rule change;
- Categories completed: mean number of completed categories;
- Trials: Mean number of trials to complete 6 categories.

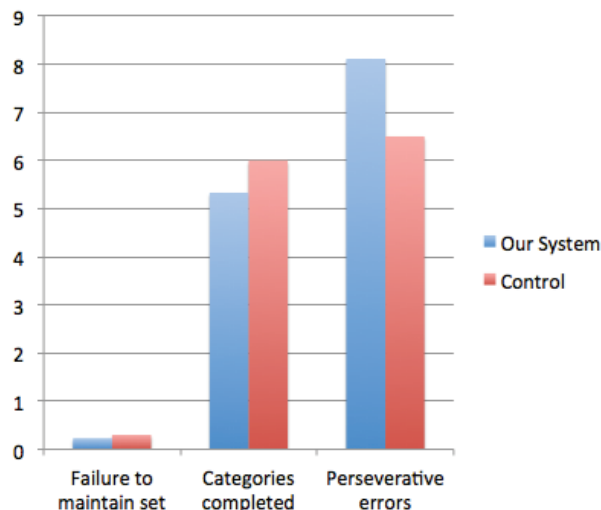


Figure 3: Mean « Failure to maintain set » errors, Perseverative errors and categories completed by our system (mean number for a hundred trials) and by a control group of healthy human subjects taken from Bishara *et al.*'s study (2010).

We compared our results (degree of sensitivity: 1/8) for the three first factors to those of human subjects. Results are presented in figure 3.

The fourth factor, mean number of trials to complete the task in our system was 120.41 against 101.12 in unpracticed healthy human subjects (Basso *et al.*, 1999). The maximum number of rules the system (for his 128 trials) was able to discover and apply varied between 5 and 6 (mean number being 5.33).

During simulations, we were able to observe two different types of wrong answers: those due to a failure at the stage of rule-discovery (reflective level – failure at the hypothesis testing level) and those due to a failure of inhibition (inhibition of the reactive level by the algorithmic level – failure of the working memory in the human subject).

Our system was able to reproduce the pattern of results observed in human subjects (see Figure 3); however, its perseverative tendency was worse than that of human subjects. We believe that this result reflects a failure of inhibition on the part of Control Agents, showing a deficit at the algorithmic level: a failure to inhibit the routine activity that had been established by previous trials at the reactive level. On the other hand, the low value for the « failure to maintain set » factor shows a good flexibility of the system for the selection of new rules, therefore validating the work of the reflective level. To improve the system in future works, we therefore plan to work on calibrating the message passing activity of Control Agents in the Algorithmic Group. This correction might also help increase the number of completed categories.

5. CONCLUSION AND FUTURE WORKS

In this paper, we present a cognitive architecture implementing Stanovich' Tripartite Framework. We already produced classical and semantic Stroop task

simulations to validate the two first levels of the cognitive architecture: algorithmic and reflective mind (Larue et al. 2012). In this paper, we present a validation of its computational soundness as a whole. The WCST simulation we perform on this architecture is a concrete demonstration of how the three levels can interact to produce a complex behaviour involving different levels of cognition in one single structurally unified tool. The WCST allows a study of the interaction between the algorithmic and reflective levels, and more specifically the initiation of simulation/decoupling processes. Decoupling is a key process that sustains deliberative behaviour by allowing hypothesis testing. In this task, automatic cognition is observed in the initial reaction of the system to stimuli, algorithmic cognition allows the system to achieve trial-and-error adaptation and hypothesis testing (e.g. simulation) initialized thanks to the Reflective level from which will emerge the action to be performed by the system. To implement the dual nature (sequentiality/rule-following and reactivity/dynamicity) of human cognition, and thus meet the duality challenge, we combine aspects of classical symbolic (sequentiality/rule-following) and connectionism (reactivity/dynamicity) into a structurally unified cognitive architecture. We were able to combine those two approaches (dynamicity and sequentiality) using one single computational intelligence paradigm (Multi-agent system). Results were encouraging concerning the interaction of reflective and reactive levels of the system, but they we will need in the future to improve the system's inhibition ability at the algorithmic level.. In future work, we plan on addressing this issue by introducing a new variable to the system: neuromodulations. Neuromodulations would modulate communications between agents inside a group and between agents in different groups of the system, therefore modifying the system's general dynamics (allowing us finer-grained control on the dynamics but also, for example, to induce pathological behaviour in the system). Also, this task enabled us to test hypothesis testing abilities which are primordial for a deliberative behaviour (especially selection of the good strategy); however, the strategies among which the system had to chose where very simple strategies. In the future work, we plan on using the system with more complex strategies.

REFERENCES

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, Mass.: Harvard University Press.
- Basso, M.R., Bornstein, R.A., & Lang, J.M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, 13.
- Bishara, A.J., Kruschke, J.K., Stout, J.C., Bechara, A., McCabe, D.P., Bussemeyer, J.R. (2010), *Sequential Learning Models for the Wisconsin Card Sort Task*, *J Math Psychol*.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., et Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Rev*, 108(3), 624-652.
- Carruthers, P. (2006). *The Architecture of Mind*. Oxford: Oxford University Press.
- Cardon, A. (2005). *La complexité organisée*. Paris: Hermès.
- Dehaene, S., & Changeux, J. (1991). WCST: Theoretical analysis and modeling. *Cerebral Cortex*, 1
- Dehaene S. et Changeux, J., «The Wisconsin Card Sorting Test: Theoretical analysis and modelling in a neuronal network », *Cerebral Cortex*, 1991.
- Eliasmith, C. (2005). Cognition with neurons: A large-scale, biologically realistic model of the Wason task. In L. B. Bara G.; Barsalou, M. (Ed.), *Cognitive Science 2005*.
- Engle, R. W. (2010). Role of Working Memory capacity in Cognitive Control. *Current Anthropology*, 51.
- Evans, J. (1984), Heuristic and analytic processes in reasoning. *British J of Psych*, 75.
- Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*. 13(3), 378-395.
- Evans, J. (2008). Dual-processing accounts of reasoning, judgement and social cognition. *An Rev Psy*, 59.
- Ferber, J., Gutknecht, O., et Michel, F. (2003). *From Agents to Organizations : an Organizational View of MultiAgent Systems*. Paper presented at the Agent-Oriented Software Engineering (AOSE) IV, Melbourne.
- Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. MIT Press.
- Frankish, K., et Evans, J. S. (2009). The duality of mind. In Frankish & Evans (Eds.), *In Two Minds*.
- Fuster, Joaquin M. (2008). *The Prefrontal Cortex*, Fourth Edition. Boston: Academic Press.
- Goel, V. (2009). *Cognitive Neuroscience of Thinking*. In G. Berntson & J. Cacioppo (Eds.), *Handbook of Neuroscience for the Behavioural Sciences*: Wiley
- Grant, D. A., et Berg, E. A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting in a Weigl-type card-sorting problem. *J of Exp Psych*, 38.
- Havasi, C., Speer, R., et Alonso, J. (2007). *ConceptNet 3*. *Natural Languages Processing 2007*.
- Heaton, R., Chelune, G., Talley J., Kay, G., Curtiss (1993) *Wisconsin Card Sorting Test Manual: Revised and expanded*. Psychological Assessment Resources, Odessa, FL.
- Kahneman, D. (2011) *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kaplan, G.B., Sengör, N.S., et al. (2006). A composite neural network model for perseveration and distractibility in the WCST. *Neural Networks*, 19.

- Kimberg, D.Y. and Farah, M.J. (1993). An unified account of cognitive impairments following frontal lobe damage. *Journal of Exp. Psych.*, 122.
- Larue, O., Poirier, P., Nkambou, R. (2012). A Three level Cognitive architecture for the simulation of cognitive phenomena. CAI (Canadian Artificial Intelligence)
- Lieberman, M. (2009). What zombies can't do : a social cognitive neuroscience approach to the irreducibility of reflective consciousness. In K. Frankish & J. S. Evans (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 293-316): Oxford University Press.
- Nagahama, Y., Okina, T., Suzuki, N., Nabatame, H., et Matsuda, M. (2005). The cerebral correlates of different types of perseveration in the WCST.J *Neuro Psych*, 76(2).
- Nichols, S. and Stich, S. (2000) "A Cognitive Theory of Pretense." *Cognition*, 74, 115-147.
- Nyhus,E., Barcelo,F.(2009).WCST and the cognitive assessment of prefrontal executive functions. *Brain Cogn*, 71(3).
- Sherry, D. F. and Schacter, D. L. (1987). The evolution of multiple memory systems. *Psych Rev*, Vol 94(4).
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22
- Sloman, A. Chrisley, R.(2005). More things than are dreamt of in your biology: Information-processing in biologically inspired robots. *Cognitive Systems Research* 6(2).
- Stanovich, K. E. (2010). *Rationality and the reflective mind*: Oxford University Press.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Intelligence and rationality. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge Handbook of Intelligence* (pp. 784-826). New York: Cambridge University Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, 17(3).

philosophy of science (mainly of cognitive science and of neuroscience).

Roger Nkambou (Ph.D.) is a professor of Computer Science at the University of Quebec at Montreal, Director of its Cognitive Computer Science doctoral program and Director of the GDAC (Knowledge Management Research) Laboratory (<http://gdac.dinfo.uqam.ca>). His research interests include knowledge representation, intelligent tutoring systems, intelligent software agents, ontology engineering, student modeling and affective computing.

AUTHORS BIOGRAPHY

Othalia Larue is a PhD student in Cognitive Computer Science at the Université du Québec à Montréal, and member of the GDAC (Knowledge Management Research) Research Laboratory. Her research interests include cognitive simulations, cognitive architectures and affective computing.

Pierre Poirier (Ph.D.) is a professor of philosophy at the Université du Québec à Montréal and director of its Cognitive Science Institute. His interests are the

FROM PATIENTS' NEEDS TO HOSPITAL PHARMACY MANAGEMENT: AN HOLISTIC APPROACH TO THE PROCESS MODELLING

Guida R.^(a), Iannone R.^(b), Miranda S.^(b), Riemma S.^(b), Sarno D.^(b)

^(a) Department of Pharmaceutical and Biomedical Sciences - University of Salerno, Italy

^(b) Department of Industrial Engineering - University of Salerno, Italy

^(a) desarno@unisa.it

ABSTRACT

Materials management is a great issue for healthcare systems because it influences performances of structures in terms of clinical and financial outcomes. Enhancements can be done by means of an appropriate management of data related to these materials coming from all processes in which they are involved. This paper presents an integrated and detailed description and model of the hospital materials management process, able to tackle information from patient requirements to materials usage, according to traceability and streamlined processes perspectives. The holistic approach adopted, based on hospital cases analysis, literature review and international guidelines, tends to look at all actors, materials and processes involved, with the attempt to provide academicals and practitioners with a useful guide to the various technology-related, management and business issues that can arise during the design or reengineering of this business process and the associated information system.

Keywords: hospital materials management,-healthcare process modelling,-drug inventory management, healthcare logistics

1. INTRODUCTION

Materials management is a great issue for healthcare systems because it influences performances of structures in terms of clinical (service level offered to patients, quality, timeliness, etc.) and financial outcomes (expenditures on warehousing, handling, deterioration, etc.). Enhancements can be done by means of an appropriate management of data related to these materials coming from all processes in which they are involved.

Apart from the managerial choices to take, indeed, the description and analysis of physical and information integrated flows as they may, should or have to happen, is useful to achieve medical risk reduction and materials and information traceability.

Moreover, before selecting, adapting and implementing leading or optimized practices, it is necessary to develop a good understanding of processes and activities in place (Landry and Philippe 2004) in order to lay the foundations of an efficient

reengineering of supply chain that reduces healthcare costs without affecting the quality of care (Jarrett 1998).

In literature, many collateral references on how the materials management works are reported but it is really difficult to find a detailed, integrated and exhaustive hospital materials management workflow analysis and description. The aim of this paper is to develop such a process model with an holistic approach to the system, describing actors, tasks and information needed to manage materials from their acceptance to utilization.

2. PROCESS ELEMENTS

The first step to conduct the hospital pharmacy “micro-world” reengineering process towards optimization is to identify the behaviour of this system, such as what to manage in terms of materials, actors and processes, taking into account information and legal constraints. In the remainder of this paper, a detailed description of actors, materials and macro-processes is carried out, followed by a thorough explanation of all processes and sub-processes involved.

2.1. Actors

First of all, it is fundamental to profile all the actors that take part to materials management processes, distinguishing by the level at which they participate to them, that is:

- Medical Unit actors: together with patient needing for care, physician (who delivers the care), nurse (who dispenses the care) and nurse manger (who supervisions the dispense of care) are basically the other actors involved at the medical unit materials management stage;
- Pharmacy Unit actors: international health care standards require a central pharmacy unit in hospitals that maintains and provides the inpatient pharmacy needs (Joint Commission International 2011). The pharmacy unit is usually composed by hospital pharmacists, storekeepers and transporters;
- Other organizational functions: the Medical Director, the Superintendence and Treasure's Office, the Accounting Office take part to the process.

2.2. Materials

Management of materials in healthcare involves two kinds of items' clusters, that is drugs (or medicines) and medical devices, subjected to different regulations harmonized by countries according to international guidelines.

The properties of medicines in a hospital information system may be mandatory or optional depending on the contextual workflow (IHE 2010). A fundamental "identifier" is the ATC, Anatomic – Therapeutic – Chemical classification, internationally accepted and maintained by the World Health Organization. In addition to commercial drugs, drug administrations can also refer to galenics, such as personalized medicines prepared as a "mixture" of commercialized products at the bedside, in hospital pharmacy or in another defined medical unit. In parallel, medical devices can be required as surgical kit and apparatus compounded by many of them, that can be managed as single items by pharmacists or directly supplied as a pack.

The item list (in other words, the set of medicines or medical devices that can be administered/dispensed or implanted to patient in a healthcare system) changes from hospital to hospital, depending not only on healthcare services managed, but also on physician's expertise and preferences and following pharmacoeconomics principles.

2.3. Processes

Irani et al. in Vergidis et al. (2008) sustain that businesses should not be analysed in terms of the functions in which they can be decomposed to or in terms of the output they produce, but taking into account the key processes they perform.

The hospital medication workflow is triggered by patient needing drugs and medical devices. In this perspective the Technical Framework, developed for the pharmacy domain by the initiative Integrating the Healthcare Enterprise (IHE), was born to stimulate the integration of healthcare information systems operating with different standards (such as DICOM, HL7, etc.). In the proposed pharmacy interoperability model (IHE 2011), the care path (clinical perspective) described is orthogonally combined with the supply path (logistical perspective) in the phase of distribution but - as explicitly stated - supply chain of ordering/delivering medication and stock management are out of IHE scope.

According to the Irani et al. opinion, we adopt and refine the IHE pharmacy interoperability model as an "information track" to classify and describe all the elements depending on logistics decisions in the hospital medication workflow. We readapt that model to define in more detail different care paths, distinguishing among the type of medicine or medical device employed; moreover, we explore the management area, depicting all the processes involved.

Analysing some hospital cases, literature review and international guidelines and considering the clinical/logistical perspective, we define the hospital materials management as to be composed by the main following macro-processes, that are fully described, together with their sub-processes (reported in Figure 1) in the next sections:

- a) Patient Management
- b) Medical Unit inventory Management
- c) Centralized inventory Management

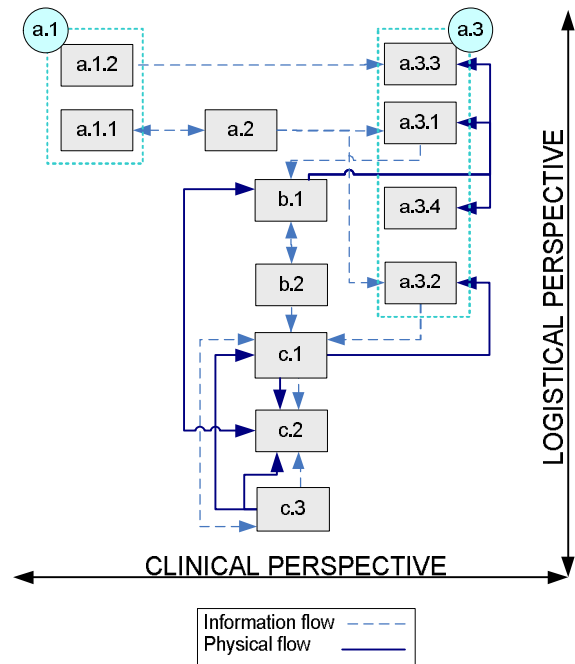


Figure 1: Process Framework of Materials Management in a Hospital (the notation used is explained later in the next sections)

3. PATIENT MANAGEMENT PROCESS (a)

The first macro-process is the patient management, since the need for care (and then the hospital medication workflow) triggers materials requirements.

When it comes to a patient admission (in emergency or outpatient department), a physician states a diagnostic hypothesis, attributing (with or without a confirmation) one or more diseases to the patient.

Depending on resources availability (beds, laboratories capacity, etc.) and urgency, (s)he also assigns a diagnostic and/or therapeutic pathway to be followed in the hospital or at home.

The hospitalisation in a medical unit can be immediate or delayed according to a queue list (for Inpatient access, Day Hospital, Day Surgery, etc.); hospitalisation or not, a patient anamnesis and a previous and current therapies recording are always required to know if they are compatibles with the pathway to perform.

The patient management process is composed by 3 sub-processes:

- a.1) Prescription
- a.2) Pharmacy prescription validation
- a.3) Materials delivery to patient

3.1. Prescription (a.1)

A prescription is the outcome of a clinical decision, is issued by a physician for one patient and may be the input of pharmaceutical validation and dispense. Sometimes, there can be a medication dispensed or administered outside the context of a prescription (for example, in case of urgency in the Emergency Department). They are considered as attached to an order session, which can be associated with a prescription.

We distinguish between two kind of prescriptions that are related to materials consumption:

- a.1.1) Drug Prescription or surgical intervention plan
- a.1.2) Exam prescription

3.1.1. Drug Prescription or surgical intervention plan (a.1.1)

In addition to pharmaceutical validation, a drug prescription is also the input of nurses' instructions to administer the drug for hospitalized patient or, in case of outside dispense, for patient home care. Variations in the content of prescriptions can occur, being dependent on country regulation, responsibilities and standard developed but, on the basis of regulations, IHE statements (2010) and interviews, the medical prescription data set is compounded, among others, by actors (Patient and Physician) Id, reason for prescribing, number of refills (for dispense), active substance and brand, dosage, frequency of intake and quantity. Alerts about prescribing restrictions and potential Intolerance, Contra-indication and Allergies (ICA) should be easily consulted.

The active substance(s), as stated before, is usually a key element because it can permit evaluations about pharmaco-economics and availability (for example, to distribute the cheapest brand name stored in satellite or central pharmacy or to prescribe what is stored in the local warehouse), but clinical consideration can imply a more rigid selection. The active substance(s), together with the pharmaceutical form, may allow to entry the medicine database and look for the most frequent disease for which it is adopted, ICA, maximum dosage per administration, frequency and times of standard intake depending on patient age, gender or weight. Physicians and pharmacists have to take into account all these data and combine them with patient characteristics to assure the best choice and to prevent mistakes.

In case of surgical intervention, some materials have to be explicitly prescribed (for example, an orthopaedic prosthesis and its size or surgical kit).

3.1.2. Exam prescription (a.1.2)

The physician can prescribes some diagnostic examinations to patients, such as laboratory exams (e.g. blood, RX) or other activities that require materials.

3.2. Pharmacy Prescription Validation (a.2)

After a prescription act, prescription information may be made available to pharmacists as a pharmacy validation can be delivered. It can be advisable but not mandatory, so that many organizations tend to jump this step. This implies no double reviews and, hence, no possibility of medical error reduction.

A detected problem can be a supply issue (suspended medication, out-of-stock., etc.), a legal issue (medication recalled or not allowed under certain conditions), or a medical issue (redundancy, interaction, contra-indication, etc. as an ICA).

3.3. Materials delivery to patient (a.3)

The diagnostic or therapeutic pathway involve materials requirements that can be indirectly related to patient because it is not convenient to address their usage to a specified activity performed ("General' goods usage") or directly related to it, as the case of "drug prescription or surgical intervention plan". Exam prescription triggering materials usage can fall in both of these classes. In more detail, four events can imply materials delivery to patient (see Figure 1) which are described in the following paragraphs:

- a.3.1) Preparation, administration or implant (inside dispense)
- a.3.2) Dispense to patient (outside dispense)
- a.3.3) Exam accomplishment
- a.3.4) "General" goods usage

3.3.1. Preparation, administration or implant (a.3.1)

When it's time to administer a medication (depending on the prescription), a message is sent to the nurse in charge and a preparation may take place. However, many hospitals do not have such an information tool or do not have an information system at all, so nurses are in charge of transcribing prescriptions on provided diaries, checking them to know when to administer. This implies risk of errors in transcription and administration execution (right time, person, etc.).

Preparation is the phase in which nurses take drugs from the stock and prepare them to be administered to patients. Some hospitals manage patient unit dose packed by robots or nurses at hospital pharmacy or medical unit warehouse, while others manage the most part of materials in the original package up to the patient bed.

Galenics (that are, as said before, extemporaneous composition of commercial drugs), are a special case because the preparation, depending on type and organization, may involve pharmacists and requires an appropriate recording of all actions and items used (information about drug id, lot number, quantity used, etc.) for traceability reasons. While an intravenous

injection of cortisone diluted in 0.9% sodium chloride solution can be prepared at patient bed, oncological galenics are usually made in a dedicated medical unit and later transported. Labelling with patient and galenic Id is the right way to reduce clinical risk while assuring traceability.

According to IHE suggestions, the assignment of a medication unit to a specific patient means a “dispense”; a galenics dispense can be considered as the instant in which labelling occurs.

All the medication are usually collected in a trolley or an already filled trolley is used, with a number of drawers containing the most frequently used drugs and other drawers destined to specific patients’ needs (this is the case described in Augusto and Xie 2009). Moreover, some medical devices such as syringes, gloves, roller bandages etc. have to be added to the trolley.

Inventory updated information are crucial, because previous checks should be done in order to assure that each patient has his own medication available at the time indicated in the prescription (naturally, this check is easy if an integrated information system is used). Together with quantity availability at the due date, expiration date control has to be done and the package with the closest expiration date should be drawn and disposed.

Administration follows. Nurse goes to patient bed and does the 7 “rights” checks (right patient, right drug, right dose, right route, right time, with the right information and documentation), administers the medication and records all the connected information, according to traceability and medical risk reduction objectives. Moreover, dispense (if done) or administration coincide with the recording of the consumption of medicine used (unit dose for commercial drugs or components for galenics and protocol – following the related bill of materials) from medical unit virtual stock so that warehouse management system is always up to date.

In case of surgical intervention, medical devices, kits or apparatus have to be used. After it, information about these materials has to be recorded as indicated before for drug administered, in order to satisfy clinical, traceability and inventory management requirements.

3.3.2. Dispense to patient (a.3.2)

In some healthcare organizations, hospital pharmacy can be charged also of materials dispense for home care (from outpatient departments or discharged patient). These materials are almost overlapping to that used by inpatients.

Outside dispense can be liken to medical unit dispense, because prescription is checked, material is taken from warehouse, recorded and given to the patient. In particular, a dispensation, among others, is composed by the refill number and the dispensing period.

3.3.3. Exam accomplishment (a.3.3)

Differently from “general goods usage” explained later, some medical devices can be directly attributable to one patient needs because a bill of materials may exist for each examination belonging to the deliverable service list of one hospital. Inventory levels are usually updated when examination is performed (back-flushing method) because examination duration is not so long to require single picking item recording.

3.3.4. “General” goods usage (a.3.4)

This process concerns medical devices utilisation not planned nor directly attributable to one patient care. This distinction is trivial because it is always possible to record each activity performed for one patient and its requirements in terms of materials, but most of the accountability/information systems does not elaborate on this function. For this reason, the medical devices belonging to this process are considered “general goods” for hospitals (examples are gloves to handle test-tube, elastic bandages).

4. MEDICAL UNIT INVENTORY MANAGEMENT PROCESS (b)

The second macro-process identified deals with the material management at medical unit level. Before going into deep of this process, we give some insight about places and policies concerning it.

First of all, regarding to places of materials management and dispensing, it is possible to distinguish between clinical and managerial perspectives. From a clinical point of view, the usefulness of having centralised patient-oriented pharmacy services to deliver professional services to patient (Carroll and Gagnon 1984) has been recognised. From a managerial perspective, many authors state the possibility to reduce logistics costs using a satellite pharmacy systems (Poley et al. 2004); numerous studies, finally, suggested the benefits of having central inventory control – rather than allowing department or region to deal with suppliers individually (O’Hagan 1995 cited in Jarrett 2006) – particularly in terms of drug waste reduction (Poley et al. 2004). In addition to a centralised point of distribution, however, it is necessary to have a decentralised materials management warehouse close to the point of delivery (Little and Coughlan 2008) to allow nurses to quickly reach stocks in the proximity of places (beds) where services (administrations) are delivered. The same holds good for medical devices, particularly in case of vendor managed inventory for surgical interventions. For these reasons, the inventory system is typically characterised by centralised and decentralised stores (de Vries 2010), with two types of warehouses: medical unit and hospital pharmacy. The first one has to deliver materials to the patients of the facility, the second one is in charge of managing items for inside (to medical unit) and outside dispense (to patients).

For what concerns policies of materials management, literature and practical cases are full of

different approaches. Going to the bottom of the issue, the two fields of intervention which need to be globally optimised are medical unit supplies and pharmacy supplies. Look-back (for example Re-Order Level, Re-Order Cycle, Just in Time), look-ahead (for instance Material Requirements Planning) or mixed (for example Vendor Managed Inventory) approaches can be used. All these approaches can be adopted inside the same healthcare system depending on the item typology (ordering prices, inventory costs, shelf life, demand mean and deviation, etc.). Cluster analysis has to be performed in order to group items that can be similarly managed. For example, galenics may have little or no stock, their components may be managed in lots. Look back approach is more popular than the other ones but it is characterised by higher inventory levels. Moreover, forecasts on aggregate data about consumptions recorded by pharmacy are influenced by medical unit materials management. Look-ahead techniques, instead, need for careful and punctual information about requirements forecast.

In what follows, the description of the two activities compounding the Medical Unit Inventory management process is reported:

- b.1) Medical unit stock management and re-ordination
- b.2) Pharmacy order assessment

4.1. Medical unit stock management and re-ordination (b.1)

Medical and surgical supplies close to the point of delivery (i.e. ward, theatre or laboratories) are usually managed by medical units, who are in charge of:

- handling incoming and materials to dispose;
- keeping warehouse management system up-to-date;
- dispensing and recording the dispense of materials triggered by a validated prescription (for preparation and administration or implant), materials for exam accomplishment, "General" goods usage;
- controlling stock levels and determining that a warehouse replenishment is needed.

A medical unit order (that is a replenishment request to hospital pharmacy) can be delivered for different motivations, depending on materials typology:

- the no-stock policy is adopted, as happens in Wong et. al (2003), where a medication ordering-dispensing and administration process triggered by patient needs is evaluated;
- there is a warehouse but a stock is not provided for this material, rarely prescribed;
- it is an out-of-stock item (also the security stock is consumed).

The order can be transmitted in form of a list to be sent to the hospital pharmacy, as a Kanban (in case of Just In Time method), as a requirement for each patient or, as happened in Augusto and Xie (2009), in the form of a mobile medicine closet to be partially refilled.

The typical materials' list, headed with list Id, date and requesting medical unit, contains items Id and quantity for each one, being comprehensive of raw materials belonging to galenics' bill of materials. According to the policy applied, in case of drugs, the item Id can refer to the brand name or the general name.

As Nicholson et al. (2004) claim, the most traditional servicing approach is the periodic review par level (or order-up to level), that requires to set the review interval and the optimal security stock (base stock level). While the second depends on therapeutic and medical constraints set by taking into account demand variability, the first has to be defined according to resources involved.

Indeed, implementing an inventory control tool which links prescription and consumption through an computerized prescriber order-entry system (considered the first step to adopt look-ahead techniques by Awaya et al. (2005)), could allow to deal with the complex and stochastic nature of demand adopting information about materials requirements of patients. Data coming from Electronic Patient Records, such as requirements and administrations concerning prescriptions for inpatients and outpatients, together with forecasts about incoming or belonging to waiting lists' patients, may be merged to develop an optimized supply and delivery planning.

Some examples of re-ordination strategies come from Kalmeijer et al. (2003), who promote the extensive use of Information Systems to manage requirements considering the default medication database as the local stock. Non-stock items are automatically ordered from pharmacy, instead.

4.2. Pharmacy order assessment (b.2)

If medical units are in charge of inventory management, pharmacists usually have to control re-orders' adequacy in terms of frequency and quantity for each material, real requirements and stock availability. Here, clarifications and changes can be made. The process ends with the release of the orders.

5. CENTRALIZED INVENTORY MANAGEMENT (c)

The final process is the centralized inventory management one, that links the internal requirements to the external supply chain. The sub-processes involved are:

- c.1) Pharmacy stock management, order disposition and supplying activities
- c.2) Internal distribution
- c.3) Materials Admission and Quality control and Payment

5.1. Pharmacy stock management, order dispositions and supplying activities (c.1)

After pharmacist order assessment, the stock management and supplying activities take place. Operationally speaking, the tasks to carry out are:

- handling of incoming order and expiration materials;
- keeping warehouse management system up-to-date;
- budget reconciliation assessment. Each cost centre/medical unit typically has its own budget to manage for each expenditure class, and the same for materials belonging to tenders, that have their specification budget. Materials have to be transferred according to them, otherwise a budget integration has to be requested - to Superintendence and Treasurer office - reporting quantities and extra-fund amount needed);
- stock levels control and authorization to dispense;
- supplying activity.

As happened to medical unit stock management, different inventory management policies can be implemented in order to take frequency/quantity supplying decisions.

The order is e-transmitted to the supplier or to a “Central of Purchasing”, whose task is to sort orders to the supplier and sends feedback information such as order acceptance or delivery due date.

An exception case is the out-of-stock of life-saving drugs or medical devices. Provided that hospital pharmacy has its own fund, a supplier different from the tender winners or more expensive than the usual ones can be contacted if provisioning lead time are supposed to be shorter.

5.2. Internal distribution (c.2)

This process is carried out starting from pharmacists authorization triggered by medical unit replenishments, eventually fulfilled by supply order (also transient materials).

Inside hospital pharmacy delivery systems have a key role in hospital’s service quality (Yurtkuran and Emel 2008). The diffusion of multi-echelon inventory problem opposed to scheduling oriented ones (Lapierre and Ruiz 2007) has often coped with in literature, and different transportation solutions or search for optimized routing problems (for example, see Augusto and Xie 2009) are presented, in which transportation costs and resources (both personnel and physical resources like means of transports etc.) are some of the main variables taken into account. The reason is the pervasiveness of satellite pharmacies and hub and spoke models in which geographical distances among medical units to be served can be great. Another issue is resource sizing that depends, inter alia, on frequency

and quantity of transport from pharmacy to medical units, quantity to handle due to arrivals, characteristics of material handling equipment.

The order preparation, performed by storekeepers, consists in printing (also virtually) the materials list released by pharmacists for a medical unit (if the list exists and no refilling method is applied), labelling and filling a bag/receptacle with all materials listed. FIFO logic should be adopted, considering materials expiration date.

The second task, that is the delivery to medical unit, can be carried out immediately or according to a schedule by transporters (that can belong to pharmacy personnel or medical units’ nurses staff).

Materials are considered as booked for medical unit till the drawing occurs, when it is subtracted from pharmacy inventory records and registered as “incoming” in medical unit warehouse system. In the end, it is virtually added to medical unit warehouse stock. This operation can be immediate or delayed depending on the information system, the traceability features involved and the other tools available for technicians.

5.3. Materials Admission and Quality control and Payment (c.3)

When the order is dispatched and the incoming materials are delivered, pharmacy personnel has to:

- Check the correspondence in type and quantity between and delivered goods;
- Accept with reserve the goods;
- Verify label conformity, item code, quantity, batch number, physical integrity;
- Transmit the definitive acceptance to the Accounting Office to perform the payment (through a Central of Purchasing or directly).

If the item is managed by pharmacy, products are handled and stored in the dedicated warehouses (chilled or not), otherwise (in case of “transient” material or belonging to a urgent order), it has to be transported (internal distribution is activated). Quantity and Lot Number are registered where materials are firstly handled.

For transient materials kept in consignment stock, the consumption corresponds to an invoice and a replenishment order release.

6. CONCLUSIONS

This study is the result of literature review, interviews and the analysis of international guidelines, presenting a new holistic approach to the hospital materials management issue, defining all materials, actors and processes involved and focusing on the coherence, traceability and integration of information and physical flows. In doing so, it attempts to provide academics and practitioners with a useful guide to the various technology-related, management and business issues

that can arise during the design or reengineering of this business process and the related information system.

The description and formalization of all tasks and processes involved in the hospital materials management - enabling data sharing and, at the same time, traceability - have a double scope. First, in the clinical perspective, it allows to enhance patient safety through medical risk reductions (for example, suggesting mandatory matching between drug prescribed and to be administered, or avoiding transcriptions), to be ready in case of recall and to efficiently manage drug shelf life and expiration dates; second, from the logistical point of view, it lays the foundations for optimising the materials management because an exhaustive, robust and flexible basis of information can permit to collect data - from materials prescriptions to their delivery, from orders to fulfillment lead times -, identify key performance indicators and compare them in different management techniques scenarios, carrying out performance evaluations.

REFERENCES

- Augusto, V., Xie, X., 2009. Redesigning pharmacy delivery processes of a health care complex. *Health Care Management Science*, 12(2), 166–178.
- Awaya, T., Ko-ichi, O., Takehiro, Y., Kuniko, Y., Toshiyuki, M., Yu-ichi, I., Yoshikazu, T., Nobumasa, H., Kazuo, M., 2005. Automation in drug inventory management saves personnel time and budget. *Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan*, 125(5), 427–432.
- Carroll, N.V., Gagnon, J.P., 1984. Consumer demand for patient-oriented pharmacy services. *American Journal of Public Health*, 74(6), 609–611.
- de Vries, J., 2011. The shaping of inventory systems in health services: A stakeholder analysis. *International Journal of Production Economics*, 133(1), 60–69.
- Integrating the Healthcare Enterprise, 2010. *IHE Pharmacy. Technical Framework Supplement. Common Parts, Trial Implementation*. Available from: http://www.ihe.net/Technical_Framework/upload/IHE_Pharmacy_Suppl_Common_Rev1-1_TI_2010-12-30.pdf [accessed 12 April 2012]
- Integrating the Healthcare Enterprise (IHE), 2011. *IHE Pharmacy. Technical Framework Supplement. Hospital Medication Workflow, Trial Implementation*. Available from: http://www.ihe.net/Technical_Framework/upload/IHE_Pharmacy_Suppl_HMW_Rev1-2_TI_2011-12-31.pdf [accessed 12 April 2012]
- Jarrett P.G., 1998. Logistics in the health care industry. *International Journal of Physical Distribution and Logistics Management*, 28(9), 741–772.
- Joint Commission International, 2011. *Accreditation Standards for Hospitals, 4th edition*. USA: 2010 Joint Commission International.
- Kalmeijer, M.D., Holtzer, W., van Dongen, R., Guchelaar, H.-J., 2003. Implementation of a Computerized Physician Medication Order Entry System at the Academic Medical Centre in Amsterdam. *Pharmacy World & Science*, 25(3), 88–93.
- Landry, S., Philippe, R., 2004. How Logistics Can Service Healthcare. *Supply Chain Forum: an International Journal*, 5(2), 24–30.
- Lapierre, S., Ruiz, A., 2007. Scheduling logistic activities to improve hospital supply systems. *Computers & Operations Research*, 34, 624–641.
- Little, J. & Coughlan, B., 2008. Optimal inventory policy within hospital space constraints. *Health Care Management Science*, 11, 177–183.
- Nicholson, L., Vakharia, A.J., Erenguc, S.S., 2004. Outsourcing inventory management decisions in healthcare: Models and application. *European Journal of Operational Research*, 154, 271–290.
- Poley, M.J., Bouwmans, C.A.M., Hanff, L.M., Roos, P.J., van Ineveld, B.M., 2004. Efficiency of different systems for medication distribution in an academic children's hospital in the Netherlands. *Pharmacy World & Science*, 26(2), 83–89.
- Vergidis, K., Tiwari, A., Majeed, B., 2008. Business Process Analysis and Optimization: Beyond Reengineering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(1), 69–82.
- Wong, C., Geiger, G., Derman, Y.D., Busby, C.R., Carter, M.W., 2003. Redesigning the medication ordering, dispensing, and administration process in an acute care academic health sciences centre. *Proceedings of the Winter Simulation Conference 2003, IEEE*, 1894–1902 vol. 2. December 7-10, New Orleans (Louisiana, USA).
- Yurtkuran, A., Emel, E., 2008. Simulation based decision-making for hospital pharmacy management. *Proceedings of the Winter Simulation Conference 2008, IEEE*, 1539–1546. December 7-10, Miami (Florida, USA).

IMPROVING PATIENT'S WAITING TIME AT A HEALTH SCREENING CENTER

Wongsammacheep T^(a), Pichitlamken J^(a), Weerawat W^(b*)

^(a)Department of Industrial Engineering, Kasetsart University, Bangkok, Thailand

^(b)Department of Industrial Engineering, Mahidol University, Nakhon Pathom, Thailand

^(a)w.thanon@gmail.com, ^(a)Juta.p@ku.ac.th, ^(b*)egwvr@mahidol.ac.th

ABSTRACT

The health screening center is the first department that patients come into contact before going to other departments. Patients sometimes complain about long waiting times at this center. We develop a discrete-event simulation model of the health screening center to support the decision making process of the hospital management. It is designed such that it can readily be used for testing any scenario inside the health screening center. Input data is collected from electronic records and interviews with staff. The simulation model is validated by considering the average total times in the system of one health checkup package. In this paper, we use the simulation model to test our patient's adaptive policies to see if patient's waiting time can be reduced.

Keywords: discrete-event simulation model, health screening center, healthcare simulation

1. INTRODUCTION

Private hospital business is expected to grow because more and more patients focus on preventive healthcare. Thus, a health screening center will receive greater number of patients because its main purpose is to identify future health risks so that the action can be taken immediately or highlight any problem areas. In addition, health screening can create or improve self awareness of health and fitness and provide referrals for further care when necessary. It is the first department that many patients come into contact with before going to other departments.

We study the health screening center at one of the forefront private hospitals of Thailand. The medical personnel are renowned for their expertise and cutting-edge medical treatments. The health screening center is open from 6 AM to 5 PM. For check-ups, patients cannot eat or drink at least nine hours before; therefore, most patients arrive to the health check up center during 7 AM to 10 AM. Some medical tests take a long time, resulting in occasionally long waiting time for patients. The current target of average total waiting time is less than 1 hour, but the actual value is approximately 2 hours. The satisfaction surveys done by the hospital indicate that it is an issue that patients are not satisfied with. Some factors that affect the waiting time are the variation of medical tests in the health screening

packages, the number of hospital staff, and the patient appointments. The hospital management would like to have a simulation model so that it can experiment with improvement schemes.

Discrete-event simulation (DES) models are computer programs that model the logical flow of complex processes occurring at discrete times (Kelton et al 2009). DES has become one of the most widely used Operations Research tools, including in healthcare applications. Everett (2002) describes simulation models that support the decision making process for scheduling of patients. Klein et al (1993), Jun et al (1999), and Jacobson et al (2006) provide a comprehensive literature review on simulation modeling and applications to healthcare.

2. DESCRIPTION OF THE HEALTH SCREENING CENTER

The health screening center under study consists of 6 stations. Figure 1 shows the layout of health screening center. Station 1 is the pre-assessment to inquire patients. Station 2 is the document preparation for medical exams. Station 3 is staffed by cashiers for making a payment. Station 4 is the vital sign checks and blood draw examination. Station 5 has two parts: the first part is the cardiac assessment consisting of the electrocardiogram (EKG) and the exercise stress test (EST), and the second part is radiology (Imaging services) consisting of chest x-ray, ultrasound of abdomen, a breast cancer exam and a mammogram. Station 6 is the physical examination and diagnosis, consisting of physical examination (PE), eye examination, pap smear and pelvic examination.

This health screening center offers 7 packages (see Table 1 for details of medical exams in each package). Patients make appointments before arrivals, or they are walk-in patients. When patients arrive to the health screening center, they go to Station 1 to register and do pre-assessment. If patients have made appointments, all the documents are readily available for them to sign when they arrive. Patients proceed to Station 3 to make a payment after getting the documents. Actual medical exams begin at Station 4 with the vital sign check and blood test. When patients finish with Station 4, they are split into two paths: the first one leads to Station 5 and then Station 6, whereas the second path goes to Station 6 first and Station 5 later, but patients need to return to

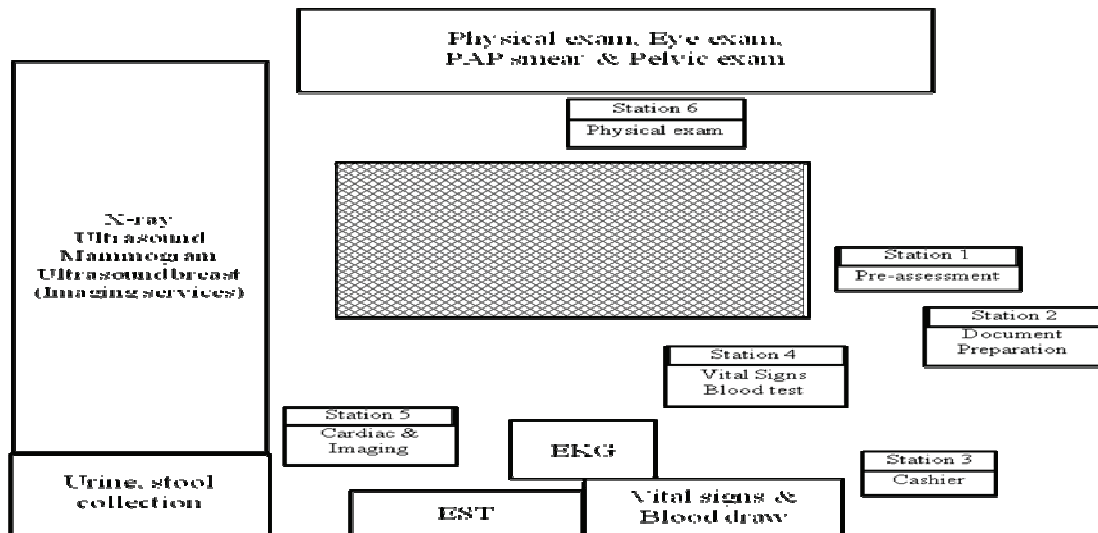


Figure 1: Layout of Health Screening Center

Station 6 again to see the same physicians for diagnosis. Patients are directed to the second path when there are more than 10 persons in the ultrasound queue. Because Station 5 consists of cardiac assessment and imaging services (the order does not matter), patients are sent to the shortest queue. The patient flow diagram is summarized in Figure 2.

3. SIMULATION MODEL DEVELOPMENT

We describe the data collection process in Section 3.1, the overview of the simulation model in Section 3.2 and the model validation in Section 3.3.

3.1. Data Collection

Empirical data was collected from computer recorded files, by interviews with doctors, nurses, and other hospital staff, and by direct observations, during July to December 2011. For the first two months, we learn about the health checkup operations, such as job details at each station, staff's work schedule, and how the staff makes decisions. Then for the subsequent month, we construct the patient's flow diagram. During the remaining 3 to 4 months, we go to the health screening center 4-5 days a week to collect time data, to interview with the staff, and to retrieve data from their patient record system (Amalga[®]).

Many issues arise during data collection. For example, hospital staff are sometimes reluctant to provide us data because they think we will use it to catch their mistakes. Therefore, we discuss with the department manager what we needed and what they would get from our work. We also have to create good personal relationship with the staff so that they are willing to give us interview time, and they can also help us on data collection.

For this model, we require input models for the patients' arrivals, service times at each station, resource availability, and fractions of patients undergoing each checkup package. The distribution of patients on checkup packages is as follows: the Regular package accounts for 10% of patients, Executive package: 17%,

Executive Female: 7%, Comprehensive Male without EST: 10%, Comprehensive Male: 27%, Comprehensive Female younger than 40: 8%, and Comprehensive Female older than 40: 21%. For the number of patient arrivals, we consider only weekdays and no holiday. The data is examined for seasonality and trends, and we do not find significant trends. We assume that the arrivals are independent of the day of the week and time of the year. The average number of patients is 170 per day, and we simulate their arrivals with the non-stationary Poisson Process with the following distribution: 2% of patients arrive during 6-7 AM, 22% during 7-8 AM, and the remaining hourly fractions until the center stops receiving new patients at 1 PM are 25%, 22%, 15%, 10% and 4%, respectively.

The service times at each station are collected by a stopwatch time study. Table 2 shows the input models in Arena[®] expressions. When patients have ultrasound tests, they need to be full bladder. The ultrasound test is done, on the lower and upper abdomen, one at a time. The physical examination consists of 2 parts: consultation and diagnosis, both of which have to be done by the same physician.

The resources at the health screening center are *registered nurses* (RN) who do pre-assessment at the registration, *nurse aids* (NA), *clinic associates* (CA) who are coordinators, *technicians* and *radiologists* work on imaging, and *physicians*. We assume that all staff of the same position are equally skilled. Resource availability during each time period is shown in Table 3.

3.2. Simulation Model Development

Waiting time in queue is the primary key performance indicator (KPI) that the hospital management is keenly concerned. Other KPIs include the total time in the health checkup center, average waiting time by package and by time period, total time by package and by time period, average total waiting time and total time of each path.

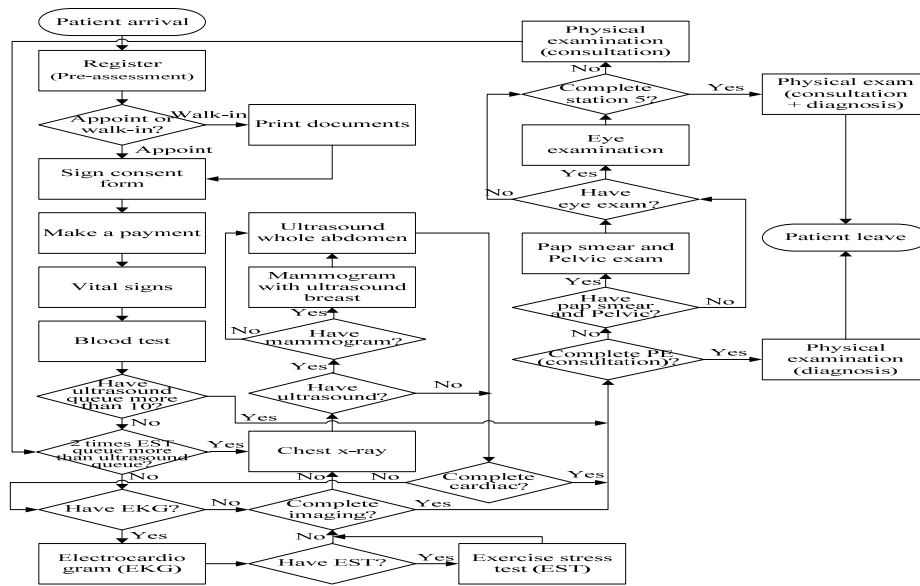


Figure 2: Patient Flow Diagram

Table 1: Examination in Each Checkup Package

	Regular	Executive	Executive female	Comprehensive male without EST	Comprehensive male	Comprehensive female younger than 40	Comprehensive female older than 40
Vital Signs	●	●	●	●	●	●	●
Blood test	●	●	●	●	●	●	●
Electrocardiogram		●	●	●	●	●	●
Exercise stress test					●		
Chest x-ray	●	●	●	●	●	●	●
Ultrasound whole abdomen		●	●	●	●	●	●
Mammogram with ultrasound							●
Pap smear & Pelvic exam			●			●	●
Eye exam				●	●	●	●
Physical examination	●	●	●	●	●	●	●

Table 2: Process Time Distribution (Unit: Min)

Process	Arena Expression
Signing consent forms	NORM(5,2)
Printing consent forms	0.5 + EXPO(1.04)
Printing order documents	0.5 + EXPO(0.857)
Charging program	DISC(0.95, 1, 0.99, 2, 1.0, 3)
Assessment for appointment patients	0.5 + LOGN(2.35, 2.36)
Assessment for walk-in patients	0.5 + GAMM(1.71, 1.53)
Cashier	8 * BETA(3.54, 8.24)
Vital signs	2 + GAMM(0.745, 2.43)
Drawing blood	1.1 + LOGN(1.48, s0.743)
Electrocardiogram (EKG)	2 + WEIB(1.82, 1.5)
Exercise stress test (EST)	NORM(24.9, 3.59)
X-ray	0.08 + LOGN(1.75, 0.696)
Mammogram	NORM(8.78, 2.14)
Ultrasound whole abdomen	5 + 26 * BETA(1.47, 1.77)
Ultrasound upper abdomen	9 + WEIB(9.44, 1.45)
Ultrasound lower abdomen	3 + 11 * BETA(1.32, 2.34)
Ultrasound breast	TRIA(6, 8.25, 18)
Eye exam	NORM(4.57, 2.14)
PAP smear and Pelvic exam	NORM(5.95, 2.46)
Physical examination (consultation)	NORM(5.71, 1.69)
Physical examination (diagnosis)	NORM(8.00, 2.74)
Physical examination (consultation and diagnosis)	NORM(10.30, 6.36)

Table 3: Resource Availability by Time Period

	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18
RN at Station 1	3	5	5	3	3	2	1	1	1	1	1	0
CA at Station 1	1	5	4	3	1	0	0	0	0	0	0	0
CA at Station 2	3	4	4	4	4	4	4	4	4	4	4	0
CA at Station 3	1	3	3	3	3	3	3	3	3	2	1	0
Vital signs NA	1	3	3	3	3	1	1	1	1	1	1	0
Blood draw NA	1	3	3	3	3	1	1	1	1	1	1	0
EKG NA	1	3	3	3	3	2	2	2	2	1	1	0
EST NA	1	2	2	4	3	3	3	3	3	1	1	0
EST RN	1	1	1	4	3	3	3	3	3	1	1	0
X-ray technician	1	1	1	1	1	1	1	1	1	1	1	0
Ultrasound technician	0	4	7	8	8	8	8	6	6	3	3	0
Ultrasound radiologist	0	2	3	4	4	4	4	2	2	1	1	0
Mammogram technician	1	1	1	1	1	1	1	1	1	1	1	0
Mammogram radiologist	1	1	1	1	1	1	1	1	1	1	1	0
Eye physician	0	1	1	1	1	1	1	1	1	1	1	1
PAP physician	0	1	1	1	1	1	1	1	1	1	1	1
Station 6 physician	0	1	1	1	1	1	1	1	1	1	1	1

The simulation model is developed on Arena version 12 (Rockwell Software) with the run length of 720 minutes (from 6 AM to 6 PM) and no warm up period. One-hour additional time is to allow the department to clear patients out of the system; it does not close until the last patient leaves. We ignore some details of the real setting, e.g., patients can request physicians whom they would like to see. Sometimes they arrive to a health checkup center in a group. If they cannot wait for the diagnostic results, the hospital can send them by e-mails or by post. We make the following simplifications in our simulation model: we assume that patients do not make physician requests; patients arrive one at a time, and patients wait for their diagnosis results, except when they arrive at Station 6 after 4 PM

3.3. Model Validation

Validation is the process of checking if the simulation model can adequately represent the real system. In this paper, we use the average total time of comprehensive male package going the first path (Station 5 first and Station 6 later) for validation. Rossetti (2010) describes the two sample *t*-test for comparing two means to validate the data from the simulation model with that from the actual system. Based on the sample size of 30 replications, the *p*-value of the test is 0.1; therefore, the simulation model can be used to model the actual system.

3.4. Scenario Comparison

We are interested in reducing total waiting time in the system. We compare the following five scenarios:

- *Policy 1*: As-is condition.

- *Policy 2*: To have the ultrasound test only once for the whole abdomen. Currently, if patients do not have full bladder, they will have ultrasound tests on the upper abdomen first. Then they keep drinking water until feeling full bladder and return to have ultrasound tests on the lower abdomen. Thus, the ultrasound radiologists work on them twice. But if they wait until being full bladder, the ultrasound radiologists will meet them only once.
- *Policy 3*: To assign diagnosis physicians at Station 6 on the first-come, first-serve basis. At present, physicians are assigned to patients when they check in at the registration, in order to balance the physicians' work load. When patients are finally due for diagnosis, they may experience long waiting time, even though some physicians at Station 6 may be available at that moment.
- *Policy 4*: To evenly schedule patients' arrivals throughout the day. Due to food and drink fasting, most patients arrive to the health screening center in the early morning, resulting in heavy traffic that propagates down the system. But if we can set patients' appointments such that patients evenly arrive, the health screening center will face less traffic variation and thus less system congestion.
- *Policy 5*: To simultaneously implement all policies.

With the run length of 720 minutes and 30 replications, the patient's total waiting times are shown in Table 4. As expected, Scenario 5 is most effective in reducing the total waiting time because all policies are included. However, if these policies are examined individually, Policy 4 of having equal arrival rate per

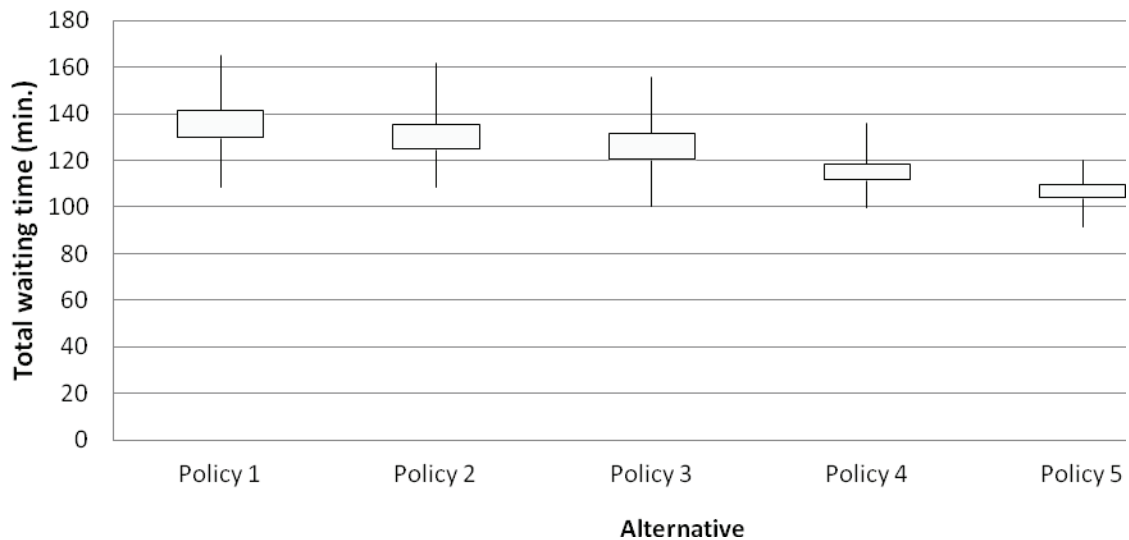


Figure 3: Alternative Comparison in Terms of Total Waiting Time

Table 4: Comparison of Total Waiting Time

	Total waiting time (min)	95% confidence interval (min)
Scenario 1	135.64	(129.81,141.47)
Scenario 2	130.08	(125.05,135.11)
Scenario 3	126.16	(120.59,131.73)
Scenario 4	114.9	(111.51,118.29)
Scenario 5	106.64	(103.87,109.41)

hour can reduce total waiting time more than other policies (Figure 3).

Policy 2 of having ultrasound tests once for each patient cannot reduce the waiting time by much because the process time of two ultrasound tests and one test do not differ by much. Policy 3 is more effective than Policy 2 because physician resources are pooled; patients are assigned diagnosis doctors on the first-come first-serve basis.

4. CONCLUSION

Simulation models are very useful for what-if analysis. In this paper, we develop our simulation model for experimenting with improvement plans at the health screening center. We validate it with a statistical test and show that it can adequately represent the actual system.

ACKNOWLEDGMENTS

The simulation model used in this paper has previously appeared in the Proceedings of the 4th International Conference on Applied Operational Research (ICAOR) Conference. This research is financially supported by the Office of the Higher Education Commission (Thailand) and Mahidol University under the National Research Universities Initiative.

REFERENCES

Everett, J.E. 2002. A decision support simulation model for the management of an elective surgery waiting system. *Health Care Management Science*, 5, 89–95.

- Jacobson, S.H., Henderson, S.N., and Swisher, J.R., 2006. Discrete-event simulation of health care systems. In: R.W. Hall, ed. *Patient flow: reducing delay in healthcare delivery*. *Springer International Series in Operations Research and Management Science*, 211–252.
- Jun, J.B., Jacobson, S.H., and Swisher, J.R., 1999. Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society*, 50 (2), 109–123.
- Kelton, W.D., Sadowski, R.P., and Swets, N.B., 2009. *Simulation with Arena*. 5th ed. McGraw-Hill Companies. Singapore.
- Klein, R., Dittus, R., Roberts, S., and Wilson J. R., 1993. Simulation modeling and health-care decision making (annotated bibliography). *Medical Decision Making*, 13, 347–354.
- Rossetti, M.D. 2010. *Simulation Modeling and Arena*. John Wiley & Sons Ltd. United States of America.

AUTHORS' BIOGRAPHIES

Wongsammacheep T. is a Research Assistant in the Department of Industrial Engineering, Kasetsart University, Bangkok, Thailand. He earned Bachelor of Engineering (Chemical Engineering) from Kasetsart University, Thailand.

Pichitlamken J. is an Assistant Professor in the Department of Industrial Engineering, Kasetsart University. Her research interests include ranking and selection procedures, simulation optimization, spreadsheet simulation, and stochastic processes.

Weerawat W. is an Assistant Professor in the Department of Industrial Engineering, Mahidol University. Her research interests include discrete-event simulation, logistics, and healthcare.

REDUCTION OF TURNAROUND TIME IN A HOSPITAL'S CLINICAL LABORATORY BY SIMULATION MODELING

Luangmul K^(a), Pichitlamken J^(b) and Weerawat W^(c)

^{(a)(b)}Department of Industrial Engineering, Kasetsart University, Bangkok, Thailand

^(c)Department of Industrial Engineering, Mahidol University, Nakhon Pathom, Thailand

^(a)natural_star@hotmail.com, ^(b)Juta.p@ku.ac.th, ^(c)egwvr@mahidol.ac.th

ABSTRACT

We develop a simulation model of a clinical laboratory for a complete blood count (CBC) test in a large private hospital. The model will be used for experimenting with new lab layouts and new work processes for the CBC test to reduce the turnaround time which is defined as total time in process. The average value from the simulation model is 72.07 ± 0.9 minute compared with 73.08 ± 22.9 minute from the empirical data. A new CBC test layout was created by the Relationship Diagram method to improve the work flow of medical technologists. The new turnaround time is reduced to 57.03 ± 1.12 minutes.

Keywords: Turnaround Time, Clinical Laboratory, Stochastic Discrete-Event Simulation, Relationship Diagram

1. INTRODUCTION

The Royal Thai government announced a policy to promote Thailand as a medical hub of Asia. The number of foreign and Thai patients has increased steadily due to a growing interest in personal care. The number of foreign patients was 1,103,095 in 2005 compared with 973,532 in 2004 and 630,000 people in 2003 (Department of International Trade Promotion, Ministry of Commerce, the Royal Thai Government).

Discrete-event simulation utilizes a mathematical or logical model to represent the actual system. Both the nature of the state change and the time at which the change occurs require precise descriptions (Albrecht 2012). Simulation modeling is chosen as our analysis tool because it is able to model complex systems. For example, Ahmed and Alkhamis (2008), Pirolo et al (2009), and Venkatadri et al (2011) use simulation models to evaluate staffing policies. Perkiang (2010) study causes and remedies for reducing the return of blood specimen to the center and then assign solutions to the relevant agencies.

Simulation modeling has been applied to solve various healthcare problems, including scheduling (Proctor 1996), admissions policy evaluation and operational improvements in outpatient facilities (Swisher et al 2001, and Duguay and Chetouane 2007). Blasak et al (2003), and Sinreich and Marmor (2005) use simulation models to reproduce the behavior of a healthcare system in order to evaluate its performance and analyze the outcome of different scenarios of the

emergency department. Deb and Bhattacharyya (2005) use the relationship diagram to design manufacturing facilities.

The hospital under study is a world-renowned private hospital in Thailand, offering many specialized service such as cancer centers, a dialysis center, a doctor golf clinic, and a spine center, serving both Thai and international patients. Questionnaires reveal that patients are not satisfied with long waiting time. Because doctors' prognosis require lab results, if the lab turnaround time is reduced from the current average of 73.08 minutes, the patient waiting times will be reduced. Note that the target turnaround time is 60 minutes, 17.8% smaller than the current average.

The laboratory has three test groups: composite hematology, immunology, and biochemistry. In this paper, we only consider the complete blood count (CBC) test which is one type of the hematology tests. The CBC test process is semi-automatic, involving an automatic machine with lab staff. The CBC test includes counting the number of white blood cells, determination of hemoglobin guides, counting the type of blood cells, and estimating the number of platelets. The CBC test is done to determine the body abnormalities for the purpose of timely treatments. The CBC lab layout is shown in Figure 1, and the test procedure is as follows:

1. The Specimen is sent to the laboratory, and the assistant medical technologist (AMT) receives the specimen into the computer system by reading the barcode affixed to the specimen.
2. The AMT places the specimen at the station where the medical technologist (MT) is waiting.
3. The MT puts the specimen into the CBC test machine. The procedure depends on the order types:
 - 3.1. When the order is urgent, the MT places the specimen at Station Blood Smear Manual. She chooses if she wants to do the *tint slide auto* or the *tint slide manual*. The *tint slide manual* is done at the Station Blood Smear Manual. The *tint slide auto* is done at the CBC test machine.
 - 3.2. When the order is not urgent, the MT holds a specimen to be put in front of the

CBC test machine, which runs automatically. If there are special cases, proceed to Step 4. If the test result is higher than the upper limit, start blood smear and tint by an auto machine. If the blood smear machine fails, or the doctor wants to look at slides, the MT does the blood smear manual.

4. The specimen undergoes the blood smear process
 - 4.1. Slides are placed at the station for blowing and wiping.
 - 4.2. Slides are moved to the microscope station. The MT scans the microscope, and results are recorded into the computer system.
5. MT verifies the results and input into computer system.

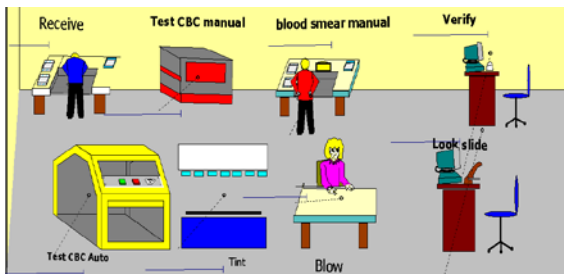


Figure 1: Physical Layout of the CBC Lab

2. SIMULATION MODEL DEVELOPMENT

We consider the data collection in Section 2.1, the simulation model construction in Section 2.2, and the model verification and validation in Section 2.3.

2.1. Data Collection

The data under consideration consists of 615 blood samples that were collected over three months. The arrival process is divided into 24 one-hour intervals. We collect the number of resources such as MTs, AMTs, and the CBC test machine by interviews. The times required for various activities associated with the CBC test were collected by the stop-watch time study, i.e., manually recording the process time with a stop-watch. The raw data was used to fit parametric probability distribution (Table 1). Table 2 shows the distance between work stations under the current lab layout. Figure 2 shows arrival rates to the CBC test, and the arrival rate during 07.00 AM – 01.00 PM is the highest.

The major problem in data collection is the number of timekeepers; there was only the author who did the job. To get their collaboration, we also need to build personal relationship with the MT and AMT by explaining the benefits to be gained from our work.

2.2. Simulation Model Construction

The simulation model was developed using the Arena software package, version 12.0. The Arena's *Entity* that flows through our simulation model is the blood specimen for the CBC test. The resources are MTs, AMTs, the blood smear machine, the CBC test machine, and a microscope. Staffs work on 2 shifts: the first shift is 07.01 AM – 04.00 PM and the second shift is 04.01 PM – 07.00 AM. The numbers of MT and AMT are 4 MTs in the first shift and 2 people in the second shift; 2 AMTs in the first shift and 1 AMT in the second shift. Other number of resources stays constant: 2 blood smear machines, 2 CBC test machines, and 1 microscope.

Other Input data for the simulation model include the arrival process of specimens, the MT and AMT schedules, activity time distribution, and walking distance from one station to another.

Table 1: Activity Time Distribution (unit: minute)

Process	Arena's expression
Receiving specimen (1)	TRIA(0, 0.26, 4.73)
Manual blood smearing (3.1)	TRIA(0.15,0.23,2.7)
Auto CBC test (3.2)	NORM(6.75,1.97)
Auto blood smearing (3.2)	NORM(3.22,0.36)
Auto tinting (3.2)	6.24+Expo(0.074)
Blowing and wiping (4.1)	NORM(0.275,0.129)
Scanning electron microscopy and results record (4.2)	TRIA(0.01,0.034,2.72)
Verifying (5)	0.07 + LOGN(0.153, 0.0972)

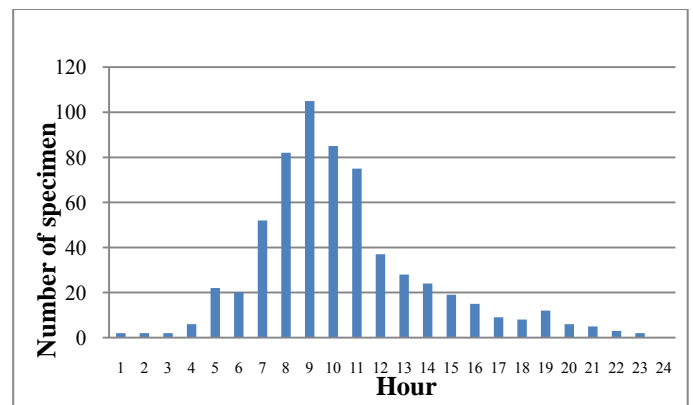


Figure 2: Arrival Rate in CBC Test

2.3. Model Verification and Validation

Verification and validation are key steps in simulation model development. Verification is done to determine if the computer simulation model matches to the analysts' concept. Our model was verified by explaining the model logics to the lab staff.

The validation step is the comparison of turnaround time received from the model and empirical values of the actual system. We use a two-sample *t*-test for validation (see Rosetti 2010 for details). Table 3 shows the averages and standard deviation of turnaround time from simulation and that of actual data.

The *p*-value of the 2 sample *t*-test is 0.8 which is more than the significance level 0.05.

Table 2: Distance in the CBC test lab

From	To	Distance (m)	
		Current Layout	Proposed Layout
Receiving	Waiting for MT	5.20	1.25
Waiting for MT	Manual CBC test	6.15	2.81
Waiting MT	Auto CBC test	12.00	1.20
Waiting for MT	Manual blood smearing	5.65	2.35
Manual CBC test	Manual blood smearing	0.50	1.01
Manual CBC test	Verifying	2.00	1.00
Manual blood smearing	Auto tinting	8.00	0.70
Manual blood smearing	Microscope scanning	2.50	0.80
Auto tinting	Microscope scanning	6.10	1.61
Auto tinting	Verifying	1.30	2.41
Microscope scanning	Verifying	6.00	1.00

Table 3: Comparison of Turnaround Time from Simulation and the Actual System

Data type	Sample size	Average (min)	Standard deviation
Simulation	400	72.07	0.9
Actual data	13	73.08	22.9

3. LAB LAYOUT DESIGN

We use the Activity Relationship Chart and Relationship Diagram (Table 4) to design a new layout. The relationship diagram is a tool for identifying relationships between areas in a plant or pair activities by showing how closely two areas should be located with a rating level (A, E, I, O, U and X). The Relationship Diagram shows a location of each activity by converting the Activity Relationship Chart of “closeness-ratings” into a diagram (Muther 1961). and In this paper, closeness scores are provided by the MT. Layout planning rests on the three fundamental of relationship between the activity in layout, space for each activity area, and adjustment of relationship and space interested near of each stations. Closely related activities are assigned to be in proximity of one another.

Table 2 shows the distance between stations of the new layout and the old one. Stations that have high closeness ratings are placed closer to one other. For example, the distance between the manual blood smearing station and the microscope scanning station decreases from 8.00 meter to 0.70 meter.

Table 4: Activity Relationship Chart

process CBC test		1	2	3	4	5	6	7	8	9	10	11	12		
		Receiving specimen	Receiving computer	Waiting for MT	Manual blood smearing	Manual tint	Manual CBC test	Auto test CBC	Auto blood smearing	Auto tint	Blow slide	Microscope scanning	Verifying		
1	Receiving specimen	X	A	I	U	U	U	U	U	U	X	U	U		
2	Receiving computer system		X	A	U	U	U	U	U	U	X	U	U		
3	Waiting for MT			X	U	U	U	I	U	U	X	U	U		
4	Manual blood smearing				X	A	I	U	U	U	A	I	U		
5	Manual tint					X	O	U	U	U	A	I	U		
6	Manual CBC test						X	U	U	U	U	I	I		
7	Auto test CBC							X	U	U	U	I	I		
8	Auto blood smearing								X	A	U	I	U		
9	Auto tint										X	I	U		
10	Blow slide											X	I		
11	Microscope scanning													X	
12	Verifying														X

Defination
 A :Absolutely Necessary
 E :Especially Important
 I :Important
 O :Ordinary Closeness
 U :Unimportant
 X :Undesirable

4. PERFORMANCE MEASURES

The simulation run set ups are as follows: 400 replications, the run length of 24 hour and no warm up period. Tables 5 and 6 shows the comparison of simulation results under the current layout and the proposed layout. Under the proposed layout, the transfer times are significantly reduced. Turnaround time and total waiting time in queue decrease because the work flow of MT is changed. As is, after the MT obtains the microscope results, she writes it on a sheet of paper and leaves it in a basket for another MT who keys a batch of results into the computer system. We propose that MT who reads the microscope subsequently input the results into computer system right away, not placing it in a batch.

The simulation model returns low utilization of human resources because these numbers reflect the fraction of times they actually work on specimens. In reality, they do not have that much free time since they also perform other duties, such as filling out paper work or talking on the phone.

Table 5: Comparison of the Current Layout and the Proposed Layout

Time	Current layout (minute)	New lab layout and work process (minute)	Percent decrease
Turnaround time	72.07 ± 0.9	57.03 ± 1.12	21%
Total waiting time in queue	44.27 ± 3.2	30.00 ± 0.98	32%
Transfer time	3.37 ± 0.1	0.28 ± 0.2	92%

Table 6: Utilization of Resources (%) in CBC Test

Resources	Old Layout	New Layout
AMT in the 1 st shift	25 ± 1.0	26 ± 1.5
AMT in the 2 nd shift	60 ± 0.1	65 ± 1.1
MT in the 1 st shift	12 ± 0.7	18 ± 2.3
MT in the 2 nd shift	22 ± 0.1	29 ± 0.3
MT at microscope station in the 1 st shift	8 ± 0.8	12 ± 1.0
MT at microscope station in the 2 nd shift	49 ± 1.0	60 ± 0.1

5. CONCLUSION AND DEVELOPMENT APPROACH

We show that our simulation model can adequately represent the actual system. We use it to test our ideas of reducing turnaround time by modifying the lab layout to decrease the walk time of staff. We use relationship diagram for create new lab layout. The turnaround time of new lab layout decreases from 72.07 ± 0.9 minute to 57.03 ± 1.12 minute or 18.4% with the CBC test capacity of 615 specimens per day. Reducing the turnaround

time will increase CBC test's capacity to meet the demands of increased patient volumes.

ACKNOWLEDGEMENTS

The simulation model used in this paper has previously appeared in the Proceedings of the 4th International Conference on Applied Operational Research (ICAOR). This research is financially supported by the Office of the Higher Education Commission and Mahidol University under the National Research Universities Initiative.

REFERENCES

- Ahmed, A. and Alkhamis, M., 2008. Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research*, 198, 936–942.
- Albrecht, MC. Available from: <http://www.albrechts.com/mike/DES/index.html> [accessed 8 March 2012]
- Banks, J., Carson, J.S., Nelson, B.L. and Nicol, D.M., 2005. *Discrete-event System Simulation*. 4th ed. U.S.A.: New Jersey.
- Blasak, R.E., Armel, W.S., Starks, D.W. and Hayduk, M.C., 2003. The use of simulation to evaluate hospital operations between the emergency department and a medical Telemetry unit. In: S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, ed. *Proceedings of the 2003 Winter Simulation Conference*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 1887–1893.
- Deb, S.K., and Bhattacharyya, B., 2005. Fuzzy decision support system for manufacturing facilities layout planning. *Decision Support Systems*, 40, 305–314.
- Department of International Trade Promotion. Available from: <http://www.ditp.go.th/Default.aspx> [accessed 1 September 2011]
- Duguay, C. and Chetouane, F., 2007. Modeling and Improving Emergency Department Systems Using Discrete Event Simulation. *Simulation*, 83, 311–320.
- Muther, R., 1961. *Systematic Layout Planning*. Industrial Education Institute. Boston.
- Perkliang, M., San-ae, S., Chawaku, D., Rutti, T., Prasongsab, T., Pocathikom, A. 2010. Root Cause Analysis of Specimen Rejection for Hematology Testing. *Songkla Medical Journal*, 28, 267-274.
- Pirollo, J., Ray, A., Gadzinski, M., Manese, M., Garvert, B., Scoville, G., Walpole, H., Amland, B., Boos, R., Mamminga, I., Brown, J. and Donlon, K., 2009. Utilization of Discrete event simulation in the Prospective Determination of Optimal Cardiovascular Lab Processes. In: M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, ed. *Proceedings of the 2003 Winter Simulation Conference*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 1916-1926.

- Proctor, T., 1996. Simulation in Hospitals. *Health Manpower Management*, 22, 40-44.
- Rossetti, M.D. 2010. *Simulation Modeling and Arena*. John Wiley & Sons Ltd. United States of America.
- Sinreich, D., and Marmor, Y., 2005. Emergency department operations the basis for developing a simulation tool. *IIE Transactions*, 37, 233–245.
- Swisher, J., Jacobson, S., Jun, B. and Balci, O., 2001. Modeling and Analyzing a Physician Clinic Environment Using Discrete-Event (Visual) Simulation. *Computers and Operations Research*, 28, 105-125.
- Venkatadri, V., Raghavan, V.A., Kesavakumaran, V., Lam, S.S. and Srihari, K. 2011. Simulation based alternatives for overall process improvement at the cardiac catheterization lab. *Simulation Modeling Practice and Theory*, 19, 1544-1557.

AUTHORS' BIOGRAPHIES

Luangmul K. is a Research Assistant in the Department of Industrial Engineering, Kasetsart University, Bangkok, Thailand. She earned Bachelor of Engineering (Industrial Engineering) from Chiang Mai University, Thailand.

Pichitlamken J. is an Assistant Professor in the Department of Industrial Engineering, Kasetsart University. Her research interests include ranking and selection procedures, simulation optimization, spreadsheet simulation, and stochastic processes.

Weerawat W. is an Assistant Professor in the Department of Industrial Engineering, Mahidol University. Her research interests include discrete-event simulation, logistics, and healthcare.

A NEW HYBRID ALGORITHM BASED ON WATERSHED METHOD, CONFIDENCE CONNECTED THRESHOLDING AND REGION MERGING AS PREPROCESSING FOR STATISTICAL CLASSIFICATION OF GENERAL MEDICAL IMAGES

Gerald Zwettler^(a), Werner Backfrieder^(a,b)

^(a)Bio- and Medical Informatics, Research and Development Department,
Upper Austria University of Applied Sciences, Austria

^(b)School of Informatics, Communication and Media, Upper Austria University of Applied Sciences, Austria

^(a)gerald.zwettler@fh-hagenberg.at, ^(b)werner.backfrieder@fh-hagenberg.at

ABSTRACT

Segmentation of morphology in medical image data is a highly context specific and differs from various imaging modalities, necessitating the use of sophisticated mathematical models and algorithms to achieve good results. In this work an algorithm is presented for pre-segmentation of general medical input data, based on a watershed-segmentation strategy utilizing both, original intensities and derived gradient magnitudes for region growing. The number of resulting pre-classified regions is iteratively reduced to a user-defined threshold using merge metrics, accounting for the similarity of intensity profiles of two neighboring regions to merge, as well as the height of the gradient barriers to overcome and geometric aspects like sphericity and size of the border area with respect to the total region size. Based on such a context-independent pre-segmentation, the resulting manageable number of regions can be further merged and classified, utilizing texture features and a priori statistical models. Results are presented from brainweb database.

Keywords: watershed-segmentation, statistical image classification, texture features

1. INTRODUCTION

Accurate and automated segmentation of medical image data is of high importance in a very broad range of medical applications. The evaluation of geometric properties like position, size and extent of anatomical structures necessitates previously performed image segmentation and is of high relevance in the domain of surgery planning as well as disease progression. Based on such an anatomical classification, e.g. in case of liver tumors, the position of the lesion with respect to the supplying vessel systems as well as the parenchyma size and shape can be assessed prior to liver lobe resection (Zwetter, Backfrieder, Swoboda, and Pfeifer 2009). When combining data from different imaging modalities like CT/MRI for high anatomical resolution and SPECT/PET from the functional imaging domain, metabolic activity can be quantified utilizing segmentation masks of the corresponding anatomical

region (Beyer, Schwenzer, Bisdas, Claussen, and Pichler 2010). Evaluation of therapy success and disease progression gets feasible in a quantitative way if morphological segmentations of the same patient at different points in time are available (Kuhnigk, Dicken, Bornemann, Bakai, Wormanns, Krass, and Peitgen 2006). Furthermore, in virtual reality scenarios, patient-specific segmentation models based 3D tracking and navigation, true 3D vision and rapid prototyping of haptic patient models (Wulf, Vitt, Gehl, and Busch 2001; Torres, Staskiewicz, Sniezynski, Drop, and Maciejewski 2001) are highly applicable for the task of surgical training and planning (Stone 2011).

To achieve accurate segmentation of particular organs and anatomical structures, semi-automated concepts like region growing (Gonzalez and Wintz 1987) or live-wire contour definition (Barrett and Mortensen 1997) can be applied. Due to the high demand for user-interaction these concepts are improper for clinical use. Although requiring rarely no a priori knowledge and thus favoring generic segmentation, utilizing these strategies only rather awkwardly shaped structures with homogenous intensities can be segmented. Although not directly applicable for medical applications, these concepts are of high relevance preparing manual reference segmentations for model-driven segmentation approaches.

Utilizing deformable models (McInerney and Terzopoulos 1996) and incorporating a priori knowledge, particular structures with low anatomical variability can be precisely segmented. The generic use of deformable models for segmentation of arbitrary anatomical structures is not practicable, because the model parameters need proper adjustment to the target morphology and fine structures showing inhomogenous intensities remain hard to segment. Statistical Shape Models (Cootes, Taylor, Cooper, and Graham 1992) can be automatically derived from a set of reference segmentations, allowing for the use of geometric features of the target structures to segment. If trying to model several structures utilizing statistical shape models, the problem of shape overlapping as well as topological changes remain unconsidered. Furthermore,

modeling very thin and sparsely connected structures with respect to low imaging resolution or partial volume effects cannot be achieved with geometric shape modeling. Active appearance models (Cootes, Edwards, and Taylor 1998) introduce statistical properties of the targets structure expected intensity profile besides geometric features, thus being more robust in case of anatomical variability. Segmentation of small, loosely connected structures and changes in topology remain still unaddressed. Utilizing level sets (Osher and Sethian 1988), topological changes and anatomical variability can be handled, but adapting the steering parameters curvature, propagation and advection, only rather compact structures can be modeled. Furthermore, the level set parameters must be adjusted for each structure to segment, rather than deriving them from a set of reference segmentations.

To facilitate generic segmentation of arbitrary anatomical structures, a context and modality invariant pre-segmentation is required as fundament for texture based merging for classification. Classifiers like fuzzy connectedness (Gammage and Chaudhary 2006) and confidence connected thresholding (Ibanez, Schroeder, Ng, and Cates 2003) necessitate initial seed points for separating clusters of homogenous intensity distributions. For application of k-means clustering (Kanungo, Mount, Netanyahu, Piatko, Silverman, and Wu 2002), no seeds are required but regions of the expected k clusters are defined according to expected constant intensity profiles within the entire input image volume.

In this work we present a generic concept for pre-segmentation on arbitrary medical image input data. Starting at local minima positions in a gradient magnitude representation of the image volume, regions are grown similar to the watershed algorithm (Vincent and Soille 1991; Beare and Lehmann 2006). To overcome limitations of the watershed algorithm along homogenously and slightly increasing areas, the input image is incorporated during the region growing process. The initially classified regions (catchment basins) are not merged solely due to watershed level tolerance, but a metric, also considering region neighborhood, geometric properties and similarity of intensity profiles. Thus, an arbitrary input image can be pre-classified independent of context and imaging modality at a user-specified number of target regions, defining granularity of the pre-processing step. Based on this pre-segmentation, the final classification can be performed, utilizing statistical feature values automatically extracted from the set of reference segmentations.

2. DATA

For testing of generic pre-segmentation, a number of $n=20$ T1-weighted MRI datasets from the simulated *brainweb* database (Cocosco, Kollokian, Kwan, and Evans 1997; Kwan, Evans, and Pike 1999) and associated reference segmentations are used.

Further test runs and validations are performed utilizing $n=12$ anonymous multi-modal patient studies, comprising morphologic image acquisitions (T1, T2, PD, ...) as well as functional imaging (SPECT, PET). For patient data, the reference segmentations are achieved in a semi-automated way by applying image processing pipelines modeled and evaluated using the MeVisLab image processing platform (Ritter 2007; MeVis 2012). The image segmentation pipeline comprises filtering, region growing, mathematical morphology, image arithmetics and live-wire contour definition, see chapter 3.6 for more details.

3. METHODOLOGY

For the pre-segmentation strategy, in a first step input image data is filtered to smooth the intensity topography and suppress noise and artifacts. Later, the gradient magnitude is extracted as derivation of the original intensities for balancing differences in intensity level and reducing region definitions to their boundaries. Based on the gradient magnitude image, local minima regions are detected and enlarged utilizing region growing. Local minima forming autonomous regions are iteratively dilated with respect to a stepwise increased water surface. Finally for the grown regions image statistics are calculated, essential for pair-wise region merging until the number of pre-segmented regions falls below the target region count. For illustration of the processing chain, see Fig. 1.

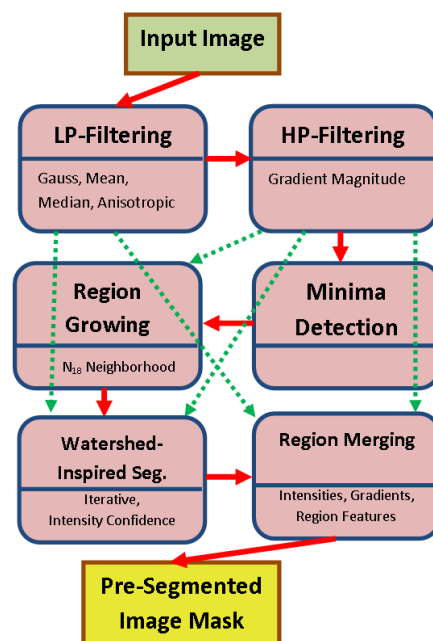


Figure 1: The sequential process chain (solid red arrows) comprises low-pass filtering, high-pass filtering, minima detection region growing, the watershed-type segmentation procedure and finally region merging until the convergence criterion is reached. Filtered input image and the gradient magnitude representation are required as input for several particular process steps (dashed green arrows).

3.1. Filtering

As all local minima of input image derived gradient magnitude representation result in an additional initial region to grow, their number has to be limited by applying some lowpass filtering for smoothing the intensities. Limiting the number of autonomous regions prior to the final merge procedure is essential with respect to runtime and memory resources.

For lowpass filtering classic filter methods like mean, Gaussian and the median rank filter are evaluated. As the image intensity smoothing should not be performed at the expense of weakened edges and reduced inter object contrast, diffusion filtering has to be applied. The applied filters are taken from the ITK image processing library (Ibanez, Schroeder, Ng, and Cates 2003).

The filter configuration discussed in the following section refers to an input image with iso-voxel spacing of $1mm$ in each dimension. In case of data with different spacing configuration, additional noise or artifacts to handle, an individual additional pre-processing is required to prepare the data at sufficient quality and with properties needed.

The major smoothing and fine tuning is accomplished applying several diffusion filter runs. Modules `itk::GradientAnisotropicDiffusion` and `itk::CurvatureAnisotropicDiffusion` are sequentially applied with parameterization $(6;0.125;10.0)$ and $(3;0.125;10.0)$ for *iteration*, *time-step* and *conductance* value. Utilizing both, gradient and curvature preservation aspects in diffusion equation, leads to balanced results, compared to an increased number of iterations for only one diffusion filter type in the image processing chain.

Besides the diffusion filtering, median filter module `itk::MedianImageFilter` is utilized with *radius 1* at the end of the filtering pipeline to minimize local outliers, preserved during diffusion filtering.

3.2. Gradient Magnitude Calculation

Besides the input image intensities, a gradient representation indicating closed borders between neighboring separated regions is required.

For choice of the gradient calculation strategy, impermeability between two neighboring regions to separate is the most important criterion. Impermeability is achieved, if the gradient values of an enclosed region don't show any kind of low gradient value by-pass connections to neighboring regions. In case of smooth intensity changes, only gradients of low magnitude remain to separate the regions, see Fig. 2. The weakest gradient borders delimit applicability of gradient-based segmentation strategies. To check the level of impermeability, the possible flood-fill level for well defined regions is evaluated. For details please refer to the results section.

The `itkGradientMagnitudeImageFilter` filter algorithm is applied for deriving the edge representation from input image. Gradient magnitude values are calculated at floating-point precision. As

discrete values are required for further merging neighboring pixels of same value into larger regions, the gradient magnitude is normalized to gradient range $[0;maxVal]$, with $maxVal=200$ and rounded to integer type. This normalization ensures a constant level of segmentation granularity, independent from pixel type of input image data.

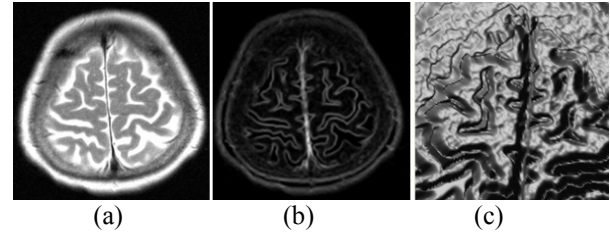


Figure 2: Transversal brain slice (a) and calculated gradient magnitude (b). Although brain windings are mostly enclosed by high gradient magnitude values, the smooth intensity decline in ventral direction leads to smaller gradients along the borders, as illustrated utilizing a depth profile (c).

3.3. Local Minima Detection

Each local gradient minimum in N_{26} is a possible starting position of a new region to be grown in the course of watershed-type pre-segmentation. For local minimum detection, all pixels in the gradient magnitude representation of the input image are evaluated with respect to their direct neighbors. If all neighbors show equal or larger values, the current position is marked as local minimum, see Equ. 1.

In areas of homogeneity gradient values, an over-classification of local minima is inevitable. Consequently, the marked positions are only region start positions for following initial region growing and candidates to form up a new region. The number of regions to grow is therefore significantly smaller compared to the total number of detected minima.

$$LM(p) := \forall q \in N_{26}(p) | I(q) \geq I(p) \quad (1)$$

3.4. Initial Region Growing

The detected local minima positions are seed point candidates for growing regions of identical gradient magnitude utilizing N_{18} neighborhood.

Besides gradient magnitude as region growing criterion, similarity in filtered intensities is incorporated as second aspect for an adjacent voxel to be added to a particular region. The intensity topography must be accounted for, as regions of same gradient magnitude can show huge differences in intensity values due to marginal but steady changes.

A threshold T is defined for the acceptable difference in intensity of a region to be grown with respect to matching gradient magnitude values. A voxel v can be added to an adjoined region R , if intensity limits after union are still within the defined tolerance interval T .

$$\text{grow}(v, R) := \begin{matrix} N_{18}(v, R) \wedge \text{grad}(v) == \text{grad}(R) \wedge \\ \left(|\max I(v \cup R) - \min I(v \cup R)| \leq T \right) \end{matrix} \quad (2)$$

Due to incorporation of the input image intensities, large differences in intensities of a grown region can be prevented.

In arbitrary medical image data, the variance of local intensities is expected to be quite inhomogeneous, e.g. lower variance σ for background intensity distribution compared to objects and structures belonging to the patient's anatomy. Standard deviation of local intensities can be calculated for all voxel positions, thus the generation of a standard deviation map, indicating image areas with slightly homogenous voxel values and areas with high variance, is performed. To balance these local differences in the intensity profile of the entire image I , an adaptive region growing algorithm will segment the initial regions based on a tolerance interval T_R , individually calculated and updated for each region R .

$$\begin{aligned} \text{aGrow}(v, R) &:= \begin{matrix} N_{18}(v, R) \wedge \text{grad}(v) == \text{grad}(R) \wedge \\ \left(|\max I(v \cup R) - \min I(v \cup R)| \leq T(r) \right) \end{matrix} \quad (3) \\ T(R) &:= T \cdot \left(\frac{1}{2} + \frac{\text{meanStdDev}(R)}{2 \cdot \text{meanStdDev}(I)} \right) \end{aligned}$$

The standard deviation value for each pixel is calculated as mask operation with radius $r=5$, incorporating all neighboring voxels at a maximum Euclidean distance of r . Results are smoothed applying an $11 \times 11 \times 11$ average kernel. With this adaptive region growing strategy, similar granularity of pre-segmented regions can be accomplished, balancing local variation in intensity standard deviation.

3.5. Hybrid Watershed-Inspired Pre-Segmentation

Based on the regions resulting from initial region growing, the tolerance level for gradients is iteratively increased, thus processing the border elements of the regions and growing neighbors with a matching gradient value. Only voxels with a gradient value of exactly the particular region start gradient value (from the local minima) plus the current gradient increment are candidates for growing. The gradient tolerance level is increased when none of the regions can grow at the current gradient tolerance level. The described algorithm refers to classic watershed segmentation and is labeled as GRG_WS.

To stronger incorporate the intensity features, the grow conditions described in Equ. 2 and Equ.3 are also applied for growing the regions, resulting in a static intensity interval segmentation strategy (SII_WS) and an adaptive intensity interval segmentation strategy (AII_WS), see Code Listing 1.

Introducing the new intensity conditions for the growing process, it is not assured that all voxels can be classified. Remaining voxels are iteratively assigned to the neighboring region with lowest difference in

intensity. The intensity tolerance limit is set to 2 and is temporarily increased, whenever in the course of one full iteration, not a single voxel can be assigned to one of the existing regions, see Code Listing 2.

```
regions = initialRegionGrowing;
int gradientTolerance = 1;
while(gradientTolerance < maxGradVal)
  while(stillChanges)
    for(region r : regions)
      for(voxel u : r.border)
        for(voxel v : u.N18)
          if(unclassified(v) &&
             grad(v) == (r.initGrad
                          + gradientTolerance) &&
             aGrow(v, R))
            r.addToRegion(v);
            v.classVal = r.regID;
            r.addToBorderVect(v);
            stillChanges = true;
gradientTolerance++;
for(region r : regions)
  r.updateBorderVect(gradientTolerance);
```

Code Listing 1: Illustration of the watershed-inspired pre-segmentation algorithm based on adaptive intensity intervals (AII_WS).

```
int intensityTolerance = 2;
while(unclassifiedVoxel.size > 0)
  bool changes = false;
  for(voxel u : unclassifiedVoxel)
    for(voxel v : u.N18)
      region r = region(v);
      if(r != NULL &&
         insideTolerance(u.intVal, r,
                         intensityTolerance))
        r.addToRegion(u);
        u.classVal = r.regID;
        r.updateIntToleranceInterval();
        changes = true;
  if(changes)
    intensityTolerance = 2;
  else
    intensityTolerance += 2;
```

Code Listing 2: Algorithm for iteratively assigning the voxels remaining unclassified during AII_WS to a neighboring region.

3.6. Calculation of Region Features for Merging

Based on the results of the pre-segmentation, the number of regions has to be significantly reduced applying region merging strategies. For classic watershed algorithm, a tolerance value for differences in the starting gradient value of neighboring regions is the key criterion for merging two separated regions u and v into one. This concept lacks steering of the resulting number of regions to remain.

For the presented intensity interval and adaptive intensity interval pre-segmentation strategy, besides the gradient value $gradW(u,v)$, also differences in the regions intensities as $intW(u,v)$ and length of the borders as geometric property $borderW(u,v)$ and indicator for sphericity are taken into account, with $surf(u,v)$ as voxel count on the border between regions u and v and $surf(u)$ as the total count of neighboring region voxels of a region u .

$$\text{intW}(u, v) := \frac{(\text{meanInt}(u) - \text{meanInt}(v))^2}{\text{maxIntTolerance}}, \quad (4)$$

$$\text{maxIntTolerance} = \frac{(\text{maxInt}(I) - \text{minInt}(I))}{10}$$

$$\text{gradW}(u, v) := \frac{\max \left(\begin{array}{l} \text{mean}G(u, v) - \min G(u), \\ \text{mean}G(u, v) - \min G(v) \end{array} \right)}{\text{maxGradTolerance}}, \quad (5)$$

$$\text{maxGradTolerance} = \frac{(\text{max}G(I) - \text{min}G(I))}{10}$$

$$\text{borderW}(u, v) := \min \left(\begin{array}{l} \frac{\text{surf}(u) - \text{surf}(u, v)}{\text{surf}(u)}, \\ \frac{\text{surf}(v) - \text{surf}(u, v)}{\text{surf}(v)} \end{array} \right) \quad (6)$$

Thus, for every pair of neighboring regions u and v connected at a border b , the discussed features can be evaluated and a numerical merge feature value $\text{mergeW}(u, v)$ can be evaluated, thus indicating similarity and weight for merging regions u and v . The weighting factors $w_g=2$, $w_i=7$ and $w_b=1$ lead to good results as combining the different aspects and outperforming each single criterion.

$$\text{mergeW}(u, v) := w_g \cdot \text{gradW}(u, v) + w_i \cdot \text{intW}(u, v) + w_b \cdot \text{borderW}(u, v). \quad (7)$$

Based on a sorted and steadily updated list of merge values, the total region count can be reduced by particularly merging the region pair showing the lowest merge value, until the target number of R remaining regions is reached.

3.7. Semi-Automated Reference Segmentation

For semi-automated generation of reference segmentations, MeVisLab (MeVis 2012) rapid prototyping platform is utilized. Particular anatomical structures are segmented using the LiveWire module (Handels 2009; Barrett and Mortensen 1997) if the shape is well defined and RegionGrowing module is applied for segmentation of smaller structures lacking insularity. Besides the segmentation modules, anisotropic filtering for initial smoothing and morphological post-processing (opening/closing utilizing Morphology) in case of artifact occurrence are additional process steps in the image processing pipeline.

In contrast to brainweb reference segmentations, not the entire dataset will be semi-automatically segmented. For each imaging modality and test case, only particular anatomical structures, like gray matter, white matter and ventricle for brain MRI data, are chosen for algorithmic validation.

4. RESULTS

The presented watershed-inspired pre-segmentation strategy is tested utilizing the first $n=6$ samples and associated reference segmentations of the brainweb database. The image data is scaled to 8bit and iso-voxel spacing at size $256 \times 307 \times 256$ using *sinc*-window-based Lanczos 3 interpolator (Burger and Burge 2008).

4.1. Evaluation of the Filter Chain

At first the filter chain is tuned with respect to the filter types and parameterization. As one quality criterion, the image mean intensity level should be kept constant. The smoothing effect can be estimated, evaluating changes on the voxel intensity's standard deviation and the average voxel change. Besides noise reduction, the key goal of this filtering process is to balance inhomogeneity in local voxel neighborhood and thus, significantly reducing the number of initial regions to be detected during the first region growing run. Finally, the classification result quality after executing the filter chain is evaluated utilizing the reference segmentations. A comparison of different filter chain settings is given in Table 1.

Based on the low number of initial regions, a small difference in mean intensity and preferably high precision, the chains Ga10 (0.01;145,622;0.98) and Ga6Ca3M1 (0.16;119,805;0.966) are best balancing the criterions, as charted in Fig. 3.

Table 1: Evaluation of different filter chains, utilizing average (A), median (M), gradient anisotropic diffusion (Ga), curvature anisotropic diffusion (Ca) and Gaussian filtering (G) at different kernel sizes.

filter chain	ΔI	σI	ΔI_{mean}	#reg	prec
unfiltered	0.00	0.00	0.00	899,386	1.000
A2	6.05	8.28	0.50	148,393	0.974
M1A1	4.39	5.93	0.70	222,375	0.987
A3	7.66	10.76	0.51	119,767	0.952
Ga10	2.46	2.23	0.01	145,622	0.980
Ca10	8.95	34.07	5.09	898,315	0.982
Ga6Ca3M1	3.35	3.95	0.16	119,805	0.966
G2	6.97	9.71	0.51	110,991	0.964
G1Ga6Gc2	5.59	7.39	0.52	96,122	0.937

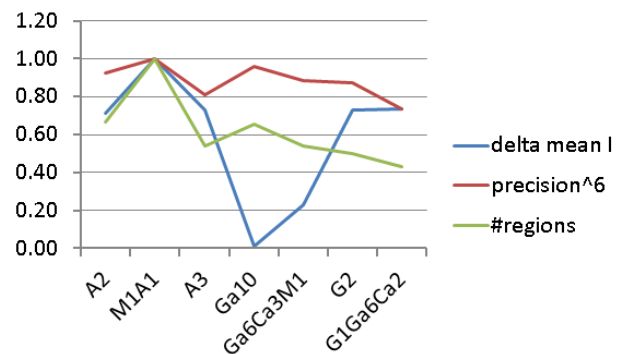


Figure 3: Charting of the three key features ΔI_{mean} , #regions and precision^6 as ratio with respect to filter chain M1A1. Despite precision, a low feature value is desired.

With the discussed filter chain parameterization Ga6Ca3M1 smoothing of local intensity homogeneity can be achieved but borders and gradients are preserved, see Fig. 4.

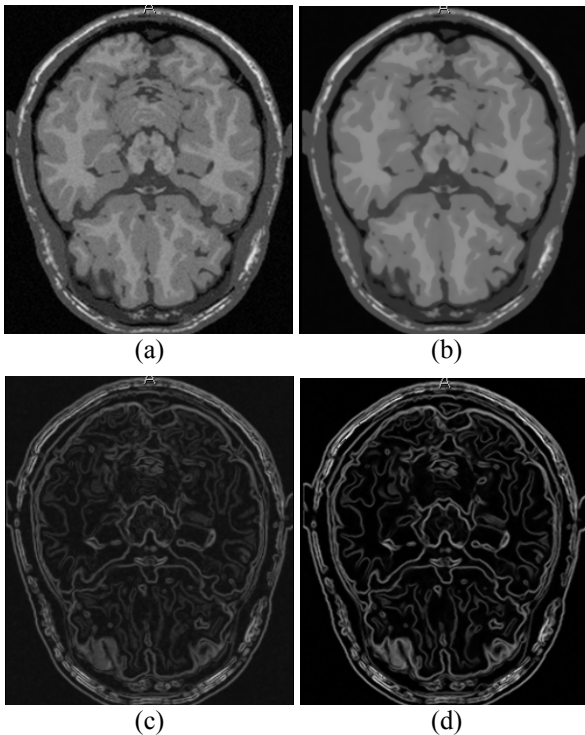


Figure 4: Input image intensities (a), slice #76 of *subject04*, are significantly smoothed applying the filter chain (b). Furthermore, results of gradient magnitude calculation after smoothing (d) show less variance due to suppressed noise compared to gradient magnitude calculation on the original image data (c).

4.2. Tests on Initial Region Growing

Region growing is performed based on the gradient magnitude representation after processing the filter chain. For common watershed segmentation, neighbors in N_{18} with identical gradient value are grouped together. N_{18} adjacency demands two neighboring voxels to share at least one edge and thus excludes the eight corner diagonal members of N_{26} . Focusing on the gradient values leads to high variability in intensities, see Table 2. With our presented strategies for initial region growing, intra-region intensity homogeneity can be enforced.

For calculation of the voxel standard deviation map to be used for adaptive region growing (AII), a radius of $r=5$ with subsequent $11 \times 11 \times 11$ average kernel application is sufficient for handling local intensity aspects, see Fig. 6.

As presented in Table 2, the restrictions on the intensity range for initial region growing significantly improve quality of the results. Despite a slightly increased number of regions, configuration AII4 best balances the different criterions, see Fig. 5. Enforcing regions to show identical gradient and intensity values (SII0) does not lead to improved results but an

increased number of regions to handle in the next processing steps.

Table 2: Comparison of gradient-based region growing (GRG) with static intensity interval restriction (SII) and adaptive intensity interval restriction (AII) at different interval sizes. Maximum inner region intensity difference ($\max \Delta I$) and mean region intensity difference ($\text{mean } \Delta I$) indicate homogeneity of intensities. The grow ratio reflects, how many voxels are processed during this first classification run.

RG strategy	$\max \Delta I$	$\text{mean } \Delta I$	#reg	grow ratio	prec
GRG	151	0.703	119,805	0.592	0.966
SII6	6	0.458	123,056	0.589	0.967
SII4	4	0.405	134,569	0.584	0.968
SII0	0	0.000	196,984	0.542	0.972
AII6	15	0.469	125,495	0.585	0.970
AII4	10	0.401	138,323	0.592	0.972

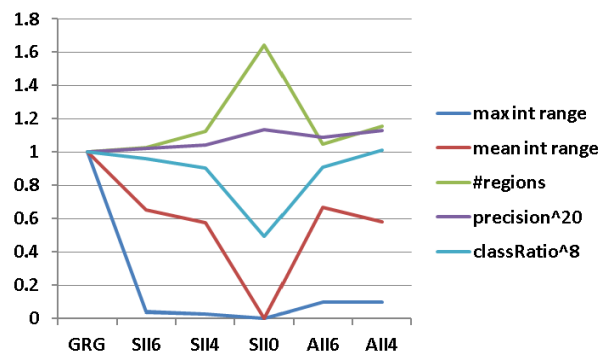


Figure 5: Charting of the features from Table 2 as ratio with respect to region growing strategy GRG. Despite precision and the classification ratio, a low feature value is desired.

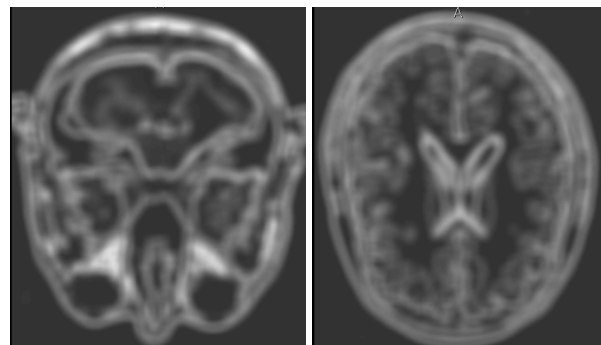


Figure 6: Two slices of the calculated standard deviation map at Euclidean radius $r=5$ with subsequent $11 \times 11 \times 11$ average filter application. Results are smooth enough to adequately reflect local intensity variation.

4.3. Testing Watershed-Inspired Pre-Segmentation

Based on the results of initial region growing for GRG, SII6 and AII4, the results after classification of all voxels are analyzed, see Table 3. Both, static intensity interval and adaptive intensity interval significantly improve results compared to common gradient-based watershed segmentation. Furthermore, the mean inter-

region intensity range (mean ΔI) is reduced by a factor of 2.0, relevant for subsequent neighbor region merge process. The key criterions precision and mean inter-region intensity range can be both improved compared to GRG_WS, see Fig. 7.

Table 3: Comparison of common gradient-based watershed segmentation (GRG_WS) with the newly developed concepts for static intensity interval restriction (SII6_WS) and adaptive intensity interval restriction (AII4_WS). Maximum inner region intensity difference (max ΔI) and mean region intensity difference (mean ΔI) indicate homogeneity of intensities. The grow ratio reflects, how many voxels are processed during this first classification run.

RG strategy	max ΔI	mean ΔI	#reg	grow ratio	prec
GRG_WS	215.5	21.974	119,955	1.00	0.883
SII6_WS	166.2	10.818	123,056	0.89	0.910
AII4_WS	168.2	10.768	130,466	0.87	0.920

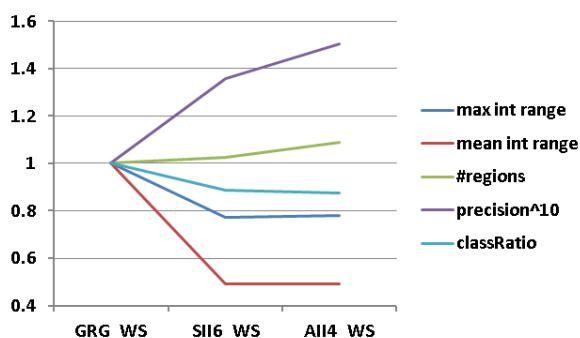


Figure 7: Charting of the features from Table 3 as ratio with respect to GRG_WS watershed segmentation. High precision and a low value for both, mean and max intensity range are highly recommended.

Evaluating results in detail, GRG_WS is outperformed with respect to precision for each particular dataset. Furthermore, SII6_WS is outperformed by the adaptive AII4_WS for each particular brainweb sample, see Table 4 for the first $n=6$ samples.

Table 4: Precision and mean ΔI for the first $n=6$ brainweb datasets. Precision of SII6_WS higher compared to GRG_WS and precision of AII4_WS higher compared to SII6_WS. Same for decrease in mean ΔI , exempt from sample [5]sj20.

data	GRG WS		SII6 WS		AII4 WS	
	mean ΔI	prec	mean ΔI	prec	mean ΔI	prec
[1]sj04	22.1	0.881	11.0	0.910	10.9	0.915
[2]sj05	20.6	0.888	10.4	0.915	10.3	0.924
[3]sj06	21.3	0.849	11.0	0.883	10.6	0.908
[4]sj18	24.9	0.885	11.7	0.913	11.7	0.920
[5]sj20	21.9	0.904	10.3	0.928	10.5	0.931
[6]sj38	21.1	0.889	10.6	0.913	10.6	0.921

Results after AII4_WS are presented in Fig. 8. Regions showing rather related region labels are visualized in similar colors. Due to a still large number of regions ($>100,000$), the coloring reflects a trend of growing IDs from left to right according to the processing order.

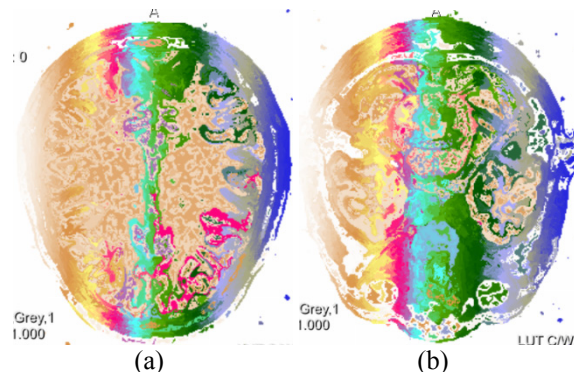


Figure 8: Color representation of regions resulting from AII4_WS. The MRI morphology is cognizable for slices #146 (a) and #55 (b) of dataset [1]sj04.

Segmentation errors produced by GRG_WS and AII4_WS are compared in Fig. 9. Most of the wrongly classified voxels result from filtering effects and deficits of the reference segmentation, like differentiating between air outside and inside the head, see Fig. 10.

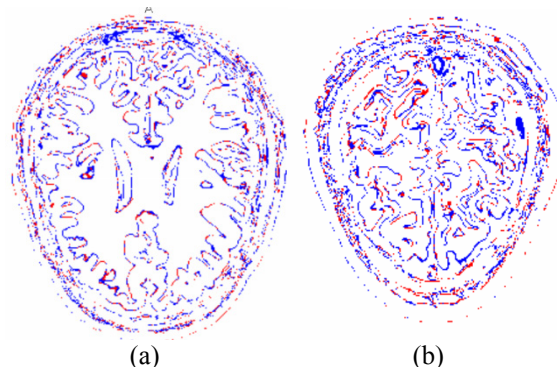


Figure 9: Color coded differences of errors produced from GRG_WS and AII4_WS for slices #140 and #182 of dataset [1]sj04. Discrepancies only appearing for GRG_WS shown in blue and those only resulting from AII4_WS in red.

4.4. Testing Calculation of Merge Features

Tests show that in contrast to classic watershed segmentation approach, the target number of remaining regions can be precisely pre-defined by the user.

Due to combination of different types of features, the achievable precision can be significantly increased, see Table 5 and Fig. 11. For common watershed segmentation only the tolerance in gradient level at the region borders is utilized as merge criterion. Results in Table 5 show that the watershed region merge metric is outperformed by incorporating intensity level and border ratio.

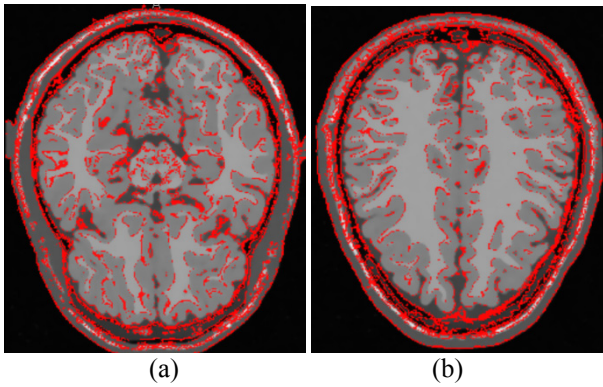


Figure 10: Error (red) of AII4_WS with respect to the original image data at slices #70 and #155 of dataset [1]sj04.

Table 5: Comparison of result quality utilizing different types of merge features. Analysis runs are performed based on AII4_WS pre-segmentation results. For the cumulated weight defined in Equation 7, the weights for intensity level similarity (squared diff in intensities), difference in gradient level and the border ratio are combined, thus leading to increased precision. The difference in mean region gradient level (squared diff in gradients) showed to be no proper metric for region merging.

merge strategy	#reg= 2500	#reg= 5000	reg= 7500
random (region IDs)	0.474	0.458	0.477
squared diff in intensities	0.809	0.803	0.836
squared diff in gradients	0.514	0.497	0.517
difference in gradient level	0.735	0.736	0.766
border ratio	0.622	0.603	0.629
cumulated weight	0.862	0.831	0.866

5. DISCUSSION

The incorporation of original image intensities in the process of region growing and gradient-based watershed-like segmentation showed to improve results significantly.

Due to the fact that the number of local gradient minima is low in anatomically relevant areas of high pixel intensity variance, further region seed candidates will be evaluated in future. Besides the gradient values, local minima will also be searched for, utilizing the input image intensity voxel mask. It is thereby expected to increase the number of region seed candidates by around 20-30%, mainly in the sparse area of high variance.

Furthermore the algorithm for iteratively assigning the voxels remaining unclassified during AII4_WS algorithm execution will be improved. If unclassified voxels show higher similarity to their neighbors being unclassified too, additional regions should be formed to keep the intra-region intensity range low.

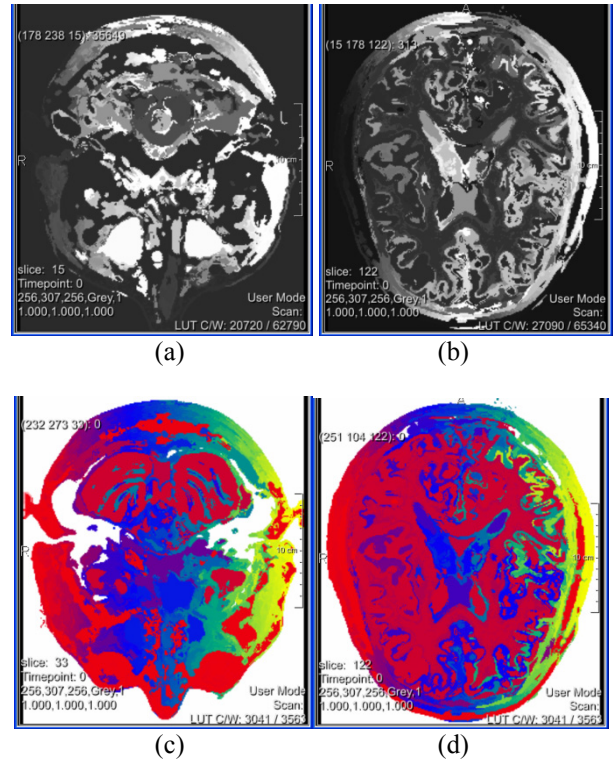


Figure 11: Final merged regions of dataset [1]sj04 after AII4_WS pre-segmentation and region merging utilizing cumulated metrics. Visualization with gray value colormap (a-b) and in color (c-d). Compared to Fig. 8, the color trend from left to right is replaced by solid and well defined regions representing anatomical structures.

ACKNOWLEDGMENTS

Thanks to our medical partners from the Wagner-Jauregg state mental hospital, Linz, Upper Austria, at the institute for neuro-nuclear medicine headed by Primarius Dr.Dr. Robert Pichler for providing medical image data and for valuable discussion.

This research is part of the INVERSIA project (<http://inversia.fh-linz.at>) which was funded by the European Regional Development Fund (ERDF) in cooperation with the Upper Austrian state government (REGIO13).



REFERENCES

- Barrett, W.A., Mortensen, E.N., 1997. Interactive Live-Wire Boundary Extraction. *Medical Image Analysis* 1:331-341.
- Beare, R., and Lehmann, G., 2006. The watershed transform in ITK – discussion and new developments. *Insight Journal*, January – June 2006.
- Beyer, T., Schwenzer, N., Bisdas, S., Claussen C.D., and Pichler, B.J., 2010. *MR/PET – Hybrid Imaging for the Next Decade. MAGNETOM Flash* 3/2010.

- Burger, W., and Burge, M.J., 2008. *Digital Image Processing – An Algorithmic Introduction Using Java*. New York: Springer, 402pp.
- Gonzalez, R.C., Wintz, P., 1987. *Digital Image Processing*. London: Addison Wesley.
- McNerney, T., Terzopoulos, D., 1996. Deformable Models in Medical Image Analysis : A Survey. *Medical Image Analysis* 1 (2):91-108.
- Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Evans, A.C., 1997. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage* 5(4):425.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1992. Training Models of Shape from Sets of Examples. *Proceedings of the British Machine Vision Conference*, 9-18. September 22-24, Leeds (United Kingdom).
- Cootes, T.F., Edwards, G.J., Taylor, C.J., 1998. Active Appearance Models. *Proceedings of the 5th European Conference on Computer Vision*, 484-498. June 2-6, Freiburg (Germany).
- Gammage, C., and Chaudhary, V., 2006. On Optimization and Parallelization of Fuzzy Connected Segmentation for Medical Imaging. *Proc. of the 20th International Conf. on Advanced Information Networking and Applications*, 623-627. April 18-20, Vienna (Austria).
- Handels, H., 2009. Medizinische Bildverarbeitung. Wiesbaden (Germany) : 116-124.
- Ibanez, L., Schroeder, W., Ng, L., and Cates, J., 2003. *The ITK Software Guide: The Insight Segmentation and Registration Toolkit*. New York (United States) : Kitware Inc.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., and Wu, A.Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 24:881-892.
- Kuhnigk, J.-M., Dicken, V., Bornemann, L., Bakai, A., Wormanns, D., Krass, S., and Peitgen, H.-O., 2006. Morphological Segmentation and Partial Volume Analysis for Volumetry of Solid Pulmonary Lesions in Thoracic CT Scans. *IEEE Transactions on Medical Imaging* 25(4):417-434.
- Kwan, R.K.-S., Evans, A.C., Pike, G.B., 1999. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging* 18(11):1085-1097.
- MeVis, 2012. *MeVisLab : medical image processing and visualization*, MeVis Medical Solutions, Bremen (Germany). Available from: <http://www.mevislab.de/developer/documentation/> [04/2012].
- Osher, S., Sethian, J.A., 1988. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics* 79:12-49.
- Ritter, F., 2007. Visual Programming for Prototyping of Medical Applications. *IEEE Visualization 2007 workshop*.
- Stone, R.J., 2011. The (human) science of medical virtual learning environments. *Philosophical Transactions of the Royal Society* 366:276-285.
- Torres, K., Staskiewicz, G., Sniezynski, M., Drop, A., and Maciejewski, R., 2011. Application of rapid prototyping techniques for modeling of anatomical structures in medical training and education. *Folia morphologica* 70(1):1-4.
- Vincent, L., and Soille, P., 1991. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(6):583-598.
- Wulf, J., Vitt, K.D., Gehl, H.-B., and Busch, L.C., 2001. Anatomic Accuracy in Medical 3D Modeling. *Phidias Rapid Prototyping in Medicine Newsletter* No. 7, December 2011.
- Zwettler, G., Backfrieder, W., Swoboda, R., Pfeifer, F., 2009. Fast Fully-automated Model-driven Liver Segmentation Utilizing Slice-wise Applied Levels on Large CT Data. *Proceedings of the 21st European Modeling and Simulation Symposium*, 161-166. September 23-25, Tenerife (Canary Islands, Spain).

AUTHORS BIOGRAPHY

Gerald A. Zwettler was born in Wels, Austria and attended the Upper Austrian University of Applied Sciences, Campus Hagenberg where he studied software engineering for medicine and graduated Dipl.-Ing.(FH) in 2005 and the follow up master studies in software engineering in 2009. In 2010 he has started his PhD studies at the University of Vienna at the Institute of Scientific Computing. Since 2005 he is working as research and teaching assistant at the Upper Austrian University of Applied Sciences at the school of informatics, communications and media at the Campus Hagenberg in the field of medical image analysis and software engineering with focus on computer-based diagnostics support and medical applications. His e-mail address is gerald.zwettler@fh-hagenberg.at and the research web page of the Research & Development department at campus Hagenberg, he is employed at, can be found under the link <http://www.fh-ooe.at/fe/forschung>.

Werner Backfrieder received his degree in technical physics at the Vienna University of Technology in 1992. Then he was with the Department of Biomedical Engineering and Physics of the Medical University of Vienna, where he reached a tenure position in 2002. Since 2002 he is with the University of Applied Sciences Upper Austria at the division of Biomedical Informatics. His research focus is on Medical Physics and Medical Image Processing in Nuclear Medicine and Radiology with emphasis to high performance computing. Recently research efforts were laid on virtual reality techniques in the context of surgical planning and navigation.

ADAPTIVE BEHAVIOR IN COMPLEX HEALTHCARE INTERVENTIONS: ASSESSMENT USING COMPUTER SIMULATION

Jean-Christophe Chiêm^(a), Thérèse Van Durme^(b), Florence Vandendorpe^(c),
Olivier Schmitz^(d), Niko Speybroeck^(e), Sophie Cès^(f),
Jean Macq^(g)

^(a-g) Institute of Health and Society (IRSS),
Université Catholique de Louvain (UCL)

^(a)Jean-Christophe.Chiem@uclouvain.be, ^(b)Therese.Vandurme@uclouvain.be, ^(c)Florence.Vandendorpe@uclouvain.be,
^(d)Olivier.Schmitz@uclouvain.be, ^(e)Niko.Speybroeck@uclouvain.be, ^(f)Sophie.Ces@uclouvain.be,
^(g)Jean.Macq@uclouvain.be

ABSTRACT

Projects of case management of older persons were implemented in Belgium. This type of long-term healthcare interventions presents many aspects of complexity including human interactions and contextual effects. Computer simulation enables to mix advantages of classic qualitative and quantitative analyses to assess why differences in the implementation of these projects led to failure or success. A simulation model of case management was designed using a Rule-Based Methodology. Based on the research material, variables and rules definitions were elicited into a simple conceptual frame. The simulation was then confronted to field experts. The rigorous formulation clarified the description of the interventions. Parametric analyses were performed with the experts. Problems and adaptation processes were reported. Eventually, a number of recommendations were addressed. The simulation model presented should be seen primarily as a tool for thinking and learning. The computer simulation can help to unfold causal mechanisms of complex interventions.

Keywords: Computer Simulation, Participatory Research, Complex Healthcare Intervention, Case Management

1. INTRODUCTION

1.1. Complexity in Healthcare Interventions

Interventions aimed at improving chronic and long-term healthcare present many aspects of complexity. Their action mechanisms are often characterized by feedback loops, delays and non-linearity (Plsek and Wilson 2001). These effects cannot be modeled properly by linear quantitative tools, leading to inconclusive results when assessed with classical effectiveness studies such as quantitative randomized experimental designs (Eldridge et al. 2005). For this reason, qualitative studies are usually preferred, to take advantage of their explorative power.

Quantitative and qualitative methodologies are often presented in opposition to each other. However, both approaches present different advantages and rely on empirical data of different natures. On the one hand, quantitative studies establish relationships within a limited set of variables, following a mathematical normative approach. On the other hand, qualitative studies build on perceptions and interpretations to find a contextual meaning and propose a causation mechanism (Greenhalgh and Russel 2010). The contrast between these perspectives makes a formalization combining both approaches difficult to construct.

In addition, long-term interventions are characterized by non-linearity and non-determinism. This implies that the collected data represent only one specific instance of reality in one particular context. Hence, a proper data collection would ideally involve data on a very long term, with an exhaustive set of variables. The cost and feasibility of such studies make them hard to set up. Besides, the mere fact that such data are highly context-sensitive makes them hard to handle in a generalization process.

1.2. Computer Simulation: a Possible Answer?

Computer simulation has been argued to be one possible solution to overcome the complex aspects of healthcare studies (Auchincloss and Diez Roux 2008; Diez Roux and Auchincloss 2009). While already being used in the modeling of infectious disease (Epstein 2009), computer simulation has not yet been exploited to its full potential in health system research. However, further advances in this field may be encouraged by simulation applications from other fields.

Firstly, simulation research in the field of project management might bring useful insights for the evaluation of complex interventions in health care. Indeed, the simulation of projects modeling heterogeneous workers within teams and organizations has helped managers to take step forwards in decision-

making processes (Peculis, Rogers and Campbell 2007; Araúzo, Pajares and Lopez-Paredes 2010).

Secondly, the simulation strategy has also offered many opportunities in organizational and public policy modeling. The analysis of multiple scenarios enables to build more robust solution by considering policy decisions as an adaptive response that evolves over time (Lempert 2002).

Finally, simulation as a tool of participatory research might also shed light on health system research applications. Indeed, the opinions of field experts can prove to be an important and valuable source of information when data are scarce and submitted to the constraints mentioned earlier. These opinions can be gathered, confronted and aggregated using a simulation model. This results in the building of a formalized and generalized knowledge. Interesting applications in the fields of ecology have been designed using role-playing games and companion modeling methodology (Bousquet et al. 2007; Barreteau and Le Page 2011).

All these computer simulations typically use an Agent-Based or, more generally, Rule-Based Modeling (RBM) approach, which can also be used to assess complex interventions. The RBM consists in the definition of a set of variables and their rules of interactions (Collopy and Armstrong 1992). These rules can be expressed using a logical language such as if-else statements, which is well suited to translate experts' opinions. Other quantitative data can also feed the model. Moreover, random effects can be introduced and parametric analyses performed.

1.3. The Case of a Complex Intervention

The National Institute of Health and Disability Insurance in Belgium (NIHDI) has financed 62 innovative projects in order to enable the elderly to remain in their own residences (Protocol 3). A consortium of universities was asked to perform a scientific evaluation of all projects, in order to identify the types of projects likely to provide the most effective support to the elderly and help them to remain at home in good conditions.

However, the types of interventions covered a large scope of care services (case management, night care at home, occupational therapy, psychological and psychosocial support). In addition, the projects were submitted to diverse levels of embeddedness in the local healthcare system. Hence a classic comparison process between the projects was not an option.

Preliminary analyses of seemingly identical case management interventions (N=19) seem to reveal that they have adopted different patterns of adaptation leading to success or failure in achieving their pre-set goals. This may be explained by the high number of different actors and interactions involved in such interventions. Indeed, case management involves interactions with patients, healthcare providers and case managers, in assessing needs, planning, coordinating, carrying out and evaluating care.

Typical assessment of this hypothesis would require comprehensive case studies to assess and test emergent practice theories (Polit and Beck 2008). However, computer simulation provides a rigorous language that can help to build a structured assessment. In addition, parametric analyses can be performed, and different scenarios can be implemented. The whole implementation of the model can generate valuable insights regarding this theory, which can then be transferred with a reduced ambiguity.

In this paper, we show that computer simulation can help to unfold causal mechanisms of a complex intervention (i.e. case management in home-dwelling frail elderly) through the conception and the test of a conceptual frame. Moreover, the building of this type of framework supported by both quantitative and qualitative inputs might represent a valuable and reliable output of studies in a complex context.

2. METHODOLOGY

2.1. Gathering of Data

Several sources of data were available to perform computer simulation for the assessment of case management action on elderly depression. Firstly, empirical data about the health status of the elderly person were available from the BelRAI database. BelRAI is the adapted version of a multidisciplinary comprehensive geriatric assessment tool for the elderly, and is part of the international InterRAI project. The projects managers were asked to fill out these data via an on-line health questionnaire regarding the elderly people included in their project. In the context of our study, the level of depression state was particularly used as an indicator of well-being (Depression Rating Scale (Burrows et al. 2000)).

Secondly, the initial project proposal description, was used as the first key step in conceptualizing the project.

Thirdly, a yearly questionnaire relating to the description and implementation of the project was used to assess the current evolution of the project. Finally, scientific researchers conducted specific case studies analyses for 4 out of 19 projects to develop an in-depth understanding of the implementation: what was the planned intervention? With what result? For what target population? By whom? With what means? Did the project achieve its goals? How and why? How did the project adapt to the specific challenges of the context throughout the implementation process?

2.2. Conceptual Frame

Based on the preliminary interpretation of the researchers, a conceptual frame of the typical description of a project was built following a Rule-Based Methodology. Three agents were defined succinctly with a corresponding set of variables.

- Case Manager (CM): Knowledge, Informal Knowledge and Organizational Skills.

- Caregiver (CG): Knowledge and Capacity.
- Older person (ELD): Needs and Demands.

The links between the agents (topology) are represented in Figure 1.

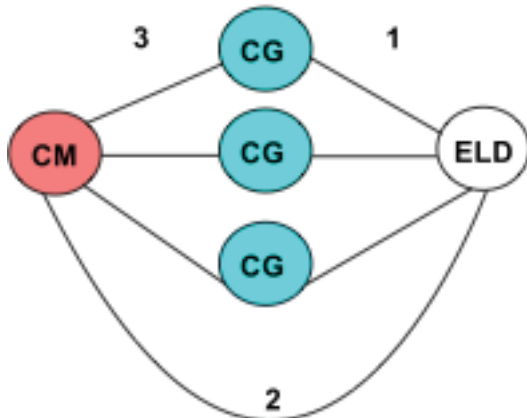


Figure 1: Conceptual Frame Representing the Interactions between Three Types of Agents: Case Manager (CM), Caregiver (CG), Older Person (ELD). The number of agents of each type can be varied.

Using these variables simple rules were designed integrating the qualitative material available, representing the opinions of stakeholders and researchers. Three parameters were used to determine 'reasonable' Demands, 'good' Organizational Skills and 'good' Informal Knowledge. The main rules were defined as follows.

1. The Needs of the ELD were set to increase constantly as a natural evolution.
2. Interaction CG-ELD (1): if the Needs of the ELD are met by the Capacity of the CG, these Needs decrease; else the Needs and Demands of the ELD increase. In addition, in the case where the Needs of the ELD are not met by the Capacity of the CG, then if the Demands of the ELD are reasonable, the CG increases his Capacity and his Knowledge (reflexive process); otherwise the CG decreases his Capacity.
3. Interaction CM-ELD (2): if the Informal Knowledge of the CM is good then he decreases the Demands of the ELD when they are not reasonable. When the Needs of the ELD are not met, the CM increases his Knowledge, Informal Knowledge and Organizational Skills (reflexive process).
4. Interaction CM-CG (3): if the CM have good Organizational Skills he can identify the ELDs for whom needs are not met, identify the associated CGs, and increase their Capacity and Knowledge.

The model was implemented in Netlogo, a free open source software providing a user-friendly

graphical interface (source code and accompanying implementation details are available at the following link <https://sourceforge.net/projects/cmsimulation/>).

This preliminary implementation was designed and presented to four experts in elderly case management. The definition of their project, the problems encountered and the possible mechanisms of adaptation were translated in terms of the variables and rules used in the conceptual frame.

Different scenarios could be investigated using the number of each type of agent, the initial states of their variables and the good/reasonable parameters. Different simulations were performed to construct the conceptual frame step by step beginning with the older persons only, and successively adding caregivers and a case manager. The group of experts and the researchers discussed each scenario.

3. RESULTS

3.1. Use of Empirical Data

As an initial step, raw Depressive Rating Scales data were retrieved from the BelRai. They were averaged over all projects and their evolution was succinctly investigated. Obviously finer analyses would be required to enhance the descriptive power of these data, possibly including confounding variables, and discussed for each project. However, the assessment of this single variable already revealed the inherent resulting blur when comparing across projects, and the difficulty of establishing a possible causation process from a before-after assessment.

In parallel, while designing the rules, poor quality in the data filled out in the questionnaire was observed. This revealed a problem in both the filling out of the questionnaires and the perception of the evaluation process by the stakeholders. Hence, these observations emphasize the problem of the reliability of the data in a complex context integrating human response under evaluation. To counter this issue of reliability, a triangulation of the data by means of case studies analyses in four projects enabled the clarification of questions and the purpose of the evaluation process.

Both these problems emphasize the need for other methodological approaches.

3.2. Ontological Step

The qualitative material was used to define in explicit details the variables used in the conceptual frame (see Table 1). While the definitions and language elaborated could not be exhaustive, this ontological step was considered as an essential common dictionary in the discussion with the experts. In addition, in terms of operationalization, it clarified the type of case management (e.g. according to the typology of Lee (Lee et al. 1998), such as brokerage model, integrated case management model, self-managed care, or a combination of them).

Table 1: Ontological Dictionary. Three Agents are defined: Case Managers (CM), Caregiver (CG), and the Older Person (ELD). Each Agent (1st column) is Associated with Corresponding Variables (2nd column) and their Details (3rd column).

CM	Knowledge	Evaluation of the ELD's needs: Use of diagnostic tools to measure physiological state; evaluate the capacity of the older person
	Informal Knowledge	Consider the specific demands of the ELD: Analyse and consider the sociodemographic context; take into account the resources and will of the older person; establish a good acquaintance with the older person
	Organizational Skills	Organization of a fine-tuned envelop of care: Enable an access to differentiated care and services; setup efficient and durable partnership; coordinate and work at minimal cost
CG	Capacity	Deliver care/services required by the CM: Deliver quickly efficient care and service; be accessible
	Knowledge	Knowledge of the physiological state and capacity of the ELD: Fill out and use adequately ELD information; share information and collaborate with other CGs
ELD	Needs	State of "health/frailty": Regarding physiological, functional, psychological, socially, relational aspects
	Demands	Life habits and preferences: Influence the capacity to comply to the proposed care/service; ex: organization of daily life, cultural aspects, incomes, general ability

3.3. Confrontation to the Experts

First, the conceptual frame was presented to the experts in elderly case management, together with the ontology of the variables. Quite naturally, the simplistic and

possibly controversial definitions of the variables were subject to discussion, refining the concepts behind the terms. For example, the experts would have exchanged the terms 'Demands' and 'Needs' or would have renamed the informal demands as 'Preferences'. In the subsequent discussion, the terminology was then specified in 'Formal Needs' and 'Informal Demands'. Despite the possible lack of political correctness of the terms, the experts could accept and handle the meaning of the variables.

Next, the experts were invited to define their own projects within the conceptual frame, in terms of the established variables. The experts were able to discuss, compare their projects, and position themselves within the schema with surprising differences in terms of topologies (links between actors), with different roles, profiles and missions of the case manager. For example, one expert positioned himself higher ('on the left') of the CM. Another described the agent ELD as being a group of elderly people. Another spontaneously declared that "this exercise allows a definition of factors and characteristics which look similar through projects, but that are actually totally different".

Then, the rules of interactions were presented and discussed. The written and spoken formulations appeared quite difficult to understand. However, running the simulation and visualizing the rules dynamics into graph (Figure 2) enabled a better comprehension.

Finally, the step-by-step reconstruction of the conceptual frame and playing with the parameters trained them to interpret simulations, similar to the one in figure 2, as follows. While the Needs of the ELDs increase, the Knowledge, Informal Knowledge and Organizational Skills of the CM increase as well. When Organizational Skills and Informal Knowledge become good, the CM is able to increase the Capacity of the CGs and to decrease the Demands of the ELDs. This results in a stabilization of the Needs of the ELDs.

The discussion of these simulations enlightened by insights of both experts and researchers led to the definition of problems, and the way the project adapted to them, according to the context in which the case management was delivered. Eventually, a number of recommendations for the implementation of case-management projects could be addressed.

3.4. Formal Needs and Informal Demands

Through all simulations, experts pointed out that making the distinction between formal needs (assessment and coordination) and informal demands (understanding and interpretation) was one big challenge for an effective case management. This justified a posteriori the two kinds of knowledge of the case manager.

In addition, the experts noted that a state of equilibrium in the different variables was reached after a certain time. While this is in fact a numerical property of the simulation, the experts interpreted this as the time of adaptation required for the case manager to increase

their informal knowledge about the elderly individuals and their caregivers, which appears to be a reality of the field.

Further, while the question of the time of adaptation was raised, it came up as a recommendation that it is better to reduce the length of the monitoring in favor of the frequency of visits for both caregiver and case manager. While this intuitively enables to monitor for potential random accidents, it also contributes better to the understanding of the informal demands of the older person.

Moreover, recourse to persons previously acquainted with the older person (nurses, GPs, informal caregivers) or structures may help in a phase of needs assessment by better understanding the informal demands. While further management skills will still be required, this step may help to decrease the period of adaptation when attempting to organize or implement case management interventions.

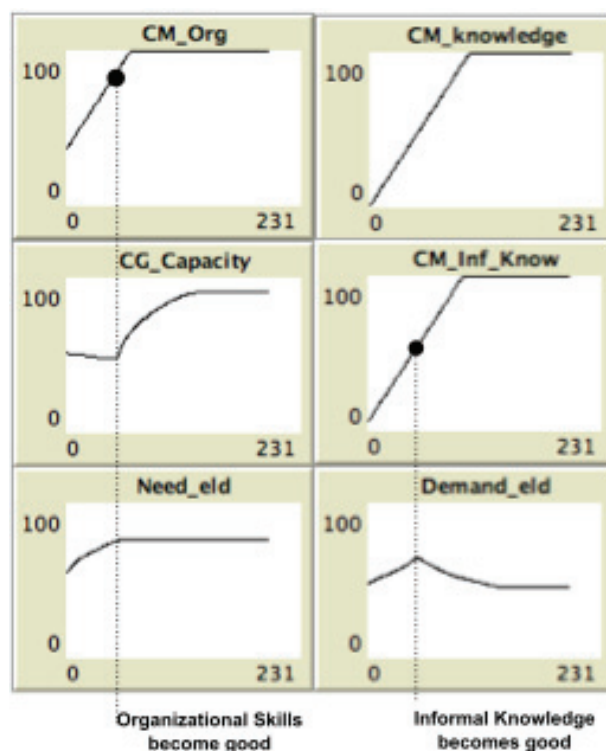


Figure 2: Illustrative Output. Mean Evolutions of the Case Manager's Organizational Skills, Knowledge and Informal Knowledge (CM_Org, CM_knowledge, CM_Inf_Know); the Caregivers' capacity (CG_Capacity); the Needs and Demands of the Older persons (Need_eld and Demand_eld).

3.5. Information Flow

It emerged in the discussion that in fact, in an ideal setup, there is no need for a case manager. This ideal situation would suppose that perfect information about each elderly person is distributed among all caregivers in real time. This was formulated by one of the expert, referring to the conceptual frame (Figure 1): "Somehow, interaction 2 is the same as interaction 1, only modulated by interaction 3".

These remarks highlighted possible enhancements in the design of our conceptual frame. First, the caregiver may be able to understand the informal demands as well. In fact, the limit between a caregiver and a case manager may be thin and modulated by the transfer of information between the different actors. Second, links between the caregivers could be usefully used to consider the sharing of information and the organization of the care.

A possible setup to enhance the transmission of the information could involve a team of caregivers performing case management via intercommunication and meetings. However, this scenario appeared too ideal. But in most projects, this transfer of information has been implemented via a software solution. In other cases, the entire purpose of a project was to ensure that information is dispatched among the care network of the elderly person.

3.6. Institutional Context

Discussing with the experts, it appeared that the institutional context (political and financial) was responsible for the creation of obliged topologies and profile distributions between case manager and caregivers, which were not always optimal. Two cases were reported. In one case, financial incentives created competition between the caregivers. In another, political competencies required that the organization skills and the knowledge skills (needs assessment of the older person) could not be attached to one same case manager. This might result from the Belgian-specific repartition of competencies among different authorities.

Hence, a good comprehension of the initial context and its evolution is an advantage for every case manager. Finally, it appears that a harmonization of the institutional context could help.

Regarding this point and the preceding ones, it appears that one specific ideal style of case manager does not exist. A case manager should be suited to the institutional context and take into account the capacities of each actor of the care. Moreover, mandating too explicitly the ideal profile of the case manager might be counterproductive.

3.7. Organization of Care

In answer to the low quality of care and increasing needs of patients in some simulated patterns, some adaptation processes were described as experienced by the projects.

First, one project created a distinction between older persons with urgent needs and persons requiring less attention. Second, some projects decreased the size of the targeted population. Finally, at the opposite, in some projects supplementary caregivers needed to be added to fulfill the needs of the population.

These patterns highlight the fact that the model was missing information about the turnover of the caregivers and elderly persons. This might also be an important indicator regarding the adaptive patterns of the project.

4. DISCUSSION

4.1. A Valuable Cognitive Process

The simulation model presented here should be seen primarily as a tool for thinking and learning about the case management of home-dwelling older persons (Edmonds 2010). Several aspects of the methodology contribute to a valuable cognitive process.

First, the conceptual frame expressed in terms of variables provides a vocabulary and an architecture to properly elicit rules of dynamic interaction. The use of a formal language forces a non-ambiguous formulation of the rules and a clarification of the contextual effects.

Second, this formalism and the flexibility of RBM make it a powerful tool of communication and negotiation between experts, providing an interactive medium for social exploration (Edmonds 2001). The informative interaction between modelers and experts allows for improving and learning from the elicitation. In addition, while the experts might already have the right knowledge, the simulation can help them to phrase or adjust their internal representation. Better understanding and good communication skills can only improve their ability to manage and advise other stakeholders.

Finally, the discussion resulting from the confrontation is also beneficial to the experts. Indeed, the output of a simulation of the type of Figure 2 acts as a brainteaser on the experts' reasoning. If the dynamic of a simple rule might appear quite intuitive when taken independently, their aggregation becomes much more difficult to handle. Hence, from the output graph comes out a question, a possible answer is formulated, triggering a reinterpretation, and finally the experts associate the resulting dynamics to a case story of their own experience. This cognitive scheme enhances the expert's internal representation of the mechanisms underlying the interventions and ultimately prepares them to adapt better to unpredictable situation.

4.2. Usefulness of the Recommendations

The insights and the recommendations that emerged from the discussion might appear somewhat obvious. However, personal internal representations might lead to different interpretations of these recommendations. The participatory process involving field experts permits to ensure that each participant understands everyone's interpretations feeding a discussion eventually resulting in better generic formulations. Further, if RBM has been argued to be a tool for consensus reaching, it can also be used to further elicit points of disagreement.

Moreover the obviousness of some of the recommendations does not mean that they should be left out. Hence, the simulation model provides a structure for a systematic inventory of these recommendations, which is a valuable step in a decision-making process.

Finally, the recommendations issued from this type of exercise will likely have a strong advantage when it

comes to field integration. Indeed, policies imposed from authorities may lead to failure because of a perceived or a real gap with the field (Hunter and Marks 2002). Instead, the confrontation with field experts and the resulting refinement aim at better generalizing their knowledge and experience as anchored in their day-to-day reality. Hence, such guidelines emerge from a bottom-up approach and may trigger a peer effect, facilitating their acceptance and implementation.

4.3. Validation and Limitations

At this early stage of the model creation, the level of complexity remains limited. However, the necessity to remain slow in the design process needs to be emphasized. Firstly, the time to analyze each step contributes to structure and formalize the experts' representations. Secondly, each rule needs to undergo a strong and rigorous evaluation by the experts to assess its stand-alone mechanics before being merged into the rest of the algorithm. Hence, this assessment contributes to a validation process in the simulation context.

In addition, the benefits of the learning process of the experts are neither immediate nor quantifiable (Berland and Rand 2009). As recommendations and rules might appear intuitive or obvious depending on the a priori knowledge and experience, the lessons learned might be different for every expert. New simulations are drawn within seconds and interpreted without consequences; actual application and benefits in real life situations linked to the simulation exercise would be hard to highlight.

Finally there is an inherent destabilizing feeling when the simulation model is presented. In a sense, experts feel uneasy with the oversimplification and the cliché of huge lines of code. However, the natural language of the rules quickly frees them and leaves room for enthusiasm and imagination. In addition, as general computer literacy increases more and more, a greater interest from experts is expected in the future.

4.4. Perspectives

The creation of a simulation model as designed, confronted to and fed by experts' opinion is an iterative process. New iterations of this process are required to refine the model taking into account the experts' remarks.

The adaptation processes mentioned should be all investigated. In order to test these processes, new scenarios should be implemented to simulate potential problems and assess how robust the solutions are.

Further questions will need to be investigated, regarding economic matters, time resources and institutional embeddedness as they seem crucial to the understanding of the problems encountered by the case management.

Finally, the simulated results should be compared using the quantitative data available to possibly recreate real project case story. For example, change in depression level can be used to assess project efficacy and evolution in terms of preserving or enhancing the

well-being of the elderly person. Mixing quantitative and qualitative approaches through computer simulation methodology could only improve our insights of the causal mechanisms underlying complex interventions.

ACKNOWLEDGMENTS

The authors are grateful for the scientific evaluation grant of the innovative Protocol 3 projects by the National Institute for Health and Disability Insurance (NIHDI). They also wish to thank all colleagues of the consortium and project coordinators who contributed to this study.

REFERENCES

- Araújo, J.A., Pajares, J., Lopez-Paredes, A., 2010. Simulating the dynamic scheduling of project portfolios. *Simulation Modelling Practice and Theory* 18: 1428-1441.
- Auchincloss, A.H., Diez Roux, A.V., 2008. A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health. *American Journal of Epidemiology* 168: 1-8.
- Barreteau, O., Le Page, C., 2011. Using Social Simulation to Explore the Dynamics at Stake in Participatory Research. *Journal of Artificial Societies and Social Simulation* 14: 12.
- Berland, M., Rand, W., 2009. Participatory simulation as a tool for agent-based simulation. *Proceeding of the International Conference on Agents and Artificial Intelligence*, pp. 553-557. Porto, Portugal.
- Bousquet, F., Castella, J.-C., Trébuil, G., Barnaud, C., Boissau, S., Kam, S.P., 2007. Using multi-agent systems in a companion modelling approach for agroecosystem management in South-east Asia. *Outlook on Agriculture* 36: 57-62.
- Burrows, A.B., Morris, J.N., Simon, S.E., Hirdes, J.P., Phillips, C., 2000. Development of a minimum data set-based depression rating scale for use in nursing homes. *Age Ageing* 29: 165-172.
- Collopy, F., Armstrong, J.S., 1992. Rule-Based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations. *Management Science* 38: 1394-1414.
- Diez Roux, A.V., Auchincloss, A.H., 2009. Understanding the social determinants of behaviours: Can new methods help? *International Journal of Drug Policy* 20: 227-229.
- Edmonds, B., 2001. The Use of Models - Making MABS More Informative. *Proceeding of the Second International Workshop on Multi-Agent-Based Simulation-Revised and Additional Papers*, pp. 15-32.
- Edmonds, B., 2010. Bootstrapping Knowledge About Social Phenomena Using Simulation Models. *Journal of Artificial Societies and Social Simulation* 13: 8.
- Eldridge, S., Spencer, A., Cryer, C., Parsons, S., Underwood, M., Feder, G., 2005. Why modelling a complex intervention is an important precursor to trial design: lessons from studying an intervention to reduce falls-related injuries in older people. *Journal of Health Service Research and Policy* 10: 133-142.
- Epstein, J., 2009. Modelling to contain pandemics. *Nature* 460: 687-687.
- Greenhalgh, T., Russell, J., 2010. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. *PLoS Medicine* 7: e1000360.
- Hunter, D., Marks, L., 2002. *Decision-Making Processes for Effective Policy Implementation*. National Institute of Clinical Excellence. Available from: http://www.nice.org.uk/niceMedia/pdf/SemRef_Decision_Hunter.pdf
- Lee, D.T., Mackenzie, A.E., Dudley-Brown, S., Chin, T.M., 1998. Case management: a review of the definitions and practices. *Journal of advanced nursing* 27: 933-939.
- Lempert, R., 2002. Agent-based modeling as organizational and public policy simulators. *Proceedings of the National Academy of Sciences*, pp. 7195-7196.
- Peculis, R., Rogers, D., Campbell, P., 2008. An Agent-Based Framework for Simulating Socio-Organisational Design of Large Projects. *Proceedings of the Simulation Technical Conference 2008*. Melbourne, Victoria, Australia
- Plsek, P.E., Wilson, T., 2001. Complexity, leadership, and management in healthcare organisations. *Bmj* 323: 746-749.
- Polit, D.F., Beck, C.T., 2008. *Nursing research : generating and assessing evidence for nursing practice*. Philadelphia: Wolters Kluwer Health/lippincott Williams & Wilkins.

QUALITATIVE FEATURES OF A NOVEL BARORECEPTOR MODEL

Adam Mahdi^(a), Johnny Ottesen^(b), and Mette Olufsen^(c)

^(a,c)Department of Mathematics, North Carolina State University, Raleigh, NC

^(b)Department of Science, Systems, and Models, Roskilde University, Roskilde, Denmark

^(a)amahdi@ncsu.edu, ^(b)johnny@ruc.dk, ^(c)msolufsen@ncsu.edu

ABSTRACT

Blood pressure regulation by the cardiovascular system is a complex physiological process. Cardiovascular modeling can offer a valuable insight often beyond the reach of experiments. In this study we provide a new mathematical model of the afferent component of the baroreflex feedback system. The model takes advantage of the so-called quasi-linear viscoelastic theory, which has been widely used to describe the nonlinear viscoelastic response of living tissue. It also uses a simple integrate-and-fire model to predict the baroreceptor response and therefore takes into account the conceptual structure of the baroreceptor. Our objective is to test our new baroreceptor model for its ability to reproduce experimental data qualitatively and demonstrate known pressure-response relationships. We also highlight that the model can be coupled with an existing model of the efferent pathways, eventually predicting heart rate.

Keywords: heart rate regulation, baroreflex, blood pressure dynamics, mathematical modeling

1. INTRODUCTION

Understanding the cardiovascular control system is crucial for gaining more insight into the physiology not only for the healthy individual, but also in order to detect pathologies. Its main role is to provide adequate perfusion of all tissues, which is achieved by maintaining blood flow and pressure at a fairly constant level. To accomplish this, a number of control mechanisms are imposed regulating vascular resistance, compliance, pumping efficiency and frequency. An important contributor to this control system is the *baroreflex* (or *baroreceptor reflex*), which uses specialized neurons called *baroreceptors* for signaling. The baroreceptor neurons are activated via mechano-sensitive channels located in the aortic arch and carotid sinuses. It is believed that the baroreceptor nerves are the main contributor to the short-term regulation of vascular efferents including: heart rate, cardiac contractility, and vascular resistance and vessel tone (Levick 2010).

Prediction of heart rate from blood pressure involves two main pathways: afferent and efferent. *Afferent* pathways integrate firing of the baroreceptors in the nucleus solitary tract. *Efferent* pathways modulate sympathetic and parasympathetic signals, which lead to release or inhibition of the neuro-transmitters: acetylcho-

line and noradrenaline, which in turn modulate the effectors. In this study, we focus on regulation of heart rate. See Figure 1 for overall, conceptual division of the model predicting heart rate.

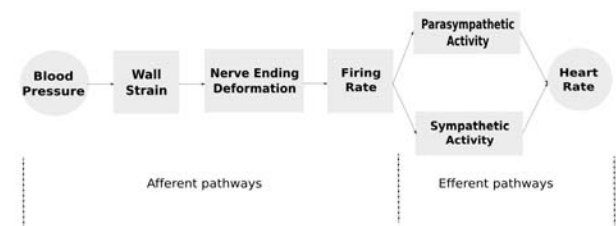


Figure 1: Conceptual division of the mathematical model for the baroreflex feedback control of heart rate.

This study focuses on analyzing qualitative aspects of a new model for the afferent signaling, whereas the efferent dynamics can be predicted using either the existing model developed by Olufsen et al. [2006], or any other model. In Section 2, after reviewing the main qualitative characteristics of the baroreceptor dynamics we introduce our model, which will be analyzed and discussed in Section 3.

2. METHODS

In this section we describe the principal physiological elements of the baroreceptors; identify the most prominent qualitative features of their discharge observed in experiments; and introduce a mathematical model that reflects all those characteristics.

2.1. Basic physiological facts and experiments

For most mammals baroreceptors are found in the aortic arch and the carotid sinuses (Sharwood 2001). These neurons are stimulated via activation of stretch receptors, which are able to detect changes in the wall strain induced by changes in blood pressure. Besides water, which makes up to 70% of arterial wall, it consists of: muscles, elastin, collagen and ground substance. The wall is commonly divided into three layers: the *tunica intima*, a thin layer of endothelial cells lining the arterial wall; the *tunica media* (the middle layer), the primary contributor to the arterial wall deformation; and the *tunica adventitia* (the outer layer) connecting the vessels to their surrounding tissue. Understanding the mechanics and the viscoelastic properties of the arterial wall is essential for modeling the baroreceptor response.

Although some characteristics of the baroreceptor response depend on their type, e.g., whether they have myelinated or unmyelinated axons (Brown 1978), it is possible to identify a number of common signaling patterns. The most prominent static and dynamic characteristics include: saturation, threshold (Seagard et al. 1990), adaptation (Brown 1980), hysteresis, post-excitatory depression (PED) (Brown 1980) and sensitivity to the rate of change of the stimulus (Landgren 1952). It is clear that any complete, baroreceptor model should be able to reflect these important features. Therefore, we first review them in more detail.

Adaptation: If the pressure changes and resets at a new constant value, the baroreceptor-firing rate follows, reaching a new steady state (Landgren 1952). The frequency of the steady discharge is the same whether the new pressure level is reached from a higher or a lower pressure (Brown et al. 1976, Sleight 1980).

Saturation: As pressure increases so does the firing rate, but after a certain increase no further firing rate increase is observed with the increase of the input (Landgren 1952). A constant decrease of the pressure results in a decrease of the firing rate, first in an almost linear and then in a hyperbolic manner tending toward some limiting value (Coleridge et. al. 1981). It has also been observed that the saturation level strongly depends on the type of the baroreceptor (Seagard et. al. 1990, Van Brederode et al. 1990).

Threshold: The increase of pressure from zero does not result in an immediate response. The nerves do not start to fire until the pressure crosses a certain threshold value.

Asymmetry: A difference in the response to a rising and falling pressure is referred to, in general, as asymmetry or hysteresis (Bronk and Stella 1932). In particular, periodic inputs (such as a triangular or sinusoidal wave) produce loops in the pressure-frequency response causing asymmetry also referred to as hysteresis (Coleridge et al. 1981).

Post-Excitatory Depression (PED): After a step decrease in pressure, firing may cease for some seconds, followed by a recovery to the steady state firing rate, commensurating with the new established pressure level (Brown 1980). It has been observed that the length of the pause (also denoted the refractory period) depends on the depth of the pressure drop (Wang et al. 1991).

2.2. Afferent dynamics – the model

One of the first attempts to quantitatively describe the activity of baroreceptors goes back to Landgren (Landgren 1952). Since then, a number of models have been proposed. Unfortunately most of them are not able to reflect all of the known qualitative features of the response (Spickler and Kezdi 1967) or have no biological foundations (Taher et al. 1988). Still other models treat the prediction of baroreflex firing purely from the mechanical perspective, and thus offering very superficial or no description of the neural part of the response (Srinivasan and Nudelman 1972, 1973).

In this study, the modeling processes involved with

predicting baroreceptor firing (described later in this section) will reflect the conceptual division of the afferent pathways including: arterial wall deformation (wall strain); nerve ending deformation (nerve strain); and action potential formation (firing rate). To be more precise, given an arterial pressure $p(t)$ we will model the viscoelastic response of the arterial wall $\varepsilon_w(t)$, which will be used as an input to obtain the current generated along the axons $I(t)$, which will be integrated to predict the baroreceptor firing rate $f(t)$. The sections below describe each of these components in detail.

Arterial wall deformation

Most models predicting the viscoelastic response of the arterial wall due to changing pressure are formulated as linear models involving springs and dashpots. It was observed by Fung (Fung 1993), that biological tissues are not elastic, but that the history of the vessel strain affects the stress. Moreover, he observed that there is a difference in stress response between loading and unloading. Generalizing the linear viscoelastic theory, Fung introduced the so-called *quasi-viscoelastic theory* (QLV), which has been successfully used in modeling stress-strain relationship of the arterial wall (Valdez-Jasso et al. 2009-2011). We shall use the QLV theory in order to model the strain of the arterial wall to changes in pressure. We proceed under the assumption the arterial wall can be modeled as a homogeneous and isotropic cylindrical vessel with a thin wall (Fung 1996). For such a vessel, the wall strain can be predicted as a function of pressure given by

$$\varepsilon_w(t) = \int_{-\infty}^t K(t-\gamma) \frac{\partial s^e[p(\gamma)]}{\partial \gamma} d\gamma,$$

where following (Valdez-Jasso, 2009) the *creep function* $K(t)$ is given by

$$K(t) = 1 - A_1 e^{-\frac{t}{b_1}},$$

with the *elastic response* $s^{(e)}[p(t)]$ defined by

$$s^{(e)}[p(t)] = 1 - \frac{\sqrt{A_0(p^k(t) + \alpha^k)}}{\sqrt{A_m p^k(t) + A_0 \alpha^k}}.$$

In this equation A_m and A_0 (cm^2) denote the maximal and zero pressure cross-sectional area of the vessel, respectively, α ($mmHg$) denotes the characteristic pressure at which the vessel starts to saturate, and k denotes the steepness of rise of the sigmoid curve.

Nerve ending deformation

The stimulation of baroreceptor firing is probably the least understood element of the afferent pathway and therefore in most models it is either omitted or treated superficially. Our modeling process is based on the observation that the baroreceptor firing rate is sensitive not only to the intrasinus mean pressure, but also to its rate of change (Ursino 1999). Moreover, we assume that the adaptation process is due to the coupling of the baroreceptor nerve with the wall. Thus, we propose to

model the stretch of the baroreceptors by the following simple dynamical system

$$\begin{aligned}\frac{dx_1}{dt} &= -\alpha_1 x_1 + \alpha_2 x_2 + \gamma_1 \varepsilon_w \\ \frac{dx_2}{dt} &= -\alpha_3 x_1 - \alpha_4 x_2 + \gamma_2 \varepsilon_w,\end{aligned}$$

where $\alpha_1, \dots, \alpha_4$ and γ_1, γ_2 are parameters. In particular, if the parameters of this system are written as

$$\begin{aligned}\alpha_1 &= \frac{E_0 + E_1}{\eta_1} + \frac{E_0}{\eta_2}, & \alpha_2 &= \frac{E_1}{\eta_1} - \frac{E_2}{\eta_2}, \\ \alpha_3 &= \frac{E_0}{\eta_2}, & \alpha_4 &= \frac{E_2}{\eta_2}, \\ \gamma_1 &= \frac{E_0(\eta_1 + \eta_2)}{\eta_1 \eta_2}, & \gamma_2 &= \frac{E_0}{\eta_2},\end{aligned}$$

it can be interpreted as a linear viscoelastic model with a spring and two Voigt bodies in parallel, where x_1, x_2 are denote the relative displacement from the rest position. Similar ideas have been used in the context of baroreceptor modeling in (Alfrey 1997, Bugenhagen et al. 2010), and before that for the modeling the muscle spindle dynamics (Houk et al. 1966, Hasan 1983). Following this idea, the strain sensed by the nerve ending of the receptors due to the changing strain of the arterial wall is given by

$$\delta(t) = \varepsilon_w(t) - x_1.$$

We incorporate a simple mechanism for the threshold, predicting the input current $I(t)$ needed to stimulate the neuron by

$$I(t) = \begin{cases} 0 & \text{if } \delta < \delta_{th} \\ s_1(\delta - s_2 \delta_{th}) & \text{if } \delta \geq \delta_{th} \end{cases},$$

where s_1, s_2 and δ_{th} are parameters.

Firing rate

Finally, the neural part of the baroreceptor is predicted using a simple integrate-and-fire model described by the first-order differential equation

$$C \frac{dv(t)}{dt} + Gv(t) = I(t),$$

where G and C represent the neuron's membrane conductance and capacitance, respectively. It is assumed that the neuron fires if the voltage on the membrane reaches the threshold value V_{th} . In other words, when $v(t) = V_{th}$ an action potential is generated and the capacitor voltage is reset to zero. Simple computation shows that for a constant stimulus, the time it takes to integrate the model from 0 to V_{th} is given by

$$T = \frac{C}{G} \log \left(\frac{I(t)}{I(t) - GV_{th}} \right).$$

It is well known that after an action potential has occurred the neuron cannot fire for several milliseconds (Izhikevich 2007). This time is generally referred to as

the *refractory period*, which we denote by t_{ref} . The refractory period t_{ref} , can be added to T giving

$$f = \frac{1}{T} = \frac{1}{\tau \log \left(\frac{I(t)}{I(t) - GV_{th}} \right) + V_{th}},$$

where $\tau = C/G$. We note that for a sufficiently large current $I(t)$, the term involving the *log* function is arbitrary small. Thus because of the refractory period, the maximum firing rate is given by (Chen 2004)

$$f_{sat} = \frac{1}{t_{ref}}.$$

2.3. Efferent responses: heart rate dynamics

For the sake of completeness we point out how the afferent and efferent parts can be used to predict the heart rate. Following (Olufsen et al. 2006) the sympathetic and parasympathetic tone can be predicted as

$$T_{par} = \frac{f}{N}, \quad T_{sym} = \frac{1 - f(t - \tau_d)}{N},$$

where N is the baseline firing rate and τ_d is a delay for sympathetic stimulation. Stimulation of sympathetic and parasympathetic outflow modulate the concentration of the neurotransmitters acetylcholine C_{ach} and noradrenaline C_{nor} , which can be predicted from the differential equations

$$\frac{dC_{ach}}{dt} = \frac{T_{par} - C_{ach}}{\tau_{ach}}, \quad \frac{dC_{nor}}{dt} = \frac{T_{sym} - C_{nor}}{\tau_{nor}},$$

where τ_i denotes characteristic timescales. Finally, heart rate H can be computed as

$$H = H_0(1 - M_{nor}C_{nor} + M_{nor}C_{nor}),$$

where H_0 is the baseline firing rate and M_i are constants weighting each of the neurotransmitters.

3. QUALITATIVE ANALYSIS

To test the ability of the baroreceptor model, introduced in Section 2 to reflect the qualitative features of the baroreceptor firing rates observed in experiments, we study the response to the following pressure inputs: a step-increase and decrease, a sine wave, and a continuous ramp. With these stimuli, it is possible to show that the model under investigation is able to reflect all the main static and dynamics features listed in Section 2.1. Figure 2 shows data (adapted from previous studies by Brown [1978, 1980] and Seagard [1990]) from the three pressure input types and Figure 3 shows similar responses obtained with the proposed model.

We begin the testing the model's response to a pressure step increase and decrease (Figures 2 and 3, top panels). A step change (either an increase or a decrease) is one of the most commonly used pressure stimuli for studying the activity of the afferent baroreceptor nerves. The response to this stimulus has been discussed in several previous studies (Franz 1971, Clarke 1968, Brown 1978, Brown 1980, Taher et al. 1988).

Figure 2 (top panel) shows a typical experimental result (from Brown 1980). For comparison, our model shows similar dynamics including overshoot, threshold, adaptation, post-excitatory depression, and saturation. It should be noted that similar to the experiments, the adaptation occurs with different time-scales in response to a step increase or decrease, respectively. One aspect, not shown here, is that the steady-state discharge depends only on the level of the stimulus, which is in accordance with the experimental observations by Brown et al. [1976].

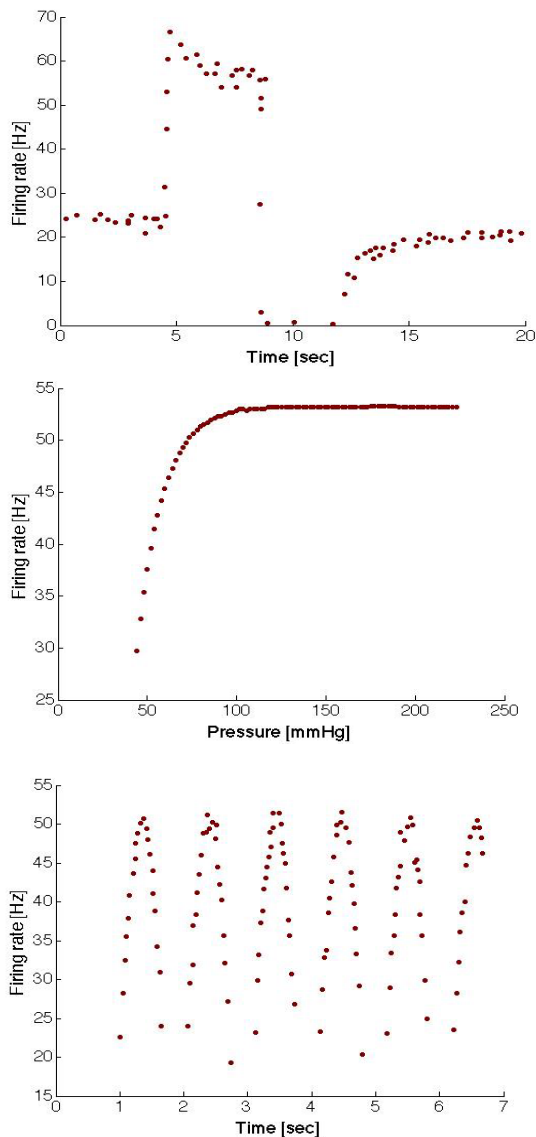


Figure 2: Typical baroreceptor response to a step (top; from Brown 1980), ramp (center; from Brown 1978), and sinusoidal (bottom; from Seagard 1990) pressure stimulus. The response to a step change (either a step increase or decrease) shows that the baroreceptor exhibits threshold, overshoot, adaptation and post-excitatory depression. The ramp response shows that the firing discharge saturates, and the sinusoidal response shows that in the lower frequencies discharge to disappear.

The model also predicts a *threshold pressure with steady discharge*, i.e., the minimum pressure value needed for obtaining a steady nonzero discharge. This feature has been observed by (Landgren 1952). Most of the model's dynamics stems from the equations predicting viscoelasticity of the nerve endings, and therefore it is typically assumed that the firing rate modulation is mediated by the coupling of the nerve ending with the arterial wall. Next we test the model's ability to predict the response to a ramp pressure profile (Figures 2 and 3, center panel).

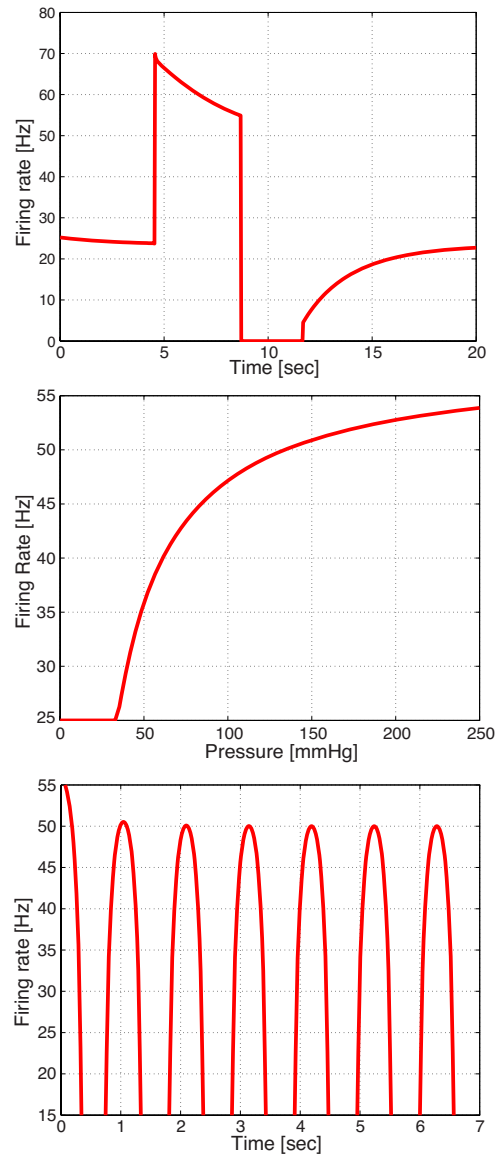


Figure 3: Baroreceptor responses obtained with the proposed model. Model responses shown here are designed to qualitatively predict similar dynamics as those observed experimentally (see Figure 2). Top shows results to a step increase and decrease; the center panel shows results to a ramped pressure stimulus, and the bottom panel shows the response to a sinusoidal stimulus.

The continuous ramped up pressure input is another commonly used stimulus, see previous studies by Clarke [1968], Coleridge et al. [1981, 1987], and Seagard et al. [1990].

As noted in the introduction, two physiological arguments can be used to explain why saturation occurs. First, the arterial wall can only deform finitely, baroreceptors are imbedded in the arterial wall, and thus, beyond a certain pressure stimuli the wall will no longer expand and neither will the baroreceptor nerve endings. This property can be incorporated, by imposing *nonlinear elastic response* within the viscoelastic model predicting arterial wall strain. Saturation has been described in previous studies, see (Kalita and Schaefer 2008, Valdez-Jasso 2009). Second, as described in the methods section, the refractory period following the firing of the neuron causes saturation. This phenomenon has also been observed previously; see (Koch and Segev 1998, Izhikevich 2007). As illustrated on Figure 3, we noted that both elements are well reflected in the modeling proposed model. This figure also shows threshold, as the discharge does not take place until the minimum value of the stimulus is reached.

Finally, we tested the dynamic response using a sine wave pressure profile as the stimulus. Comparison of Figures 2 and 3 (bottom panels), show that the model is in qualitatively agreement with experimental observations (Brown et al. 1978, Franz et al. 1971). The model predicts a time-varying firing rate that ceases when the pressure is below a given threshold. In addition to the result depicted here we tested the model using pressure waves with varying amplitudes, results showed that the firing rate is sensitive to the rate of change of the stimulus. This observation is in agreement with observations by Landgren [1952], Spickler and Kezdi [1967].

4. CONCLUSION

In this study we introduced and analyzed a new model predicting dynamics within the afferent pathways of baroreflex regulation. The main focus was on predicting firing of the afferent baroreceptor nerves. By using various pressure profiles we tested our model's ability to reproduce the main qualitative features reported in the literature, and we observed that our model is able to reproduce all known observed features including: threshold, saturation, overshoot, adaptation, post-excitatory depression, and sensitivity to the rate of change of the stimulus. We also showed that the firing rate model has potential to be coupled with an efferent model allowing prediction of heart rate. The efferent model can either be the existing model presented in section 2.3 (Olufsen et al. 2006), or any model that utilizes the afferent firing rate to predict heart rate. In future studies, we will test this model's ability to predict heart rate observed in both rat and human data.

ACKNOWLEDGMENTS

Mahdi and Olufsen were partially supported by NIH-NIGMS grant #1P50GM094503-01A0 subaward to

NCSU. Olufsen and Ottesen were supported in part by Snedkermester Sophus Jacobsen and wife Astrid Jacobsen's foundation, and Olufsen partially supported by NSF-DMS under grant #1022688.

REFERENCES

- Alfrey K.D., 1997. *Model of the aortic baroreceptor in rat*. M.S. Thesis, Rice University, Houston, TX.
- Bronk D.W. and Stella, G., 1932. Afferent impulses in the carotid sinus nerve I. The relation of the discharge from single end organs to arterial blood pressure. *J Cell Comp Physiol*, 1: 113–130.
- Bronk D.W. and G. Stella, 1935. The response to steady state pressure of single end organs in the isolated carotid sinus. *Am J Physiol*, 110: 708.
- Brown A.M., 1980. Receptors under pressure. An update on baroreceptors. *Circulation Research*, 42: 694-702.
- Brown A.M., Saum W.R., and Tuley F.H., 1976. A comparison of aortic baroreceptor discharge in normotensive and spontaneously hypertensive rats. *Circulation Research*, 39:488-496.
- Brown A.M., Saum W.R., and Yasui S. 1978. Baroreceptor dynamics and their relationship to afferent fiber type and hypertension. *Circulation Research*, 42: 694-702.
- Bugenhagen S.M, Cowley A.W. Jr., and Beard D.A., 2010. Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the Dahl SS rat. *Physiol Genomics*, 42: 23-41.
- Clarke, W.B. Static and dynamic characteristics of carotid sinus baroreceptors, 1968. Ph.D. Thesis. University of Rochester, Rochester, N.Y.
- Coleridge, H.M., Coleridge J.C.G., Kaufman M.P., and Dangel A., 1981. Operational Sensitivity and acute resetting of aortic baroreceptors in dogs. *Circulation Research*, 48: 676-684.
- Coleridge, H.M., Coleridge J.C.G., and Schultz H.D. 1987. Characteristics of C fibre baroreceptors in the carotid sinus of dogs. *J Physiol*, 394: 291-313.
- Chen, F., Zhang Y.T., and Zhang C., 2004. An integrate-and-fire based baroreceptor model. *IEEE / EMBS International Summer School on Medical Devices and Biosensors*.
- Eckberg, DL., 1977. Adaptation of human carotid baroreceptor cardiac reflex. *J Physiol, London, UK*, 269: 579-589.
- Evans, W.A., 1994. Approaches to intelligent information retrieval. *Inf Proc and Managem*, 7: 147–168.
- Franz G.N., Scher A.M., Ito C.S., 1971. Small signal characteristics of carotid sinus baroreceptors of rabbits. *J Appl Physiol*, 30: 527-535.
- Fung Y.C., 1993. *Biomechanics: mechanical properties of living tissues*. Springer Verlag, New York, NY.
- Fung Y.C., 1996. *Biomechanics: circulation*. Springer Verlag, New York, NY.
- Hasan Z., 1983. A model of spindle afferent response to muscle stretch. *J Neurophysiol*, 49: 989-1006.

- Houk J., Cornew R.W., and Stark L., 1966. A model of adaptation in amphibian spindle receptors. *J Theor Biol*, 12: 196-215.
- Izhikevich, E.M. 2007. *Dynamical systems in neuroscience. The geometry of excitability and bursting*. MIT press, Cambridge, MA.
- Kalita, P. and Schaefer R., 2008. Mechanical models of artery wall. *Arch Comput Methods Eng*, 15: 1-36.
- Koch, C and Segev I. (eds.), 1998. *Methods in neuronal modeling: from ions to networks*. MIT press, Cambridge, MA.
- Kunze D.L., Saum W.R., and Brown A.M., 1977. Sodium sensitivity of baroreceptors mediates reflex changes of blood pressure and urine flow. *Nature*, 267: 75-78.
- Landgren, S., 1952. On the excitation mechanism of the carotid baroreceptors. *Acta Physiol Scand*, 26: 1-34.
- Landgren, S. 1952. The Baroreceptor activity in the carotid sinus nerve and the distensibility of the sinus wall. *Acta Physiol Scand*, 26: 3556.
- Levick, J.R., 2010. *An introduction to cardiovascular physiology*. Hodder Arnold, New York, NY.
- Olufsen M.S., Tran, H.T., Ottesen J.T., Lipsitz, L.A., and Novak, V., 2006. Modeling baroreflex regulation of heart rate during orthostatic stress. *Am J Physiol*, 291: R1355-R1368.
- Seagard J.L., van Brederode J.F., Dean F.A., Hopp F.A., Gallenberg L.A., and Kampine J.P., 1990. Firing characteristics of single-fibre carotid sinus baroreceptors. *Circulation Research*, 66: 1499-509.
- Sleight, P., Robinson, J. L., Brooks, D. E., and Rees, P. M., 1977. Characteristics of single carotid sinus baroreceptor fibers and whole nerve activity in normotensive and renal hypertensive dog. *Circulation Research*, 41: 750758.
- Sherwood, L. 2001. *Human physiology: from cells to systems*. 7th edition. Brooks Cole, Belmont, CA.
- Spickler, J. W. and Kezdi P. 1967. Dynamic response characteristics of carotid sinus baroreceptors. *Am J Physiol*, 212: 472476.
- Srinivasan, R. and Nudelman, H.B. 1972. Modeling carotid sinus Baroreceptor. *Biophysical J*, 12: 1171-1182.
- Srinivasan, R. and Nudelman, H.B., 1973. Theoretical studies on behavior of carotid sinus baroreceptors. *Kybernetik*, 13: 144150.
- Taher, M.F., Cecchini, A.B., Allen, M.A., Gobran, S.R., Gorman, R.C., Guthrie, B.L., Lingenfelter, K.A., Rabbany, S.Y., Rolchigo, P.M., Melbin, J., and Noordergraaf, A., 1988. Baroreceptor responses derived from a fundamental concept. *Ann Biomed Eng*, 16: 429443.
- Thoren, P., Saum, W.R., and Brown, A.M., 1977. Characteristics of rat aortic baroreceptors with nonmedullated afferent nerve fibers. *Circulation Research*, 40: 231237.
- Ursino, M., 1999. A mathematical model of the carotid baroregulation in pulsating conditions. *IEEE Trans Biomed Eng*, 46: 382-392.
- Van Vliet, B. and West N.H., 1987. Response characteristics of pulmocutaneous arterial baroreceptors in the toad, *bufo marinus*. *J Physiol*, 388: 55-70.
- Van Brederode, J.F., Seagard, J.L., Dean, C. Hopp, F.A. and Kampine, J.P., 1990. Experimental and modeling study of the excitability of carotid sinus baroreceptors. *Circulation Research*, 66: 15101525.
- Valdez-Jasso, D., Bia, D., Zocalo, Y., Armentano, R.L., Banks, H.T., Haider, M.A., and Olufsen, M.S., 2009. Viscoelastic models for passive arterial wall dynamics. *Adv Appl Math Mech*, 1: 151-165.
- Valdez-Jasso, D., Haider, M.A., Banks, H.T., Bia, D., Zocalo, Y., Armentano, R.L., and Olufsen, M.S., 2009. Analysis of viscoelastic wall properties in ovine arteries. *IEEE Trans Biomed Engr*, 56: 210-219.
- Valdez-Jasso, D., Bia, D., Zocalo, Y., Armentano, R.L., Haider, M.A., and Olufsen, M.S., 2011. Linear and nonlinear viscoelastic modeling of aorta and carotid pressure-area. *Ann Biomed Eng*, 39: 1438-1456.
- Wang W., Chen J.S., and Zucker I.H., 1991. Postexcitatory depression of baroreceptors in dogs with experimental heart failure. *Am J Physiol*, 260: H1160-1165.

GLOBAL SENSITIVITY AND IDENTIFIABILITY ANALYSIS APPLIED TO A MODEL PREDICTING BAROREFLEX REGULATION DURING HEAD-UP TILT

Christian Haargaard Olsen^(a), Jesper Mehlsen^(b), Johnny T. Ottesen^(c), Hien T. Tran^(d),
Mette S. Olufsen^(e)

^(a,d,e) Department of Mathematics, North Carolina State University, Raleigh, NC

^(b) Coordinating Research Center, Frederiksberg Hospital, Denmark

^(c) Department of Science, Systems, and Models, Roskilde University, Roskilde Denmark

^(a) chaarga@ncsu.edu, ^(b) Jesper.Mehlsen@frh.regionh.dk, ^(c) johnny@ruc.dk, ^(d) tran@math.ncsu.edu,
^(e) msolufse@ncsu.edu

ABSTRACT

Mathematical modeling has long been used for prediction of biological phenomena, but practical use of mechanistic models for data analysis is not as common. One of the major obstacles is the lack of practical methods for model based data analysis. This study presents methods for predicting global sensitivities of model outcomes to model parameters and shows how global sensitivities can be used for predicting sets of globally identifiable parameters. These methods will be applied to analyze dynamics in a model designed to predict baroreceptor regulation of heart rate during head-up tilt (HUT).

Keywords: Global sensitivity analysis; Global parameter identifiability; Baroreflex regulation of heart rate; Head-up tilt.

1. INTRODUCTION

The baroreflex control system, part of the autonomic control system, is responsible for the regulation of the blood pressure and thus for organ perfusion, which is especially important in the brain. Dysfunctional autonomic regulation is associated with postural hypotension and syncope, which has negative impact on quality of life. This problem is often observed (Brignole 2001), yet few methods are available for analysis of clinical data routinely recorded in the laboratory.

The baroreceptor reflex system consists of a number of elements: Stretch sensitive baroreceptors in the wall of the carotid sinus stimulate firing of afferent nerves, which are integrated primarily in the nucleus of the solitary tract. At this location several stimuli are integrated causing modulation of parasympathetic and sympathetic outflow, which in turn determines release of neurotransmitters enabling modulation of heart rate. Several models have investigated this process (Ottesen and Olufsen 2011), though few have been successfully used for analysis of clinical data.

This paper discusses a model adapted from previous studies that predicts the heart rate, global sensitivity analysis and parameter identification. Global

sensitivities will be calculated as the average of the local sensitivities over the entire parameter space. The latter will be computed using a quasi-random Monte Carlo method based on Sobol's algorithm (Joe and Kuo 2003). Sensitive parameters will be further analyzed to predict a set of parameters that are globally identifiable. The latter requires prediction of correlations among model parameters. Two methods will be used for this part: analysis of pairwise correlations (Olufsen and Ottesen 2012) and a method that combines a principal component analysis (PCA) with analysis of orthogonality (Li, Henson and Kurtz 2004).

Sensitivity and identifiability techniques developed here will be applied to analyze measured blood pressure from a healthy young adult undergoing a head-up tilt table test.

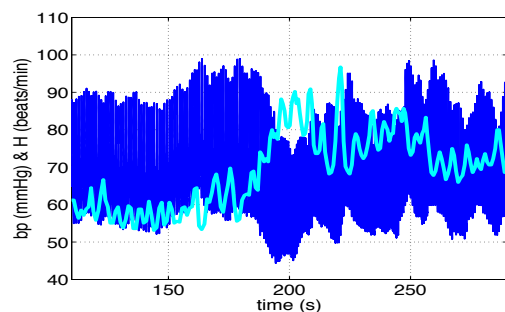


Figure 1: Blood pressure (bp) [mmHg] and heart rate (H) [beats/min] recorded during HUT. The displayed data are from a young healthy adult.

2. MATHEMATICAL MODEL

The mathematical model developed in this study incorporates the parts of the baroreflex feedback system that modulates heart rate. The model (shown in Figure 2) is an extension of previous a previous model (Olufsen, Tran, et al. 2006).

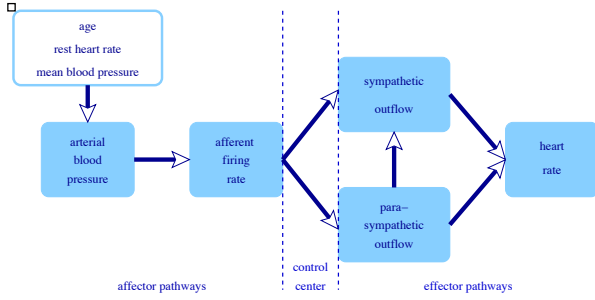


Figure 2: A schematic view of the model components included to predict baroreflex regulation of heart rate during HUT.

In summary, the model uses blood pressure data along with information about the subjects' age, gender, resting heart rate, and mean blood pressure as inputs to predict heart rate.

First, the firing rate (n) of the baroreceptors n is predicted by

$$(1) \quad \frac{dn}{dt} = k_i \frac{dp}{dt} \frac{n(M-n)}{(M/2)^2} - \frac{n-N}{\tau_i}, \quad n = \sum n_i,$$

where n_i is the firing rate of the baroreceptor of type i , M is the maximum firing rate and N is the threshold firing rate, dp/dt is the rate of change of blood pressure, k_i is a constant determining how sensitive the receptor is to changes in blood pressure, and τ_i is a time constant that determines the rate at which the firing rate decays to the threshold value.

Parasympathetic and sympathetic outflows are predicted as

$$(2) \quad T_p = \frac{n}{M}, \quad T_s = \frac{1-T_{p,d}}{1+\beta T_p},$$

where β is a constant controlling the relationship between the parasympathetic T_p and the sympathetic T_s outflow. $T_{p,d}$ is the delayed parasympathetic outflow given by

$$(3) \quad T_{p,d} = c \int_{-\infty}^t (t-s)^D e^{-\alpha(t-s)/t_d} T_p(s) ds,$$

where D and α , an integer and a real number, determines the shape of weight function for the signal, t_d is the time for the largest effect, and c is a normalization constant determined from D and t_d . This formulation is different than what is found in the literature. It has the advantage of easier numerical implementation and reflecting a distributed delay, which by the authors is believed to be more physiologically correct than a discrete delay.

The concentration of the neurotransmitters acetylcholine C_a and noradrenaline C_n are predicted from

$$(4) \quad \frac{dC_a}{dt} = \frac{T_p - C_a}{\tau_a}, \quad \frac{dC_n}{dt} = \frac{T_s - C_n}{\tau_n},$$

where τ_a and τ_n are time constants. Finally, the build up of electrical potential (ϕ) in the pacemaker cells of the heart is determined by

$$(5) \quad \frac{d\phi}{dt} = H_0 (1 - M_a C_a + M_n C_n),$$

where H_0 denotes the default depolarization rate when no neurotransmitters impact the system, and M_a , M_n denote the sensitivity to the neurotransmitters. From this expression the heart rate can be calculated, assuming that each heartbeat corresponds to an increase in the potential build up by 1.

3. SENSITIVITY ANALYSIS

Classically, sensitivities can be calculated as

$$(6) \quad S(t; \theta) = \frac{\partial y(t; \theta)}{\partial \theta},$$

where y denotes the model output (heart rate) and θ the model parameters. For linear (in the parameters) models, sensitivities can be predicted analytically, but for nonlinear models this relation is typically predicted by linearizing around current values of the parameters. For a model that depends nonlinearly on the parameters the sensitivities will change depending on the parameter values. This can produce challenges since the sensitivities are important both for determining which parameters are identifiable and for estimating parameter values that allows accurate prediction of the data. One way to get more precise information about the sensitivity is via prediction of global sensitivities, which estimate how the model output vary as an average over the parameter space. A possible approach for predicting global sensitivities is via Sobol's global sensitivity indices, which builds on an analysis of variances (ANOVA) (Kiparissides, et al. 2009). Predicting global sensitivities is computational intensive, due to the large number of evaluations of the model output. Oftentimes computational efficiency is of high importance. One method to improve efficiency is to use derivative based global sensitivity measures (DGSM), which are computed by integrating sensitivities over the parameter space (Kiparissides, et al. 2009).

Calculation of DGSM requires computation of local sensitivities, which in this study are predicted from the model and parameter Jacobians. The latter are calculated using finite differences. Both the ODEs representing the model and the local sensitivities are computed numerically using Sundials ODE suite (Serban and Hindmash 2005). Within this software, the numerical integration over the parameter space is carried out using a quasi-random Monte Carlo method. Parameter sets are generated using Sobol's algorithm, and the global sensitivities are computed as the average of the local sensitivities.

More specifically, the local sensitivities are calculated by solving the system of ODE

$$(7) \quad \begin{aligned} \frac{dS}{dt} &= J_y S(t, \theta) + J_\theta, \\ S(t, \theta) &= \frac{dy}{d\theta}(t, \theta), \quad J_y = \frac{df}{dy}, \quad J_\theta = \frac{df}{d\theta}, \end{aligned}$$

where $y(t, \theta)$ is the model output of the system with state variables $x(t, \theta)$ described by

$$(8) \quad \begin{aligned} \frac{dx}{dt} &= f(t, x; \theta), \quad x(0; \theta) = x_0, \\ y(t, \theta) &= g(t, x(t); \theta). \end{aligned}$$

In (7-8) J_y denotes the Jacobian and J_θ the parameter Jacobian. If the model consists of k output variables, m parameters that are evaluated at n time steps the local sensitivity matrix can be written as

$$(9) \quad S = \frac{dy}{d\theta} = \begin{bmatrix} \partial y_1 / \partial \theta_1(t_1) & \cdots & \partial y_1 / \partial \theta_m(t_1) \\ \vdots & \vdots & \vdots \\ \partial y_1 / \partial \theta_1(t_n) & \cdots & \partial y_1 / \partial \theta_m(t_n) \\ \vdots & \vdots & \vdots \\ \partial y_k / \partial \theta_1(t_1) & \cdots & \partial y_k / \partial \theta_m(t_1) \\ \vdots & \vdots & \vdots \\ \partial y_k / \partial \theta_1(t_n) & \cdots & \partial y_k / \partial \theta_m(t_n) \end{bmatrix}.$$

Sometimes it might be relevant to calculate relative sensitivities defined by

$$(10) \quad \tilde{S}_{i,j}(t_k) = \frac{\partial y_i}{\partial \theta_j}(t_k) \frac{\theta_j}{y_i(t_k)}.$$

Once, local sensitivities (relative or nonrelative) are computed, the global sensitivities can be found by integration over the parameter space. This is done using Sobol-sequences for sampling parameter sets within the parameter space. For each parameter set local sensitivities are computed. Global sensitivities are then found as an average of the sensitivity matrix calculated at N different points in parameter space, i.e.

$$(11) \quad S_{global}(t_k) = \frac{1}{N} \sum_{i=1}^N S_i(t_k).$$

4. IDENTIFIABILITY AND SUBSET SELECTION

Only sensitive parameters can be identified, but in addition to being sensitive parameters may be correlated (Ipsen, Kelley and Pope 2011). Parameters that are correlated cannot be estimated reliably. However, it may not be trivial to identify parameter correlations globally. In this study, we denote the process of identifying parameters that can be estimated given a model output and corresponding data for subset selection. Parameters not in a subset (i.e., those that are insensitive or correlated) will be kept constant at their *a priori* values,

while the uncorrelated sensitive parameters will be estimated as discussed in section 5.

This study adapts two previously developed methods for prediction of global sets of uncorrelated parameters: the structured correlation method (Olufsen and Ottesen 2012) and the orthogonal sensitivities method (Yao, et al. 2003, Li, Henson and Kurtz 2004). Both are originally developed for local analysis, but will be modified to work with global sensitivities. The structured correlation matrix method is based on finding pairwise correlations of the model Hessian, while the orthogonal method is based on analysis of parameters sensitivities predicted using PCA. With this method, parameters are ranked according to sensitivity from the highest to the lowest. Parameters are added to the subset one-by-one ensuring that the subset remains orthogonal using a process is similar to Gram-Schmidt orthogonalization. Subsets obtained with these methods will be compared and an ‘‘optimal’’ subset will be selected.

4.1. Structured correlation matrix method (SCMM)

The structured correlation matrix method (SCMM) is based on a principle of excluding correlated parameters from the set of parameters until no correlations remain. To evaluate the correlations the *correlation matrix*

$$(12) \quad c_{i,j} = \frac{C_{i,j}}{\sqrt{C_{i,i} C_{j,j}}}$$

is considered (Ottesen and Olufsen 2011). Here $c_{i,j}$ describes the degree of correlation between parameter i and j , with numeric value between 0 and 1. C is the correlation matrix, the inverse of the model Hessian, $C = \mathcal{H}^{-1}$, which is given by $\mathcal{H} = \sigma^{-2} S^T S$. The inversion of the model Hessian requires that S be non-singular. In this study the limiting value γ between correlation and no correlation for c is chosen to be $\gamma = 0.9$. When a correlated pair is found the least sensitive parameter is excluded from the subset, judged by the two-norm of the sensitivity,

$$(13) \quad \bar{S}_j = \sqrt{\sum_{i=0}^n [S_j(t_i)]^2}.$$

Using these measures the procedure can be described:

1. Calculate the correlation matrix c .
2. Find the parameter pair (i,j) with the highest correlation.
3. If $c_{i,j} \geq \gamma$ lock the parameter with the smallest two-norm \bar{S} . If $c_{i,j} < \gamma$ terminate the algorithm.
4. Repeat from 1.

4.2. Orthogonal sensitivities method (OSM)

In contrast to SCMM this method is based on the principle of building a subset by adding one parameter at the time. Initially the parameters are ranked by their *importance index*, which for parameter j is given by

$$(14) \quad E_j = \frac{\sum_{i=1}^m |\lambda_i Q_{i,j}|}{\sum_{i=1}^m |\lambda_i|},$$

where m is the number of parameters, and λ_i is the i th eigenvalue and $Q_{i,j}$ the j th eigenvector of the matrix

$$(15) \quad X = \tilde{S}^T \tilde{S}$$

with s denoting the relative sensitivities defined in (10). If n parameters have been selected for the subset θ_s , the *identifiability index* of parameter j is given by $I_j = E_j d_j$, where d_j is the *orthogonality index*. Denoting the column of \tilde{S} corresponding to the l th selected parameter \tilde{s}_{k_l} and assuming that \tilde{s}_{k_l} are linearly independent any vector in the n -dimensional vector field can be represented by

$$(16) \quad \tilde{s} = \sum_{l=1}^n a_{k_l} \tilde{s}_{k_l},$$

where a_k are constants. Consider another parameter h with relative sensitivity \tilde{s}_j , which has not been included in the subset, the objective is to find vector \tilde{s} closest to \tilde{s}_j by determining the coefficients a_{k_l} that solves

$$(17) \quad \min_{a_{k_l}} \frac{1}{2} (\tilde{s}_j - \tilde{s})^T (\tilde{s}_j - \tilde{s}).$$

This is equivalent to solving the linear system

$$(18) \quad \begin{bmatrix} \tilde{s}_{k_1}^T \tilde{s}_{k_1} & \cdots & \tilde{s}_{k_n}^T \tilde{s}_{k_1} \\ \vdots & \ddots & \vdots \\ \tilde{s}_{k_n}^T \tilde{s}_{k_n} & \cdots & \tilde{s}_{k_n}^T \tilde{s}_{k_n} \end{bmatrix} a = \begin{bmatrix} \tilde{s}_j^T \tilde{s}_{k_1} \\ \vdots \\ \tilde{s}_j^T \tilde{s}_{k_n} \end{bmatrix}.$$

Since the \tilde{s}_k are chosen to be linearly independent, the covariance matrix for the selected parameters is invertible and there is a solution for a . Knowing the vector \tilde{s} closest to \tilde{s}_j the orthogonality index is calculated as sine of the angle between them

$$(19) \quad d_j = \sin \left[\arccos \left(\frac{\tilde{s}_j^T \tilde{s}}{\|\tilde{s}_j\| \|\tilde{s}\|} \right) \right].$$

The procedure for ordering the parameters can then be described as follows.

1. Calculate the Importance index. Selected the parameter with the highest value for the subset.
2. Calculate the orthogonality index for the remaining parameters compared to the current subset.
3. Use the orthogonality index d_j with the importance index E_j to calculate the identifiability index I_j .
4. Add the parameter with the highest identifiability index to the subset. Repeat from step 2.

5. RESULTS

The results presented in this section are based on blood pressure data obtained from a Head-Up Tilt experiment.

5.1. Global sensitivities

The calculation of the global sensitivity estimate is based on Sobol's algorithm for integration using 5000 points in the parameter space and the intervals for the parameters given in Table 1. 5000 points was used as a preliminary investigation showed that including more points did not change the results.

The two-norm of the global sensitivity for each parameter is used to rank the parameters from most to least sensitive. This is illustrated in Figure 3.

Parameter	Min. value	Max. value
H_0	0.3	2.0
M_s	1.25	5.0
M_p	0.07	0.28
τ_{ach}	0.1	1.0
τ_{nor}	0.1	1.0
β	3.0	12
κ_1	3	30
κ_2	3	30
κ_3	3	30
τ_1	0.25	1.0
τ_2	2.5	10
τ_3	250	1000
t_d	3.0	12
$T_{par,0}$	0.3	1.0
a	1.5	1.5

Table 1: Intervals defining the parameter space. Intervals are based on the parameter values found by (Olufsen, Tran, et al. 2006).

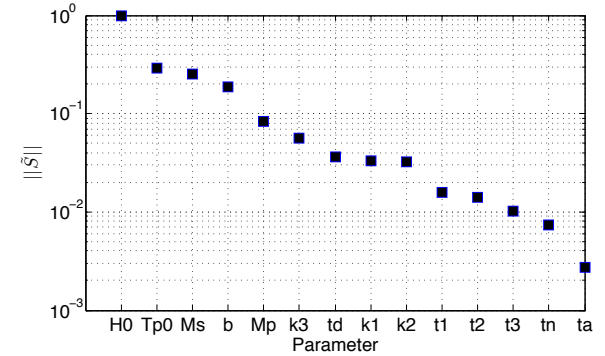


Figure 3: Parameters sorted by two-norm of the global sensitivity. Values have been scaled by the maximum value.

5.2. Subset selection

SCMM and OSM have been used to perform the subset selection. For both methods the global relative sensitivities were used. Furthermore, both analyses were performed without the possibility of including τ_3 in the subset. This parameter reflects a very long timescale in the model on the order of 500 seconds. Hence it is unlikely that this parameter can be determined from data ranging over approximately 700 seconds.

Using SCMM introduced above and a limit for correlations of 0.9 the resulting subset is

$$(20) \quad \theta = \{H_0, \kappa_1, \tau_1, \tau_2, t_d, T_{p0}\}.$$

It is worth noting that this is just one subset one can obtain using structured correlation matrix method. Other variations of the algorithm will lock several parameters at each iteration, lock the most sensitive one, or use another threshold for deciding if a pair of parameters are correlated or not.

Using OSM the parameters have been ranked as shown in Table 2. Setting the limit value of the identifiability index at 0.005 the resulting subset is

$$(21) \quad \theta = \{H_0, M_s, M_p, \kappa_2, \kappa_3, t_d, T_{p0}\}.$$

Iteration	Parameter	E_j	d_j	I_j
0	H_0	—	—	0.910
1	T_{par0}	0.259	0.354	0.092
2	M_s	0.229	0.165	0.038
3	κ_2	0.020	0.672	0.013
4	κ_3	0.040	0.151	0.006
5	t_d	0.007	0.767	0.005
6	M_p	0.075	0.061	0.005
7	κ_1	0.025	0.181	0.004
8	β	0.170	0.020	0.003
9	τ_1	0.012	0.287	0.003
10	τ_3	0.008	0.154	0.001
11	τ_2	0.002	0.231	0.000
12	τ_{nor}	0.000	0.325	0.000
13	τ_{ach}	0.000	0.290	0.000

Table 2: Parameters ordered by the identifiability index of the orthogonal sensitivities method.

It is interesting to see that the two subsets are quite different. Most notably SCMM omits the parameters M_p and M_s , which one would expect to be very important as they appear directly in the differential equation describing the potential, build up (5). Investigating the SCMM correlations in the subset produced by OSM reveals an expected correlation between the parameters H_0 and M_p .

SCMM also finds that the parameters κ_1 and κ_2 are correlated, so the fact that one of these appear in each subset is reasonable. Finally SCMM finds that κ_3 is correlated to T_{p0} , while OSM finds that these are distinguishable. It is possible to obtain different subsets within each method, but these subsets appear to be very different. This difference probably arises due to the fact that one method investigates the sensitivity of the solution to the parameter values (OSM), while the other method utilizes the inverse, the sensitivity of the parameter value to the change in model output.

While the differences are interesting it is also worth noting the parameters shared by the two subsets, H_0, t_d and T_{p0} . H_0 is a direct scaling of the potential buildup in the model, t_d is the delay associated with the sympathetic nervous system and T_{p0} is the threshold value for the firing rate of the baroreceptors. It seems

very reasonable that both methods should find these parameters to be important and identifiable.

CONCLUSION

Global sensitivities have been calculated using Sobol's algorithm for quasi-random number generation and subset selection performed using two different methods, the structured correlation matrix method (SCMM) and the orthogonal sensitivities method (OSM).

While the two methods for subset selection produced different subsets the parameters shared by both subsets appear to be very important for the model output and thus clearly identifiable.

Future work in this area will include testing the different parameter subsets through optimization against simulated data as well as real data.

ACKNOWLEDGMENTS

All authors were supported in part by NSF-DMS under grant #1022688. Additionally Olufsen was supported by NIH-NIGMS under 1P50-GM094503-01A1 and Tran was supported by NIH/NIAD under 9 R01-AI071915-05.

REFERENCES

- Brignole, M. et al., 2004. Guidelines for management (diagnosis and treatment) of syncope – update 2004. *Eur Heart J*, 22: 2054-2072.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: Efficient adaptive MCMC. *Stat. Comput.*, 16: 339-354.
- Ipser, I.C.F, Kelley, C.T., Pope, S.R., 2011, Rank-deficient nonlinear least squares problems and subset selection. *Siam J. Num. Anal.*, 49: 1244-1266.
- Joe, S., Kuo, F.Y., 2003. Remark on algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Trans Math Software*, 29: 49-54
- Kiparissides, A., Kuchenrenko, S.S., Mantalari, A., Pistokopolous E.N, 2009. Global sensitivity analysis challenges in biological systems modeling. *Ind Eng Chem Res*, 48: 7168-7180.
- Li, R., Henson, M.A., Kurt, M.J., 2004. Selection of model parameters for off-line parameter estimation. *IEEE Trans Ctrl Sy Technol*, 12: 402-412.
- Miao, H., Xia, X., Perelson, A.S., Wu, H., 2011, On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev*, 53: 3-39
- Olufsen, M.S., Tran, H.T., Ottesen, J.T., Lipsitz, L.A., Novak, V., 2006. Modeling baroreflex regulation of heart rate during orthostatic stress. *Am J Physiol*, 291:R1355-R1368.
- Olufsen, M.S., Ottesen, J.T., 2012. A practical approach to parameter estimation applied to model predicting heart rate regulation. *J Math Biol*, in press.
- Ottesen, J.T., Olufsen, M.S., 2011. Functionality of the baroreceptor nerves in heart rate regulation. *Comp Meth Prog Biomed*, 101: 208-219

- Serban, R., Hindmarsh, A.C., 2005. CVODES, the Sensitivity-enabled ODE solver in SUNDIALS. *Proc ASME Int Design Eng Technl Conf, DETC2005-85597*: 257-269
- Sobol, I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Math Phys*, 7:784-802
- Solonen, A., 2006. Monte Carlo methods in parameter estimation of nonlinear models. *Master's Thesis - Department of Information Technology - Laappennanta University of Technology, Laappennanta.*
- Yao, K.Z., Shaw, B.M., Kou, B., McAuley, K.B., Bacon, D.W., 2003. Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. *Polym React Eng*, 11:563-588

MODELLING HEALTH CARE UTILISATION: A METHOD COMPARISON

Stephanie Parragh^(a), Patrick Einzinger^(b)

^(a)Institute for Analysis and Scientific Computing, Vienna University of Technology, Austria

^(b)dwh Simulation Services, Austria

^(a)stephanie.parragh@tuwien.ac.at, ^(b)patrick.einzinger@drahtwarenhandlung.at

ABSTRACT

Models of health care utilisation help to identify influencing factors and to study the possible impact of different strategies. In addition to static analyses, dynamic models can be used to perform simulation experiments, which yield insights into the dynamical behaviour of the system. Various methods exist to establish a dynamic model. Two widely used, but substantially differing approaches are system dynamics and agent-based modelling. Their benefits and drawbacks for the application in this field are outlined and compared. To demonstrate their differences and analogies under practical application, a simplified model of health care utilisation is established and the implementation as agent-based and as system dynamics model is described. Then, the results of simulation runs are presented and discussed with regards to both the comparison of the two methods and their possible meaning.

Keywords: system dynamics, agent-based modelling, method comparison, health care utilisation

1. INTRODUCTION

Models describing the utilisation of health care services by patients are an important tool to understand and to test the impacts of different strategies concerning for example provider payment or insurance coverage. How to establish such a model and what factors are to be included depends mainly on the problem under investigation.

For the comparison of different reimbursement systems for instance, models based on game-theoretic considerations can be employed (Ma and McGuire 1997). The consequences of price shifts on the other hand can be studied with the help of economical models. Thereby, the system is described in terms of demand and supply of the idealised good “health”, influenced by the interests and preferences of patients and providers respectively (Thode et al. 2004). These models are also used to define, explain and analyse the notion of physician induced demand (Kern 2002).

Behavioural models of the social sciences represent another approach to explain the process of health care utilisation. They focus on the factors that

influence why, when and how often medical services are used. Various different variants exist. One example is Andersen’s Behavioral Model of Health Services Use (Andersen 1995), which has proven its worth as reference model many times (Thode et al. 2004).

All of these models offer a description of the system that can be used for static analyses based on statistical data. Thereby, the impacts of different factors can be studied and new trends can be identified. However, since the utilisation of health care services is a dynamic process, also the investigation of dynamic phenomena present in the system is of interest. For this purpose a dynamic model is needed and an appropriate modelling method has to be chosen.

Agent based modelling (ABM) and system dynamics (SD) are both powerful and widely used methods to model complex dynamic systems. The aim of this study is to examine their advantages and disadvantages for modelling health care utilisation. Therefore, a demonstration model is established and implemented in both methods. After a theoretical comparison of the two techniques, this model is briefly introduced and simulation results are presented.

2. MODELLING METHODS

In agent based modelling the constituent units of a system are identified and modelled as autonomous, decision-making entities called agents (Bonabeau 2002). The global behaviour of the system is not determined a priori, but it results from the (often local) interaction of the agents during simulation, which is why ABM is called a bottom-up approach. In system dynamics on the contrary, the system is described from top down, which means that the global relations and feedback mechanisms between the system components are modelled. Figure 1 shows the concept of a typical agent, cf. Macal and North (2010). Each agent normally has a set of attributes, which can be static, as for example its name, or dynamic, e.g. memory or age. Its behaviour results from methods, which specify how it acts, reacts and interacts. The main structure of an SD model according to J. W. Forrester, who developed this technique, is depicted in Fig. 2. In an SD model, the system is represented as a stock and flow diagram. The different stocks (or levels) correspond to the states of

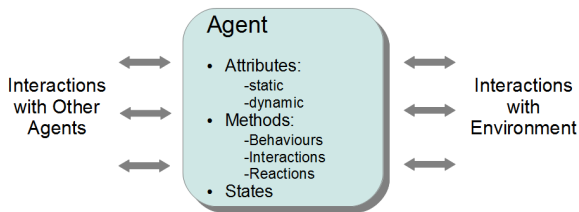


Figure 1: A Typical Agent.

the system. Their change over time is given by the difference of the in- and the outflow, the size of which depends on the information about the other stocks as well as the strategies and decisions considered.

Basically, health care utilisation corresponds to interactions between patients, who decide to use a health care service, and medical providers. Hence the acting entities of the system are individuals, which makes their representation as agents obvious. Thode et al. (2004) identified age, sex, regional aspects, subjective perception of health state and morbidity as the main influences on health care utilisation on the part of the patients. In an AB model, these individual factors can be included in a natural way as attributes of the agents. The regional aspects—or spatial arrangements in general—can be realised by defining an environment in which the agents live. For that reason, ABM is also suitable for analyses regarding the regional provision of medical services (Romstorfer et al. 2011).

SD models are based on the idea that all dynamics occur due to the accumulation of flows in stocks (Radzicki and Taylor 1997). Each stock thereby consists of homogeneous elements, thus distinctions within the elements, i.e. heterogeneities of any kind, have to be modelled by adding new stocks, representing the wanted characteristics. First of all this highly increases the complexity of the model and second this also affects the global model structure. Thus, SD does not provide the appropriate means to depict individual differences, such as varying patient preferences or variable levels of disease risk depending, for instance, on a patient's life style or his genetic disposition. Also, refinements are difficult to incorporate into an existing model subsequently, since they alter its structure. In an AB model on the contrary, extending the properties or methods of one type of agent or even adding a new type causes only local changes, i.e. the current model structure is preserved.

Moreover, dynamics caused by network effects are hard to capture by the means of system dynamics. These are deviations from the predicted aggregate behaviour caused by heterogeneous interaction patterns. Network effects may occur, whenever interactions between the modelled objects have an impact on the state of the system. This is for instance the case with infectious diseases. In SD the change of a stock over time is represented by an average, aggregated flow. As the elements within the stock are indistinguishable, this flow is based on the assumption that interactions between them as well as the distribution of information

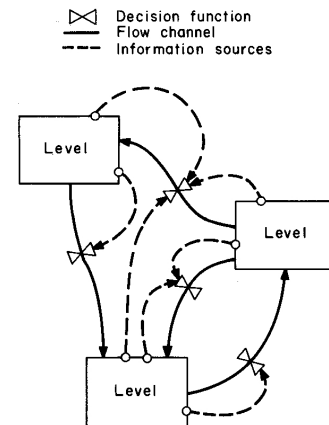


Figure 2: The Main Structure of an SD Model (taken from Forrester, 1961).

among them is homogeneous. Thus, effects due to a heterogeneous mixing are ignored. In the agent-based approach, relations between the agents, representing for example a social environment or a spatial proximity, can be easily established to specify the mode of interaction (Boneabeau 2002).

This clearly shows that ABM is advantageous over SD in modelling dynamics due to heterogeneities both in the population itself and in the way interactions take place.

One of the benefits of SD compared to ABM is that it offers a more standardised practice of modelling. An SD model can be developed from a few standard elements and the resulting stock and flow diagram automatically provides a graphical representation of the system. Furthermore, the implementation of such a model requires no, or at most little, programming knowledge, as different software tools are available that offer graphical modelling environments for this purpose.

Similar toolkits, which provide pre-implemented structures to simplify the implementation, are available for ABM as well. However, there are neither specific guidelines nor a fixed set of standard elements for establishing an AB model. For this reason no software tool is able to cover the broad range of possibilities regarding the actual realisation. Hence the implementation still requires programming knowledge. The lack of standardised ways of proceeding makes it moreover difficult for the modeller to decide on his own strategy, especially, since effects at a system-level can usually not be determined prior to simulation. SD, in contrast, approaches a system by first identifying the cause and effect relationships between the system variables and the thereby occurring feedback loops. Hence the process of modelling itself can already yield insights into the qualitative properties.

The computational power needed for simulation as well as the data requirements represent another difference between the two methods. In general, SD simulation needs less computation time, as an SD model mathematically corresponds to a system of differential

equations, which can be solved numerically in an efficient way. Simulating an AB model means simulating each agent, its interactions, reactions and actions at each discrete time step and/or event, which is computationally intensive. Especially if sensitivity analyses and/or variations of a number of parameters are needed, this constitutes a disadvantage in terms of the time required.

Regarding the data needed for parametrisation, the requirements of SD are possibly easier to meet. It is based on aggregated data and overall tendencies. These can often be obtained from national institutes. The type of data needed for the parametrisation of an AB model depends on first, what the agents represent and second, the wanted level of detail regarding their attributes and individualities. For instance, when modelling a disease, it can either be assumed that its course is the same for each individual or that variations concerning its duration or severity occur, as is usually the case in reality. Both options can be realised easily in an AB model. However, whereas averaged data can be used for the first one, the implementation of the second one requires disaggregated, individual data, which is often not available and, especially in the health sector, privacy protection can be an issue. Thus, since the capacity to take heterogeneities into account is one of the biggest advantages of ABM, data can be a limiting factor in using it to its full potential.

Even though the two methods show substantial differences, not least because in SD a system is modelled continuously whereas an AB model is in general discrete (although continuous elements may be included as well), their field of application overlaps and in some cases, the same system behaviour can be obtained with both methods. Borshchev and Filippov (2004) described a way to translate an existing SD model into a corresponding agent-based one. Figure 2 depicts the main idea of their strategy. First, the elements within a stock are taken as agents and the different stocks as their possible states, which are represented in a state chart. Thus, the current size of a stock corresponds to the number of agents in the same state (see Fig. 2 b) Then, the transitions between the states are specified, using the same conditions that determine the size of the flows. For a simple example of bass diffusion, Borshchev and Filippov demonstrated that the thereby obtained results indeed coincide.

3. DEMONSTRATION MODEL

Our model aims to describe the dynamics of health care utilisation in a simplified system, which consists of patients, who use, and providers, who offer medical services. All providers are assumed to have the same specialisation and only one type of disease with two different severity levels is considered. The patients consult a provider for regular check-ups as well as in case of disease. A comparable equivalent in real life would be for example the utilisation of dental services, where a check-up every 6 months is recommended. For

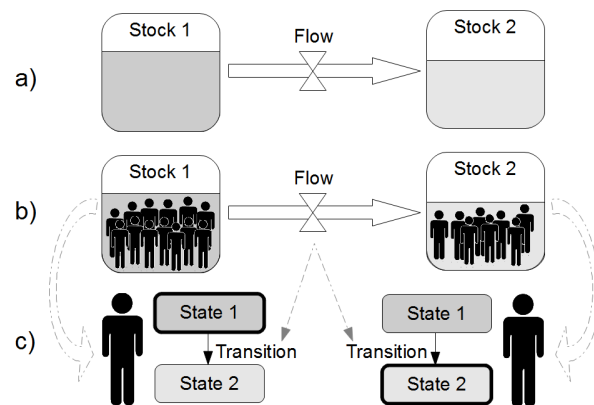


Figure 2: Translation of a Stock and Flow Diagram Into a Corresponding Agent-Based Representation.

female patients, a similar situation occurs with regards to gynaecological examinations.

Both, medical providers and patients, are supposed to have different goals, which they try to attain. The goal of the patients is to maintain a high quality of life, i.e. they are aiming for a fast recovery. For the goal of the providers, a mix between profit and patient's benefit was chosen.

A provider can offer two different levels of services, whereby one (treatment B) is considered to be an extension of the other one (treatment A) and therefore works on both severity levels, but takes longer and also costs more. Moreover he is able to diagnose the patient's severity level with a certain error rate. Providers are assumed to care about profit, which depends on the reimbursement system, and the patient's benefit from treatment, which is normalised to 1 if it was effective and 0 otherwise. In order to maximize their performance, they periodically adjust a diagnosis related threshold above which they prescribe the extended treatment to a patient.

To make the two models comparable, the same parameters and variables are used whenever possible and diagnosis, treatment decision as well as the optimisation process are realised by the same means. Both models are implemented in the multi-method simulation modelling tool AnyLogic (6.5.1) by XJ Technologies.

3.1. Implementation as AB Model

Patients and providers are modelled by different classes of agents, connected by a dynamic network, which establishes a link between each patient and his chosen provider. In addition, a static social network between the patients is set up, which enables them to exchange information about their providers.

The health state of a patient as well as his satisfaction with the provider are represented by state charts. Furthermore each patient has a quality of life, which decreases from the moment he falls ill until he recovers and thereby affects his satisfaction. Additionally, the patient's satisfaction level is reduced when the waiting time for an appointment exceeds a

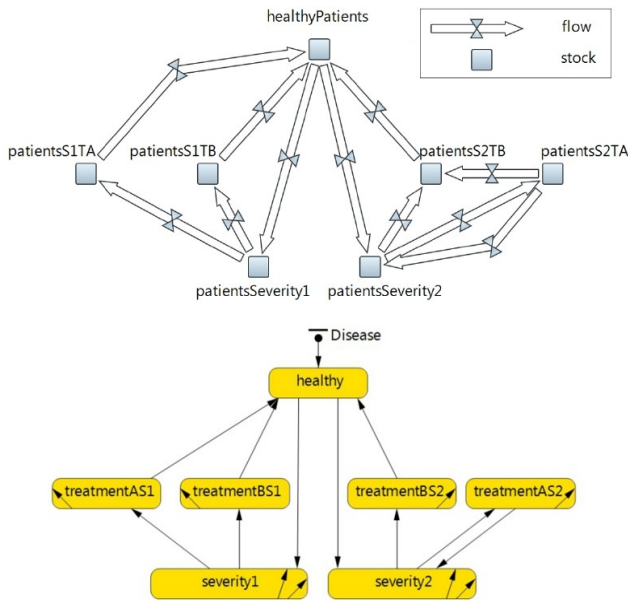


Figure 4: Correspondence Between the Stock and Flow Diagram Used in the SD Model (top) and the State Chart Used in the AB Model (bottom)

certain time span. Thus, this level reflects how a patient perceives the actions of his provider and if, in his point of view, these impede or support his attempts to keep a high quality of life. In case he becomes dissatisfied, he may change provider. To assure that this change improves his current situation, it is based on the recommendations of other agents in his social network. Only if all suggested choices, starting from the first one received, reject the patient due to full capacity utilisation, the new provider is chosen randomly. If the patient is fully satisfied for a longer period, he in turn makes a recommendation.

Each medical provider has the same number of working units per day to see patients, who come either for a regular check-up or because of illness. In the latter case, the provider has to decide which treatment to prescribe. For that purpose, he first diagnoses the patient's condition, i.e. he determines a number between 0 and 1, which is triangularly distributed with the mode depending on the patient's severity level. If the value is greater than 0.5, this indicates a higher probability for the more severe case of the disease, whereas for values less than 0.5, the opposite applies. Comparing this value to an individual threshold, above which the extended treatment is prescribed, then yields the treatment decision. Each medical provider aims at maximising his performance with respect to the patient's benefit on the one hand and his profit on the other hand by adjusting this threshold. The objective function is realised as the weighted sum of normalised income and benefit, times the utilisation ratio of the provider. A similar approach is used by Ellis (1998).

There are twofold interactions between medical providers and patients. First, the treatment chosen by the provider influences the health state of the patient

and consequently his quality of life. Second, the choice of provider by the patient influences the performance of the provider. In the first place, the patient randomly chooses a medical provider. Later on, as soon as information based on past experience becomes available, this choice is influenced by recommendations of the social network, waiting times for appointments in the past and satisfaction with the treatment.

3.2. Implementation as SD Model

In the SD model, the stocks describe the current amount of healthy and ill patients. The flows between the stocks are determined on the one hand by natural disease rates and treatment durations and on the other hand by the available supply of medical services and a factor, which represents the averaged (individual) treatment decisions of all providers, who are themselves modelled homogeneously.

The above described analogy (Borshchev and Filippov 2004) is used to assure the comparability of the two models regarding the way patients become ill as well as their recovery, see Fig. 4. Thus first, the transitions between healthy and sick, used in the AB model, should correspond to the rate-dependent flows between the level of healthy patients and the level of sick patients in the SD model. This correspondence is ensured by modelling the change of state from healthy to sick in the AB model by a transition, which is triggered after an exponentially distributed time span. If the parameter of the exponential distribution is set to the same value as the rate determining the size of the flow in the SD model, their behaviour is equivalent. Second, the transitions from the three treatment states to healthy, should correspond to the flows between the three treatment stocks to the stock "healthyPatients". It is assumed that the time needed for recovery is the same for every patient but dependent on the prescribed treatment. This can be realised by a so called "pipeline material delay" in SD and a time-out triggered transition in the AB model.

In contrast to the AB model where each agent tries to attain his goal individually, the goals of patients and providers are first averaged and then combined in one objective function, which is maximized during simulation. The providers' objective was again implemented as the weighted sum of normalised profit and patient's benefit. The goal of the patients is represented by the minimal level of quality of life, which is on average reached during the course of the disease.

The optimisation is realised by a discrete event, which might be an unusual approach because of the continuous nature of SD. But as mentioned before, this is due to the intended analogy between the models.

4. SIMULATION

The default parametrisation for the simulation uses 5000 patients and 150 medical providers. The disease rates are set to 0.03 and 0.01 for the less and for the more severe form of the disease respectively. These

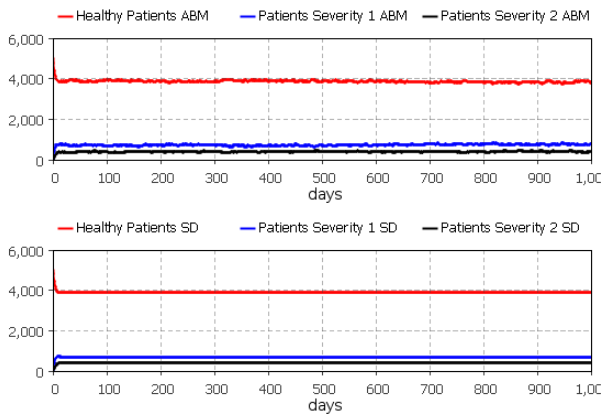


Figure 5: Number of Healthy and Sick Patients in the AB Model (top) and the SD Model (bottom) over a Simulation Period of 1000 Days.

values assume a very high density of medical providers, as they are all assumed to have the same specialisation. Moreover a person becomes ill on average about once a month. This setting may not be entirely realistic. However, the above values were chosen for practical reasons. The simulation of an AB model is computationally intensive and therefore too big samples are not convenient. Then again, if the number of agents is too small, the results show too much noise due to the stochastic elements included in the behavioural rules of the agents. Therefore, less patients, but more providers as well as a higher disease rate were taken as compromise. If one patient is assumed to represent 80 patients with the same behaviour patterns, the provider density is realistic (Habl and Bachner 2010).

The reimbursement of treatment B is supposed to yield 50% more profit than that of treatment A, and the durations of the treatments are set to 3 and 7 days for A and B respectively. How often a patient consults a medical provider is specified as, on average, once every two months for check-ups and two days after the first symptoms have occurred.

Simulation runs with the default parametrisation yield very similar results for both models. The derived average behaviour of the providers in the AB model and the assumed one in the SD model seem to correspond.

Figure 5 shows the number of healthy and sick patients (divided into two groups according to the level of severity) over time for both the AB and the SD model with the default parametrisation. The quantitative values look alike for both methods with the difference that the results obtained from the agent-based approach show fluctuations caused by the individual, stochastic behaviour of the agents. In contrast, with the SD model a steady state is reached after the optimisation process is finished.

The behaviour of the providers is represented by their treatment decision. Thus, it can be studied by looking at the number of prescribed treatments of the two different types. Figure 6 shows the total amount of health care services used per month over a simulation period of 20 years. During the first half of this period, the parameters are kept at their default values. After an

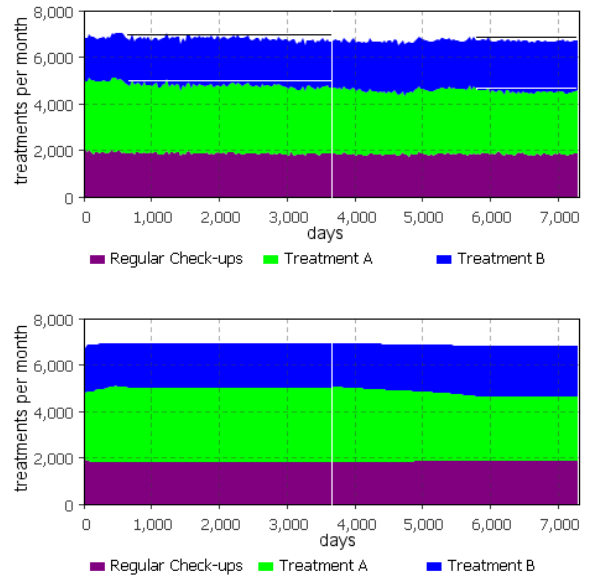


Figure 6: Total Number of Health Care Services Used per Month over a Simulation Period of 20 Years With a Raise of the Reimbursement for Treatment B after 10 Years, Marked by the White Vertical Line (AB Model Top, SD Model Bottom). The Horizontal Lines in the Top Panel Indicate the Corresponding Values of the SD Simulation, when Steady State is Reached.

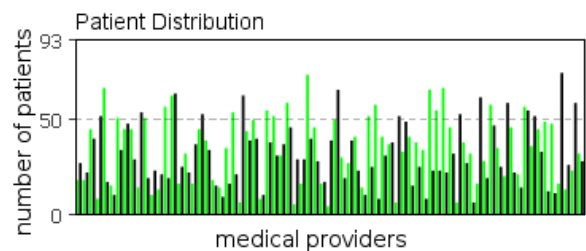


Figure 7: Competition of the Medical Providers in the AB Model

initial adaptation time, the optimal behaviour is found and hence the system stabilises. Like before, a steady state is reached in the SD simulation, whereas the values computed with the AB model show an oscillating behaviour, yet within a certain range.

After ten years, the reimbursement for treatment B is risen such that it brings 4 times the profit of treatment A. The reaction of the system looks alike. In both cases, the increased financial incentive for prescribing treatment B influences the treatment decision. However profit is not the only concern of the medical provider and also quality of life of the patients has an impact on their behaviour. In the SD model the influence of the patients is realised directly, as the objectives of providers and patients are maximised simultaneously. In the AB model, the quality of life of a patients affects his choice of a provider and thus induces a competition between the providers (see Fig. 7), which then again influences their behaviour. Because of these additional

influences on the treatment decision, treatment B is indeed prescribed more often, but again the amount stabilises at a certain level, which represents the optimum for both patients and medical providers. Furthermore, as the increase is moderate, the plus in prescriptions of treatment B concerns only cases where the diagnosis is uncertain, i.e. the number obtained by the provider is close to 0.5, which might also happen in reality.

Regarding the quantitative outcome, the obtained results for regular check-ups almost coincide, whereas those for treatment A as well as for the extended treatment show deviations. This could be explained by the differences in the way the patient-provider contact is modelled. In the SD model, the current supply of total health care services is distributed among the different stocks of patients according to their demand. This process is more complex in the AB model, as the time until a patient has an appointment with his provider depends not only on the total available supply but also on the chosen provider. Hence, the average time until a patient sees a provider and consequently the total illness duration is higher. Therefore, even though in both models the same amount of patients is sick, less treatments are prescribed in the AB model.

5. CONCLUSION

SD and ABM both have their benefits and drawbacks for modelling health care utilisation. But if these are compared, it is apparent that the strengths of the methods concern the description of totally different aspects of the system. With the agent-based approach, individual characteristics and behaviour patterns can be considered, which are impossible to include in an SD model. However, if the focus lies on the the global relations in the system, SD certainly is the better choice.

The demonstration model was implemented in a rather intuitive way and the used variables and relations are all interpretable in both methods. The simulation results show that the models describe the same qualitative behaviour. However, deviations in the quantitative outcome exist. Parts of them are caused by the stochastic elements used in the AB model and are thus inherent in the approach. The rest is due to effects emerging from the heterogeneous distribution of patients among the providers, i.e. network effects, in the AB model, which are ignored in the system dynamical implementation. Thus, even though the specified system is kept as simple as possible, the differences between the methods already carry weight.

REFERENCES

Andersen, R.M., 1995. Revisiting the Behavioral Model and Access to Medical Care: Does It Matter? *Journal of Health and Social Behavior* 36(1):1-10.

Bonabeau, E., 2002. Agent-based modeling: methods and techniques for simulation human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 7290-7287.

Borshchev, A., Filippov, A., 2004. From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. *The 22nd International Conference of the System Dynamics Society*, July 25-29, Oxford, England.

Ellis, R.P., 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics* 17:537-555.

Forrester, J.W., 1961. *Industrial Dynamics*. Cambridge, Massachusetts : M.I.T. Press.

Habl, C., Bachner, F., 2010. *Das österreichische Gesundheitswesen im internationalen Vergleich*. Bundesministerium für Gesundheit. Available from: http://public-health.meduni-graz.at/archiv/artikel/Artikel%202010/2010_Oe%20im%20Internationalen%20Vergleich.pdf [Accessed 12 July 2012].

Kern, A.O., 2002. Arztinduzierte Nachfrage in der ambulanten Versorgung-Bedeutung für eine Privatisierung von Leistungen der Gesetzlichen Krankenversicherung. *Discussion Paper Series* 225

Ma, C.A., McGuire, T.G., 1997. Optimal Health Insurance and Provider Payment. *The American Economic Review* 4:685-704.

Macal, C.M., North, M.J., 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation* 4:151-162.

Radzicki, M.J., Taylor, R.A., 1997. *Introduction to System Dynamics, A Systems Approach to Understanding Complex Policy Issues*. U.S. Department of Energy. Available from: <http://www.systemdynamics.org/DL-IntroSysDyn/start.htm>. [Accessed 13 May 2012]

Romstorfer, G., Parragh, S., Schneckeneither, G., Landsiedl, M., Einzinger, P., Scheuringer, M., 2011. Integration of GIS Data in Health Care Utilization. *Tagungsband und Fullpapers, ASIM 2011*, September, Winterthur, Schweiz.

Thode, N., Bergmann, E., Kamtsiuris, P., Kurth, B.M., 2004. *Einflussfaktoren auf die Inanspruchnahme des deutschen Gesundheitswesens und mögliche Steuerungsmechanismen*. Schlussbericht. Robert Koch Institut, Berlin.

A COMPARISON OF SYSTEM DYNAMICS AND MARKOV MODELS FOR COST-EFFECTIVENESS ANALYSIS OF CHRONIC DISEASES

P. Einzinger^(a), R. Leskovar^(b), C. Wyrzens^(c)

^(a)dwh Simulation Services, Austria

^{(b), (c)}Vienna University of Technology, Institute for Analysis and Scientific Computing, Austria

^(a)patrick.einzinger@drahtwarenhandlung.at, ^(b)e0726512@student.tuwien.ac.at, ^(c)e0825785@student.tuwien.ac.at

ABSTRACT

To simulate chronic diseases Markov modeling is often used, but however the System Dynamics (SD) methodology is also applicable to this task. Both kinds of simulations can be used to analyse costs and effects of medical technologies for diseases. In this paper we perform an exemplary cost-effectiveness analysis for a smoking cessation programme for chronic obstructive pulmonary disease (COPD) patients with a simplified Markov model based on Menn (2009) as well as an analogue SD model. Such transformations from Markov models to SD models are always possible and lead to similar results. However, only the latter is of incorporating interactions between different patient groups.

Keywords: markov model, system dynamics model, cost-effectiveness analysis, ICER

1. INTRODUCTION

Cost-effectiveness analyses are often used as basis of decision making of different kinds of medical technologies (e.g. drugs, vaccination programmes, and other treatments). Such a cost-effectiveness analysis compares the effects and the costs of different treatments, for example by calculating the ICER (incremental cost-effectiveness ratio). Various types of models allow a cost-effectiveness analysis as long as they generate the cumulative costs and effects of an intervention as output.

The ICER is defined as:

$$ICER = \frac{C_1 - C_2}{E_1 - E_2} \quad (1)$$

C_1 describes the costs and E_1 the effects of treatment 1, e.g. the actual treatment. Accordingly C_2 and E_2 are the costs and effects for treatment 2, e.g. an alternative intervention (Briggs and Sculpher 1998).

One possible modelling methodology for cost-effectiveness analyses are the so called Markov models. They are common for modelling the progression of chronic diseases through several disease stages. But this type of model is not the only possibility to simulate

such a chronic disease. The SD methodology, for example, is also applicable to this task. To show the similarities and differences of these two types of modelling as well as the generic transformation process from a Markov to a SD model we simulated the progress of the chronic disease COPD for a cohort of patients and calculated a cost-effectiveness analysis.

The basic model is a simplified version of the Markov model for COPD and two different treatments, the routine treatment on the one hand and a smoking cessation programme on the other hand, by Menn (2009). We transformed this model into a SD model such that advantages, disadvantages and various possibilities of expansions can be shown.

On PubMed we searched for other models which concern COPD in every possible way. There exist a lot of studies about COPD and a little bit less models. However, the cost-effectiveness analysis, which is realised in this paper, is very popular in the research of chronic diseases.

One model was very similar to the available model. The corresponding study performed a cost-effectiveness analysis with the help of a Markov model and a Monte Carlo simulation with the two cohorts "Smokers" and "Ex-Smokers". There are some expansions like the possibility to change the state of smoking, another discounting rate and one additional state (Atsou et al. 2011).

One cost-effectiveness analysis was performed to see the differences between the two chronic diseases COPD and Asthma in the context of countries with low and middle income (Stanciole et al. 2011).

One cost-utility analysis has been realised to research a new method to test the arterial puncture of COPD patients again with the help of a Markov model (Oddershede et al. 2011).

One model which is really different from the others was a decision tree to analyse the advanced directives of COPD patients (Hajizadeh et al. 2010).

2. COPD

COPD is a common chronic disease of the lung. The different stages of the disease and therefore COPD itself are irreversible (GOLD 2010).

COPD can be divided into four different kinds of stages. Patients are classified after their forced expiratory volume in 1 second (FEV₁) that can be measured by spirometry (GOLD 2010).

Table 1: Classification of COPD Stages by GOLD

Severity of COPD	FEV ₁
mild COPD	≥ 80%
moderate COPD	50% – 79.99%
severe COPD	30% – 49.99%
very sever COPD	< 30%

In our model mild COPD is “stage 1”, moderate COPD “stage 2”, severe COPD “stage 3” and very severe COPD is “stage 4”. A fifth possible state is “death”, which obviously is an absorbing state. The stages are distinguished by their expected life quality – these effects are quantified by QALYs (quality-adjusted life years) – and the costs of the patient’s treatment. Another feature of the stages is that it is only possibly to leave a stage by progress into the next higher stage or by death. Patients cannot skip a stage or get better, because the decline of lung function is irreversible.

The reachability of the stages can be represented by a reachability graph, shown in figure 1.

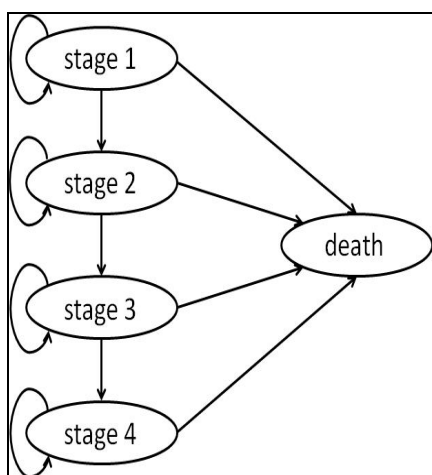


Figure 1: Reachability Graph of COPD

3. MARKOV MODEL

Markov models are based on the simulation of cohorts whose members transit through the states of the model. In this paper the simulated cohort, which contains only patients who suffer from COPD in stage 1 at the beginning, is subdivided into two different kinds of treatments: the routine treatment on the one hand and the intervention on the other hand. These cohorts are again divided into cohorts of smokers and persons who do not smoke anymore. For each of these cohorts a Markov model was calculated. The structure of the whole model is shown in figure 2, where COPD stands for the reachability graph of the different stages shown in figure 1.

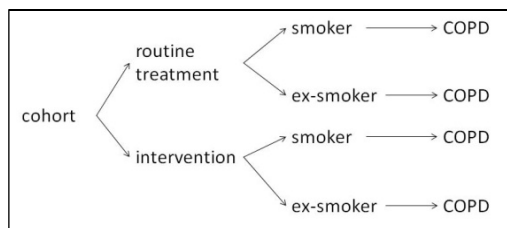


Figure 2: Structure of the Whole Model of COPD

In general, Markov models have different Markov states. These states are affiliated with each other by transition probabilities. Therefore Markov models are stochastic models (Briggs and Sculpher 1998). In our case these states correspond with the stages of COPD or the absorbing stage “death”. “Stages 1-3” are all provided with three kinds of probabilities. The first one is the probability to stay at the same state as in the cycle before, the second one is the transition probability to move over to the next state, which means to get worse, and the third one describes the probability of dying. “Stage 4” has got only two probabilities, because the patient only can stay in the state or die, as no further progress is possible. “Death” does not have any transition probability which leads away from this state, because it is an absorbing state. Therefore people who die before the end of the simulation are not considered for the costs and effects anymore and furthermore do not influence the simulation any longer (Menn 2009). However, the feature that a person can stay in exactly one stage at any time of the simulation is characteristic of Markov models (Briggs and Sculpher 1998, Sonnenberg and Beck 1993).

Furthermore, an attribute of Markov models is that the operation time is not continuous, but discrete. This means that time is partitioned into time steps, the so called Markov cycles (Briggs and Sculpher 1998, Sonnenberg and Beck 1993). The cycle length of the model should be short enough so that multiple changes in pathology, symptoms, treatment decisions, or costs within a single cycle are unlikely (Weinstein et al. 2003). A patient can only change his/her stage between two cycles and not during a cycle. Therefore the model will be assumed as constant during the single Markov cycles (Briggs and Sculpher 1998).

Another defining property of Markov Models is that each Markov model fulfils the Markov property (Sonnenberg and Beck, 1993).

$$P\{N(t) = j \mid N(s) = i, N(s_n) = i_n, \dots, N(s_0) = i_0\} = P\{N(t) = j \mid N(s) = i\} = P\{N(t) - N(s) = j - i\} \quad (2)$$

This formula means that the future state j at time t only depends on the actual state at time s . Therefore the probability of the future state does not depend on past states. Particularly, the probability of the progress of a person’s disease is independent of the states in which the person has been in the past and of the states in which the person will be in the future (Fahrmeir et al. 2012).

After simulation the ICER for the cost-effectiveness analysis can be calculated. The ICER is defined in formula (1). C_1 and E_1 represent the costs and effects of the routine treatment (smokers and “ex-smokers” combined) of COPD and C_2 and E_2 the costs and effects for the intervention to stop smoking (smokers and “ex-smokers” combined too). The costs of the intervention are equal to the costs of the routine therapy. However, every person who receives the alternative treatment adds basic costs of 596 € at the beginning of the simulation (Menn 2009).

3.1. Simulation

All parameters values for the simulation were adopted from Menn (2009).

The running time of the simulation of the Markov model was 60 years. For the simulation the assumption that all patients start at the age of 45 in “stage 1” was made. A transition during simulation from smoker to ex-smoker was not possible. The cycle length was 3 months.

Because of the cost-effectiveness analysis the number of the persons, who receive the routine treatment, is the same as the number of the persons, who receive the intervention. If joining the routine treatment, the probability of being an ex-smoker will be 0.06 and on the other hand 0.22, if joining the intervention.

Table 2 shows the transition probabilities of reaching the next higher stage.

Table 2: Transition Probability from Each Stage to the Next Higher Stage (Menn 2009)

Stage	Smoker	Ex-smoker
1	0.014	0.0025
2	0.0103	0.003
3	0.023	0.0069

The mortality for smokers and ex-smokers is dependent on time. Therefore, there are different probabilities of dying for the different ages of the persons. The mortalities of smokers are shown in table 3, for ex-smokers in table 4.

Table 3: Probability of Dying for Smokers and the “Stages 1 – 4”, Depending on Time (Menn 2009)

Age	Stage 1	Stage 2	Stage 3,4
45-49	0.0012	0.0012	0.0050
50-54	0.0021	0.0036	0.0085
55-59	0.0030	0.0052	0.0123
60-64	0.0047	0.0080	0.0190
65-69	0.0081	0.0138	0.0324
70-74	0.0136	0.0232	0.0542
75-79	0.0198	0.0335	0.078
80-84	0.0340	0.0576	0.1314
85-89	0.0473	0.0797	0.1790
≥90	0.0884	0.1467	0.3140

Table 4: Probability of Dying for Ex-smokers and the “Stages 1 – 4”, Depending on Time (Menn 2009)

Age	Stage 1	Stage 2	Stage 3,4
45-49	0.0008	0.0014	0.0032
50-54	0.0014	0.0023	0.0056
55-59	0.0025	0.0043	0.0102
60-64	0.0039	0.0067	0.0158
65-69	0.0050	0.0086	0.0204
70-74	0.0085	0.0146	0.0342
75-79	0.0132	0.0225	0.0526
80-84	0.0228	0.0388	0.0896
85-89	0.0396	0.0669	0.1516
≥90	0.0742	0.1239	0.2695

The transition of persons from one stage to another stage can now be calculated with the help of formula (3).

$$A_{i,j} = A_{i-1,j} \cdot (1 - p_{j \rightarrow j+1} - d_{i,j}) + A_{i-1,j-1} \cdot p_{j-1 \rightarrow j} \quad (3)$$

i stands for the cycle number and therefore for the time and j for the stage. $p_{j \rightarrow j+1}$ describes the probability of the possible transition from stage j to stage $j+1$ and $d_{i,j}$ stands for the mortality of stage j at time step i , $A_{i,j}$ contains the number of people in stage j at cycle i . This formula can be used to calculate all four stages for the simulation. But for “stage 1” the additional term $A_{i-1,j-1} \cdot p_{j-1 \rightarrow j}$ is 0 because there is no “stage -1”. For “stage 4” $p_{j \rightarrow j+1}$ is 0 because the persons who are in this stage can only stay in this stage or die, therefore there is no transition probability to a higher stage. The number of persons of the stage “death” in cycle i now can be calculated by summing up all the people of the other stages at this cycle and then subtract this sum from the whole number of persons in the cohort.

The costs in Euros and effects in QALYs for each stage are listed in table 5 and formula (4) shows the calculation of the costs.

$$K_i = A_{i,1} \cdot C_1 + A_{i,2} \cdot C_2 + A_{i,3} \cdot C_3 + A_{i,4} \cdot C_4 \quad (4)$$

K_i stands for the undiscounted costs of all persons of the cohort at cycle i . $A_{i,j}$ $j=1,..,4$ are the number of people in stage j at cycle i . C_j $j=1,..,4$ are the costs for each stage j , which are listed in table 5. In analogy formula (4) can be used to calculate the effects for each time step.

Table 5: Cost and Effects of “Stage 1 – 4” (Menn 2009)

Stage	Costs	Effects
1	103	0.2100
2	185	0.1975
3	367	0.1875
4	431	0.1625

Discounting of costs and effects is needed because the running time of the model is 60 years. The discounting rate is 3%. The discounted costs and effects are calculated with the help of formula (5).

$$C_d = \frac{C}{(1+d)^{(t/12)(i-1)}} \quad (5)$$

C stands for the undiscounted costs/effects, C_d for the discounted costs/effects, d for the discounting rate, t for the length of the Markov Cycle in months and i for the actual cycle of the process (Menn 2009).

4. SYSTEM DYNAMICS MODEL

As already mentioned in the introduction the SD model for COPD has been constructed on the base of the Markov model.

The general structure of a SD model consists of stocks and flows (Bossel 2004). For modelling COPD the four different stages of illness and the stage “death” were represented by stocks, which are the equivalent of the states in a Markov model.

The stage “death” could also be constructed as a “sink”, because it is an absorbing stage (Brailsford 2008).

Accordingly, flows are the equivalent of the transitions of patients between the stages in the Markov model.

In our case, each stage has one flow to the next stage and one flow to the stage “death” which resembles the possible transitions of the Markov model. The transitions are influenced on the one hand by the number of people in a stage from which the transition is going away and on the other hand by the transition rates. The transition rates can be understood in analogy to the transition probabilities of the Markov model.

The System Dynamics model is not a stochastic, in contrast to the Markov model, but a deterministic model, so not probabilities but (fractional) rates are needed. The rates can be calculated from the probabilities of the Markov model with the following formula:

$$r = -\ln(1 - p) \quad (6)$$

Here r defines the rate and p the probability (Menn 2009).

But if there is more than one possibility to change the state, as it is the case in our model (patients can die or transit to the next higher stage), it can be necessary to correct the probabilities before transforming them into rates. This is the case if the probabilities of one state represent mutually exclusive events. In our model, as people who die during one time unit reduce the number of people who can possibly transit to the next higher stage. Only the proportion $p_d + p_t(1-p_d)$ of patients would change the state in the SD model. However, formula (3) shows that in the Markov model the corresponding proportion simply equals $p_d + p_t$. To correct this, the transition probabilities have to be calculated in the following way:

$$p_t = \frac{p_t}{1 - p_d} \quad (7)$$

p_t describes the transition probability and p_d the probability of dying. After that, the probabilities can be transformed with formula (6). This calculation leads to the following transition rates for the SD model:

Table 6: Transition rates for the ex-smokers of the SD model

Cycle	Transition 1	Transition 2	Transition 3
0	0.00250514	0.00300873	0.00694622
20	0.00250664	0.00301145	0.00696304
40	0.00250941	0.0030175	0.00699552
60	0.00251294	0.0030248	0.00703546
80	0.00251294	0.00303061	0.00706862
100	0.00252462	0.00304909	0.00716998
120	0.00253666	0.00307377	0.00730974
140	0.0025616	0.00312598	0.00760795
160	0.00260648	0.00322027	0.0081662
180	0.00270402	0.00343014	0.00949048

Table 7: Transition rates for the smokers of the SD model

Cycle	Transition 1	Transition 2	Transition 3
0	0.014116	0.0103753	0.0233869
20	0.0141288	0.010391	0.0234705
40	0.0141416	0.0104078	0.0235618
60	0.014166	0.0104373	0.0237247
80	0.0142149	0.010499	0.0240572
100	0.0142947	0.0106006	0.0246186
120	0.0143843	0.0107142	0.0252566
140	0.0145979	0.0109897	0.0268363
160	0.0148041	0.0112551	0.0284145
180	0.0154768	0.0121442	0.0341026

The transition rates and the death rates are inserted in the SD model as a lookup function.

In SD models the transitions are given by ordinary differential equations. Therefore, they are continuous time models. To calculate how many people are in which stage at a certain time step one has to integrate over time (Bossel 2004). Usually this is performed with numerical solving algorithms for differential equation systems and discrete time steps. However, theoretically these time steps get arbitrarily small.

The formula how to calculate the people in one stage is exemplary given in formula (8) for stage 2.

$$stage2(T) = \int_0^T (transition1(t) - die2(t) - transition2(t)) dt \quad (8)$$

Where “transition 1” describes the flow of the people from “stage 1” to “stage 2”, “die 2” stands for the flow from “stage 2” to the stock “death” and “transition 2” represents the flow from “stage 2” to “stage 3”.

The transitions, which are the flows in the model, are calculated in this way:

$$transition2 = stage2 \cdot transitionrate2 \quad (9)$$

“Stage 2” is the number of people in stage 2 and the “transition rate 2” is the rate for the flow of people from stage 2 to stage 3. Both depend on time.

The effects and the costs are two supplementary levels. They are calculated with the same formula (4) as in the Markov model. Only the discounting of costs and effects is different because the SD model is a continuous model. So it is not possible to discount in the same way as in the Markov model. In the SD model it is necessary to discount continuously. So, the costs and the effects are discounted with the help of the exponential function:

$$K(T) = \int_0^T k(t) \cdot e^{-rt} dt \quad (10)$$

K(T) describes the discounted and k(t) the undiscounted cost or effects and r the discounting rate and T the time.

The discounting rate for the discrete case corresponds to the discounting for one year. The rate of the Markov Model stands for a year. The formula to calculate the rate for one cycle is:

$$d_c = (1 + d_y)^{m/12} - 1 \quad (11)$$

d_c stands for the rate for one cycle and d_y for the rate for one year and m stands for the length of one cycle in months. One cycle in the SD model is like 3 months in the Markov model.

After that, it is necessary to calculate the rate for a continuous model with the following formula:

$$d = \ln(1 + d_c) \quad (12)$$

d stands for the discounting rate for the SD model.

The model has been constructed two times, once for the cohort “Smokers” and another time for the cohort “Ex-Smokers”. The models are the same, except for the transition and mortality rates and the number of people.

It is also possible to simulate the model with each of the two possible treatments. Therefore an auxiliary variable exists for changing the number of people that are smokers or ex-smokers, because this is the only difference between the interventions.

The ICER is calculated the same way as in the Markov model. But the ICER is not calculated in the model itself.

In figure 3 a simplified stock and flow diagram of the stages of the SD model is shown as example. Figure 4 shows the stock and flow diagram for calculation of the costs and effects of the SD model.

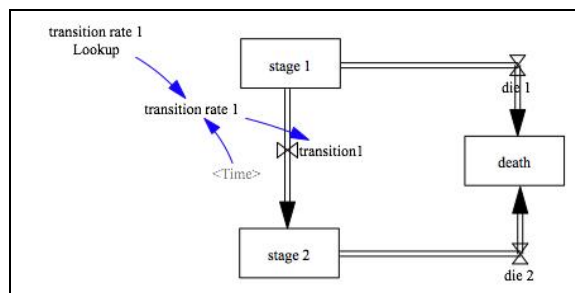


Figure 3: Stock and Flow – diagram of the stages

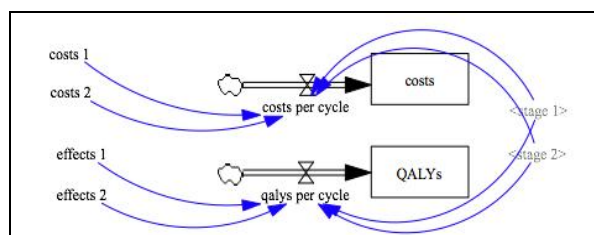


Figure 4: Stock and Flow – diagram of the costs and QALYs

5. THE DIFFERENCES

The Markov model needs probabilities to calculate the number of people in each stage (stochastic model) in contrast to the SD model (deterministic model), which needs rates.

The Markov model is a discrete time model because the transitions only can take place each time step. The SD model is a continuous-time model.

The SD model is constructed over stocks in opposition to the Markov model where the cohort is divided into states. But in spite of this, the stocks are equivalent to the states, which means that for each state there has to be a corresponding stock in the transformation of the model.

The discounting of the costs and effects in both models is different because we have on the one hand a discrete model and on the other hand a continuous model. However, the discounting factors are equal at each time step.

Table 1 shows the differences between the two models. These differences cause the different possibilities to expand the models.

Table 8: Differences/Equivalences between the Models

MARKOV	SYSTEM DYNAMICS
stochastic	deterministic
discrete time	continuous time
state	stock
transition probabilities	transition rates
discrete discounting	continuous discounting

5.1. Expansions

In both models, expansions are possible. In the SD model, one possible way is to include feedback and delays. In this research one possible expansion has been realised. People who are very ill, as in “stage 3” or “stage 4” of COPD, could influence people in “stage 1” or “stage 2”. If many people are very ill, more people

with mild cases will stop smoking than before (deterrence). This scenario can be carried out with the help of the Smooth-function, which averages information (e.g. the value of a level) over time. Naturally, the influence on people does not take place immediately, but this process needs time, for example due to the politics and news. The Smooth function implements an Information Delay, in which the delay time is provided (Forrester 1969).

In this model the delay time was 12 month. A flow from the “stage 1 smokers” to the “stage 1 ex-smokers” was added, which depended from the smoothed “stage 4 smokers”.

In the Markov model, this expansion is not possible because in a Markov model all patients are assumed to undergo an independent stochastic process. Therefore the amount of patients in one state cannot influence the stochastic process of another patient. However, it is possible to simulate a Markov model not as a cohort, but as a Monte Carlo simulation, because it is a stochastic model. In this case the patients run through the model individually and the transition probabilities are weighted in randomised probabilities. The Monte Carlo simulation also could be used as a sensitivity analysis (Briggs and Sculpher 1998).

6. RESULTS

The figures 5 and 6 show the division of the smokers and of the ex-smokers into the five stages in the SD model.

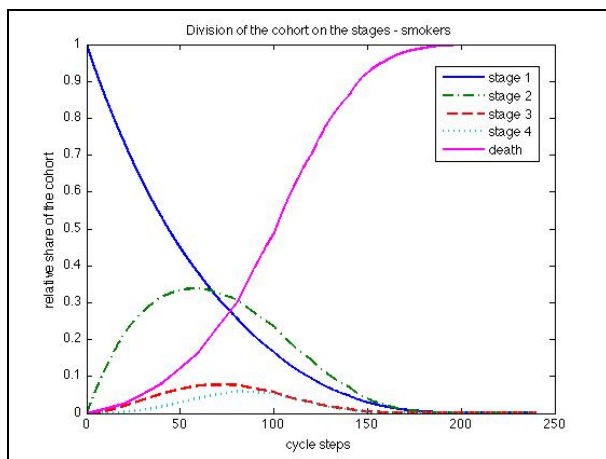


Figure 5: Division of the cohort into stages – smokers – System Dynamics model

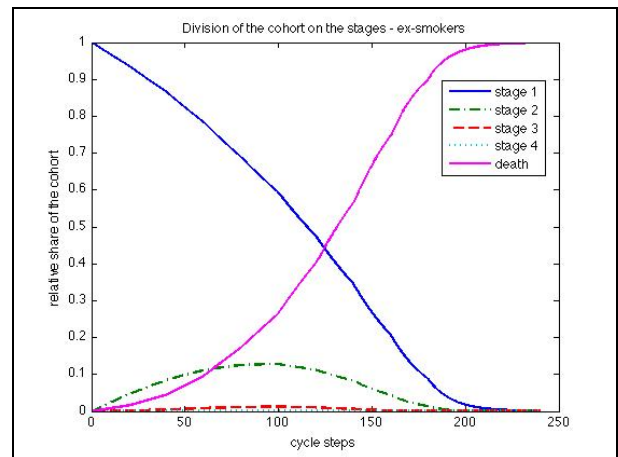


Figure 6: Division of the cohort into stages – ex-smokers – System Dynamics model

In the graphics of the SD model in opposite to the graphics of the Markov model no differences can be seen. But, in the analysis of the data are differences. The figures 7 and 8 show these differences of the division into the stages between the two models, because the absolute error of the two simulations is described. In both plots can be seen, that the biggest difference is in stage 2. In the cohort of the smokers, this difference is about 0.002 and in the cohort of the ex-smokers it is about 0.0003. Very noticeable is the oscillating behaviour of these differences, like in stage 1 in the beginning or the stage death in the end in both cohorts. Especially interesting as well is the drop of the stage 3 in both graphics and the fast increase of the stage 4 at the end in the cohort of the smokers

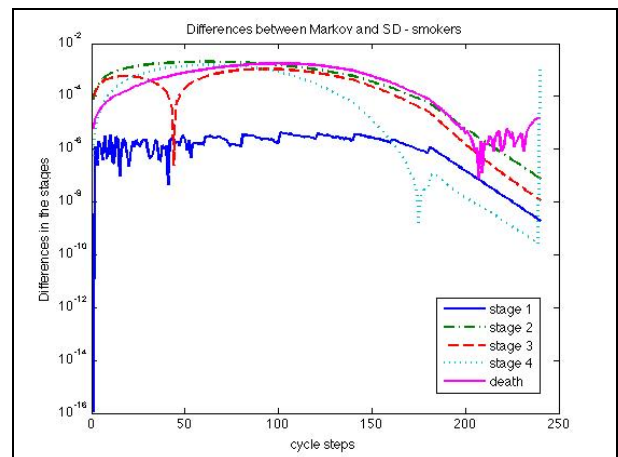


Figure 7: Differences between the Markov and SD model in the division into the stages – smokers

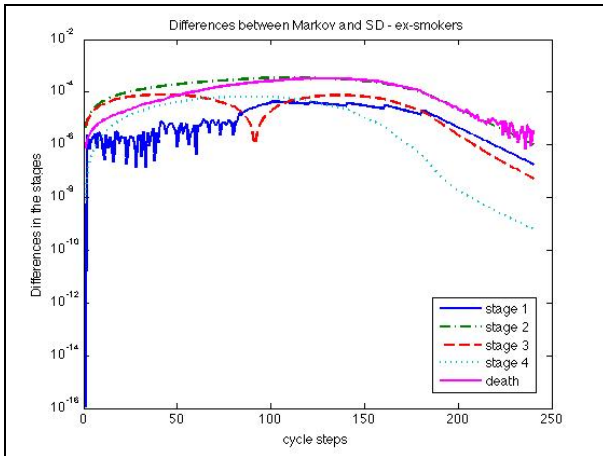


Figure 8: Differences between the Markov and SD model in the division into the stages – ex-smokers

The course of the costs and effects in the two models are very similar and no differences can be shown. But in the data are some small differences too, as for the division of the stages. The discounted values are more exactly than the undiscounted.

The ICER of the Markov model averages 838.4912 €/QUALY and the ICER of the SD model is 773.1253 €/QUALY. Why there are again some differences between the models is not yet analysed, as well not if only the differences of the stages cause the differences in the costs and effects and the ICER. Several factors could cause these deviations. Among them are the fact that although the two models should lead to equal values of states and levels at the discrete time steps, the continuous model calculates costs and effects also between these time points (and here also the discounting factor can differ), rounding errors and procedural errors of the numerical solver in the continuous case.

The following two graphics 9 and 10 show the division of the cohorts of smokers and ex-smokers in the expanded version of the SD model. It is easy to see that in the cohort of the ex-smokers are much more people than in the normal version and that they are coming from the cohort of the smokers. The people in “stage 1” of the ex-smokers increase very fast. After that, the course of the people through the stages is the same as in the normal version. In the cohort of the smokers “stage 1” shrinks in opposite to the ex-smokers and after the course is similar to the normal version.

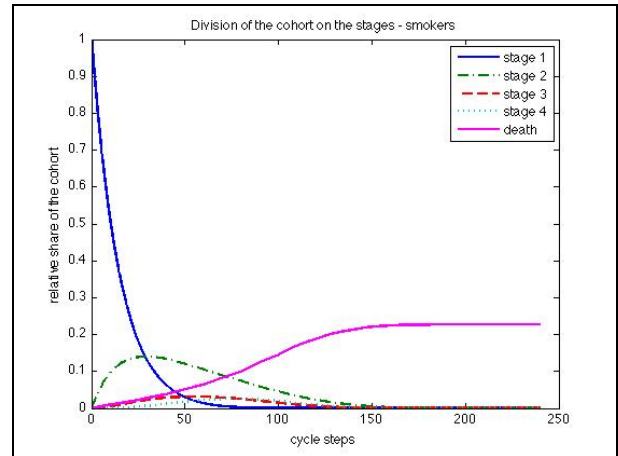


Figure 9: Division of the cohort into the stages – smokers – expanded System Dynamics model

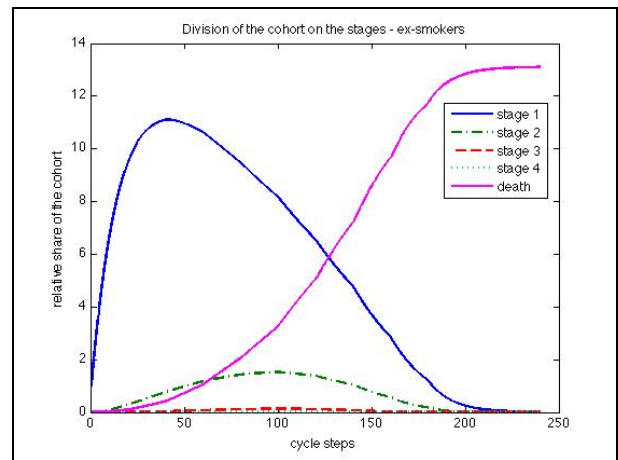


Figure 10: Division of the cohort into the stages – ex-smokers – expanded System Dynamics model

7. CONCLUSION

The approach shows that each Markov model can theoretically be transformed into an analogue SD model. In our example we got results that were slightly different, especially for the calculated ICER. There are a few possible explanations for these deviations, but this issue needs further and more detailed analyses. Naturally, there exist differences between the two methods, especially when it comes to possible expansions. For some research questions in the area of health care the limitations of Markov models are not a problem. However, a transformation of an existing Markov model into an SD formulation could allow for the incorporation of further aspects and influences, as was demonstrated in the case of deterrence for smokers in mild COPD stages.

REFERENCES

Atsou K., Chouaid C., Hejblum G., 2011. *Simulation-Based Estimates of Effectiveness and Cost-Effectiveness of Smoking Cessation in Patients with Chronic Obstructive Pulmonary Disease*. Available from:

- <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3173494/?tool=pubmed> [accessed 12 July 2012].
- Bossel, H., 2004. *Systeme, Dynamik, Simulation*. Norderstedt: Books on Demand.
- Brailsford, S.C., 2008. System Dynamics: What's in it for healthcare simulation modelers. *Proceedings of the 2008 Winter Simulation Conference*, 1478-1483. December, 7-10th 2008, Miami (Florida, USA).
- Briggs, A. and Sculpher M., 1998. An Introduction to Markov Modelling for Economic Evaluation. *Pharmacoeconomics* 13 (4): 397-409
- Fahrmeier, L., Raßer, G., Kneib, T., 2012. *Skript zur Vorlesung Stochastische Prozesse*. Ludwig-Maximilians-Universität zu München. Available from: <http://www.statistik.lmu.de/~gertheiss/Lehre/StoPr02012/skript.pdf> [accessed 3 July 2012].
- Forrester, J.W., 1969. *Industrial Dynamics*. 6th ed. Cambridge, Massachusetts.: MIT Press.
- GOLD (Global Initiative for Chronic Obstructive Lung Disease), 2010. *Spirometry for Health Care Providers*. Available from: http://www.goldcopd.org/uploads/users/files/GOLD_Spirometry_2010.pdf [accessed 6 July 2011].
- Hajizadeh N., Crothers K., Braithwaite R. S., 2010. *A theoretical decision model to help inform advance directive discussions for patients with COPD*. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3020153/?tool=pubmed> [accessed 12 July 2012].
- Menn, P., 2009. *Einsatz entscheidungsanalytischer Modelle für die ökonomische Evaluation medizinischer Verfahren am Beispiel chronisch obstruktiver Lungenerkrankungen*. Thesis (PhD). Ludwig-Maximilians-Universität zu München.
- Oddershede L., Petersen S.S., Kristensen A.K., Pedersen J. F., Rees S.E., Ehlers L., 2011. *The cost-effectiveness of venous-converted acid-base and blood gas status in pulmonary medical departments*. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169975/?tool=pubmed> [accessed 12 July 2012].
- Stanciole A. E., Ortegón M., Chisholm D., Lauer J. A., 2011. *Cost effectiveness of strategies to combat chronic obstructive pulmonary disease and asthma in sub-Saharan Africa and South East Asia: mathematical modelling study*. Available from: <http://www.bmj.com/content/344/bmj.e608?view=long&pmid=22389338> [accessed 12 July 2012].
- Sonnenberg F. A. and Beck J. R., 1993. Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making* 13 (4): 322-339
- Weinstein M.C., O'Brian B., Hornberger J., Jackson J., Johannesson M., McCabe C. and Luce B.R., 2003. Principles of Good Practice for Decision Analytic Modelling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices-Modeling Studies. *Value Health* 6 (1): 9-17

ANALYSIS AND COMPARISON OF DIFFERENT MODELLING APPROACHES BASED ON AN SIS EPIDEMIC

Andreas Bauer^(a), Carina Pöll^(b), Nina Winterer^(c), Florian Miksch^(d), Felix Breiteneker^(e)

^(a)^(b)^(c)^(e)Vienna University of Technology, Institute for Analysis and Scientific Computing, Austria
^(d)dwh Simulation Services

^(a)andreas.e101.bauer@tuwien.ac.at, ^(b)carina.poell@tuwien.ac.at, ^(c)e0726885@student.tuwien.ac.at,
^(d)florian.miksch@drahtwarenhandlung.at, ^(e)felix.breiteneker@tuwien.ac.at

ABSTRACT

An SIS epidemic is a common approach to model a contagious disease. Three different modelling approaches are used to simulate an SIS epidemic. These methods are an ordinary differential equation model, an agent-based model and a stochastic model. The aim is to compare them in equivalent settings and analyse the qualitative and quantitative similarities and differences. Furthermore, they are extended to simulate diseases based on two serotypes. For the models with one serotype the results for all three approaches are very similar. For two serotypes there are some situations in which the three models agree and some in which they do not agree. Our conclusion is that the models are not completely equivalent. However it is not possible to determine which model is the best because this always depends on the actual situation. The comparison of the three modelling approaches can help to get a better understanding of SIS epidemics.

Keywords: SIS epidemic, ODE model, agent-based model, stochastic model

1. INTRODUCTION

Models of epidemics have been of interest for humanity for a long period of time. They are important to study the mechanism by which diseases spread, predict the courses of epidemics and to evaluate plans to control them. Common questions are interventions like vaccination strategies and isolation plans.

There are different types of epidemics and different concepts to model them. One of the first epidemic models was created by Kermack and McKendrick in the 1920s (Kermack and McKendrick 1927). They used ordinary differential equations (ODEs) to model SIS epidemics, the most simplified approach for contagious diseases. Today this serves as a classical model and is still used.

This concept can also be transferred to other modelling techniques. In this paper three approaches are considered: an agent-based model (AB model), an ODE model and a stochastic model (STOCH model). The purpose is to find out if these approaches are appropriate for epidemic simulation and to get a deeper

understanding of the spread of epidemics. Since the basic characteristics of these techniques are different, it is important to have a closer look if the models are really equivalent and to compare the models in various scenarios.

The goal is to compare the three different approaches to model a simplified SIS epidemic with one and two serotypes in various scenarios and to analyse the differences if they exist.

2. SIMPLE SIS MODEL

An SIS model is a simplified concept of contagious diseases that describes the spread of an infectious disease within a constant and homogeneous population (denoted by N) of individuals. This means that there are no births, deaths and migration and all individuals are of the same type. The state of every person is either infected (denoted by I) or susceptible (denoted by S). Since the population is constant and there are only two states Equation (1) holds at every point in time t :

$$S(t) + I(t) = N \quad (1)$$

The susceptible individuals can be infected through contact with an infected individual and stay infected for a specified constant period of time. After this time, they change their state from infected to susceptible and they can potentially be infected again (see Figure 1).

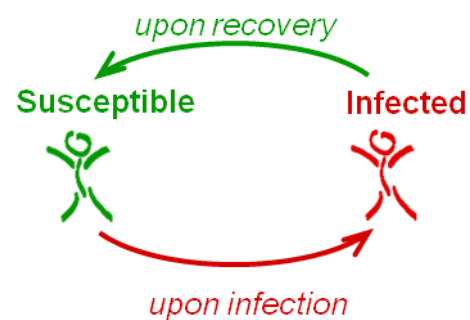


Figure 1: Illustration of the States and Transitions of the Simple SIS Model

There is no immunity or any other disease states. The transmission of infection is direct. There is no period of latency, this means the time period between the transmission of infection and the initiation of the infectivity does not exist.

2.1. Ordinary Differential Equation Model

The ODE model splits the population in two parts, the infected and the susceptible part; these are the two state variables of the model. The infected part is denoted by I , the susceptible part by S . The size of the states varies in time, because susceptible can become infected and vice versa. Figure 2 shows the two states S and I and the transitions between them. The transition rate from S to I is α , from I to S it is described by β .

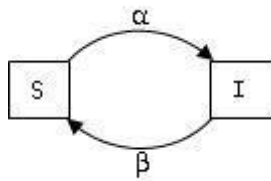


Figure 2: Illustration of the States and Transitions of the ODE Model

The parameters of the ODE model are α and β . α is the average number of transmissions per person per time unit caused by contacts between susceptible and infected individuals. α also depends on the infection probability. β is the ratio of recovery.

The ODE model consists of two ODEs (Equation (2) and (3)) with initial conditions (4) and (5) which describe the shift of the susceptible and the infected part of the population. The ODE model is similar to the model of McKendrick and Kermack.

$$\dot{I}(t) = \alpha I(t)S(t) - \beta I(t) \quad (2)$$

$$\dot{S}(t) = -\alpha I(t)S(t) + \beta I(t) \quad (3)$$

$$I(0) = I_0 > 0 \quad (4)$$

$$S(0) = N - I_0 \quad (5)$$

Equation (2) describes the change of the number of the infected part in time. The first term on the right hand side of Equation (2) represents the newly infected and the second term the recovered. Equation (3) describes the change of the number of the susceptible part in time. Since the population is constant the right hand side of Equation (3) varies only in the sign from the right hand side of Equation (2).

The two ODEs are analytically solvable and the solution of the ODE system is unique.

2.2. Agent-Based Model

The AB model describes the behaviour of the single individuals. Therefore it is an intuitive approach. An individual can either be infected or susceptible. All

agents are identical except for their states. Individuals have contacts with other individuals and these contacts are created randomly within the population. Susceptible individuals can be infected when they meet infected individuals. This process represents the contagion. For one susceptible agent the contacts and the contagion are illustrated in Figure 3.

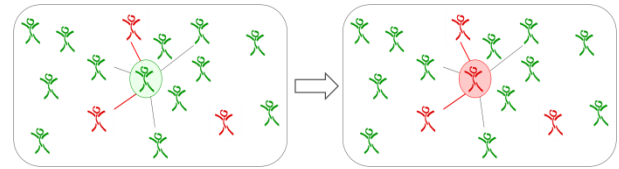


Figure 3: Illustration of the Contagion of a Susceptible Agent

After a fixed period of time an infected agent becomes susceptible again (see Figure 4).

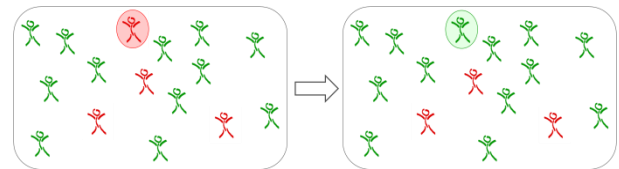


Figure 4: Illustration of the Recovery of an Infected Agent

The spread of the disease occurs upon the above-mentioned rules and is not defined explicitly.

The parameters of the model are the probability of infection ω , the number of contacts per agent κ and the duration of disease γ .

For the comparison with the other models, the numbers of infected and susceptible individuals are aggregated.

2.3. Stochastic Model

The STOCH model treats the two states susceptible and infected. The model consists of two equations, one for the number of infected individuals and one for the number of susceptible individuals.

The parameters of the model are identical to the parameters of the AB model, namely the probability of infection ω , the number of contacts per individual κ and the duration of disease γ .

The number of infected individuals for the present time step is determined dependent on the number of infected and susceptible individuals of the previous time steps (Equation (6)). In Equation (6) the newly infected individuals of time step t are represented by $ni(t)$ and $r(t)$ is the number of the recovered at time step t . The number of the susceptible individuals results from the difference of the size of the population and the number of infected individuals (Equation (7)).

$$I(t) = ni(t - 1) + I(t - 1) - r(t - 1) \quad (6)$$

$$S(t) = N - I(t) \quad (7)$$

The part of the population which changes the state is calculated in discrete time steps.

2.4. Comparison of the Models

The three models use different parameters (see Table 1).

Table 1: Parameters of the Models

ODE Model	AB Model and STOCH Model
α	ω
β	γ
	κ

In order to compare the models in equivalent settings it is important to know the relations between the parameters.

$$\alpha = \frac{\kappa \cdot \omega}{N} \quad (8)$$

$$\beta = \frac{1}{\gamma} \quad (9)$$

The Equations (8) and (9) show the relation of the parameters and therefore the models became comparable by parameters that represent the same situation.

These relations have been found by the following considerations: In the STOCH model the number of contacts between I and S individuals with transmission of infection per time unit is $\frac{\kappa \cdot I \cdot S \cdot \omega}{N}$. To obtain the parameter α from the ODE model the last expression has to be divided by $I \cdot S$ (see Equation (8)). The duration of the disease is the reciprocal of the ratio of recovery (see Equation (9)).

In general, the results of the models show that in each model a steady state is reached after an adaption phase in the beginning of the simulation.

The steady states of the three models are independent of the initial value of the number of the infected individuals and only depend on parameters that represent infection probability, recovery time and contact number.

The AB model and the STOCH model have similar adaption phases, while the adaption phase of the ODE model is different. The ODE model and the STOCH model reach the same steady states and the steady states of the AB model are slightly different.

During the adaption phase of the AB model and the STOCH model a so-called “overshooting” occurs (see Figure 6 and 7). The ODE model has no “memory” and therefore there exists no “overshooting” (see Figure 5). That means the number of the infected depend only on the number of the infected of the previous time step. The AB model and the STOCH model use numbers from earlier time steps too.

For the results shown in Figure 5, 6 and 7 the size of the population N is 10000 and the initial value of the number of infected individuals I_0 is 1000.

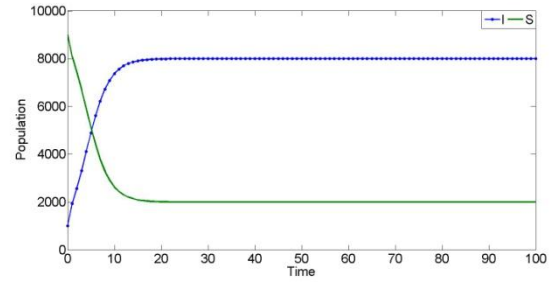


Figure 5: ODE Model with Parameters: $\alpha = 5 \cdot 10^{-5}$; $\beta = 0.1$

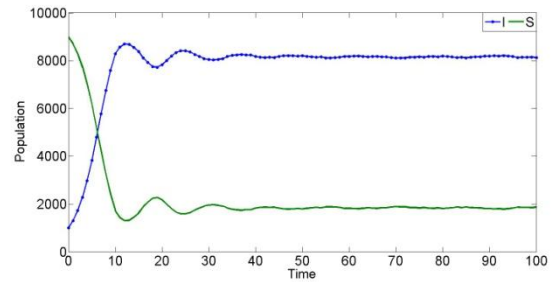


Figure 6: AB Model with Parameters: $\omega = 0.05$; $\gamma = 10$; $\kappa = 10$

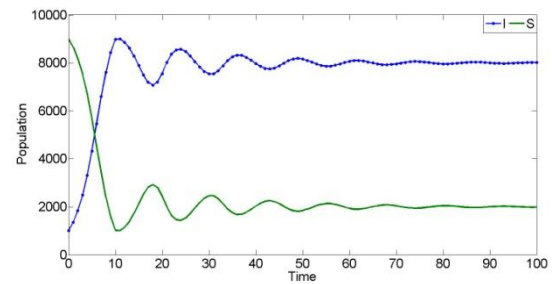


Figure 7: STOCH Model with Parameters: $\omega = 0.05$; $\gamma = 10$; $\kappa = 10$

The quantitative discrepancy between the results of the models is small and therefore the three models are appropriate methods to simulate an SIS epidemic.

3. SIS MODEL WITH TWO SEROTYPES

In contrast to the simple SIS model there are three states: individuals can either be susceptible or infected with serotype 1 or infected with serotype 2. A serotype is a variation of the pathogen of a contagious disease. The assumption for the simulated diseases is that an infected person can only be infected by one serotype at one time. An infected individual can become infected with the other serotype and then it loses the original serotype (see Figure 8). This property of the model is necessary, because models become unstable when infected individuals are not susceptible for the other serotype (Urach 2009; Bauer, Pöll, and Winterer 2011) and therefore the model is unusable.

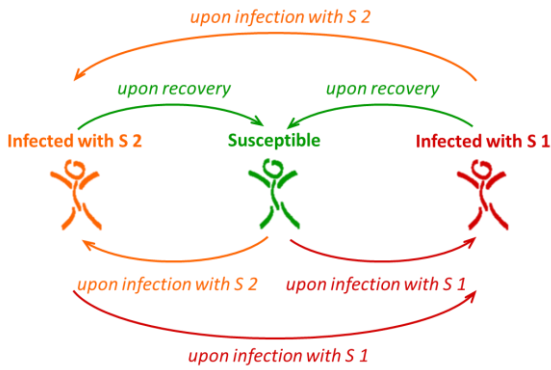


Figure 8: Illustration of the States and Transitions of the SIS Model with Two Serotypes

Further assumptions are the same as for the simple SIS model.

3.1. Ordinary Differential Equation Model

The ODE model splits the population in three state variables: I (infected with serotype 1), J (infected with serotype 2) and S (susceptible). Figure 9 shows the three states S, I and J and the transitions between them.

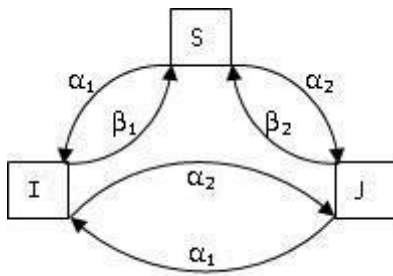


Figure 9: Illustration of the States and Transitions of the ODE Model with Two Serotypes

The parameters of the ODE model are α_1 , α_2 , β_1 and β_2 . α_1 and α_2 are the average numbers of transmissions per person per time unit caused by contacts between susceptible and infected individuals. α_1 and α_2 also depend on the infection probability of each serotype. β_1 and β_2 are the ratios of recovery of serotype 1 and 2.

The ODE model consists of three ODEs (Equation (10), (11) and (12)) with initial conditions (13), (14) and (15). Every ODE represents the change of the number of individuals of one state. The ODEs are more complex than the ODEs of the simple SIS model.

$$\dot{I}(t) = \alpha_1 I(t) S(t) - \beta_1 I(t) - \alpha_2 I(t) J(t) + \alpha_1 I(t) J(t) \quad (10)$$

$$\dot{J}(t) = \alpha_2 J(t) S(t) - \beta_2 J(t) + \alpha_2 I(t) J(t) - \alpha_1 I(t) J(t) \quad (11)$$

$$\dot{S}(t) = -\alpha_1 I(t) S(t) - \alpha_2 J(t) S(t) + \beta_1 I(t) + \beta_2 J(t) \quad (12)$$

$$I(0) = I_0 > 0 \quad (13)$$

$$J(0) = J_0 > 0 \quad (14)$$

$$S(0) = N - I_0 - J_0 \quad (15)$$

The ODE system cannot be solved analytically, but at least the solution of the ODE system is unique.

3.2. Agent-Based Model

The AB model is an extension of the AB model for one serotype. The rules for the agents are very similar. It is easy to extend the AB model, because in the model only the phase of the transmission of infection has to be adjusted, because now an infected can be infected with the other serotype.

The parameters of the model are the probabilities of infection with serotype 1 ω_1 and serotype 2 ω_2 , the number of contacts per agent κ and the duration of the diseases γ_1 and γ_2 .

3.3. Stochastic Model

The STOCH model consists of three equations, one for the number of individuals infected with serotype 1, one for the number of individuals infected with serotype 2 and one for susceptible.

The parameters of the model are identical to the parameters of the AB model, namely the probabilities of infection with serotype 1 ω_1 and serotype 2 ω_2 , the number of contacts per individual κ and the duration of the diseases γ_1 and γ_2 .

The number of infected individuals of each serotype for the present time step is determined dependent on the number of infected and susceptible individuals of the previous time steps (Equation (16) and (17)). In Equation (16) and (17) the newly infected of time step t are represented by $ni_1(t)$ and $ni_2(t)$. $r_1(t)$ and $r_2(t)$ are the numbers of the recovered at time step t . c_1 (respectively c_2) is the part of the infected individuals with serotype 1 (respectively 2) which becomes infected by serotype 2 (respectively 1). The number of the susceptible individuals results from the difference of the size of the population and the number of infected individuals (Equation (18)).

$$I(t) = ni_1(t-1) + I(t-1) - r_1(t-1) - c_1(t-1) \quad (16)$$

$$J(t) = ni_2(t-1) + J(t-1) - r_2(t-1) - c_2(t-1) \quad (17)$$

$$S(t) = N - I(t) - J(t) \quad (18)$$

The equations are more complex than the equation in the STOCH model for one serotype.

The numbers of I, J and S are calculated in discrete time steps.

3.4. Comparison of the Models

The three models use different parameters (see Table 2).

Table 2: Parameters of the Models

ODE Model	AB Model and STOCH Model
α_1	ω_1
α_2	ω_2
β_1	γ_1
β_2	γ_2
	κ

The relations of the parameters of the three models are like the relations of the parameters of the simple SIS model (see Equation (19) and (20)).

$$\alpha_1 = \frac{\kappa \cdot \omega_1}{N}, \alpha_2 = \frac{\kappa \cdot \omega_2}{N} \quad (19)$$

$$\beta_1 = \frac{1}{\gamma_1}, \beta_2 = \frac{1}{\gamma_2} \quad (20)$$

This strategy enables further comparability of the three models.

There are two different cases for the parameterisation of the three models. On the one hand, the parameters for the two serotypes are equal (equally strong serotypes), on the other hand, there is a difference in at least one parameter of the serotypes (unequally strong serotypes). The models always reach a steady state. The steady states of the AB model and the STOCH model are independent of the initial settings. In the ODE model in case of equally strong serotypes the steady states are dependent on the initial settings, for unequally strong serotypes they are independent.

The main difference between the models is that coexistence of both serotypes is more often found for the AB model (see Figure 11). The STOCH model and the ODE model have almost no situations of coexistence for unequally strong serotypes. This means that the weaker serotype usually dies out (see Figure 10 and 12).

For the results shown in Figure 10, 11 and 12 the size of the population N is 10000 and the initial value of the number of infected individuals with serotype 1 I_0 and serotype 2 J_0 is both 1000.

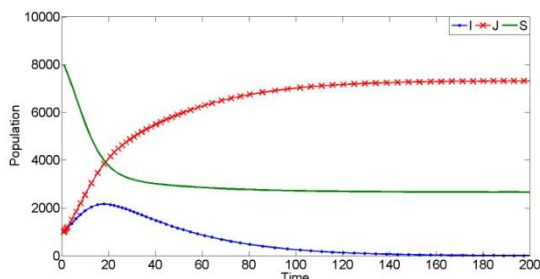


Figure 10: ODE Model with Two Serotypes with Parameters: $\alpha_1 = \alpha_2 = 2.5 \cdot 10^{-5}$; $\beta_1 = 0.1$; $\beta_2 = 1/15$

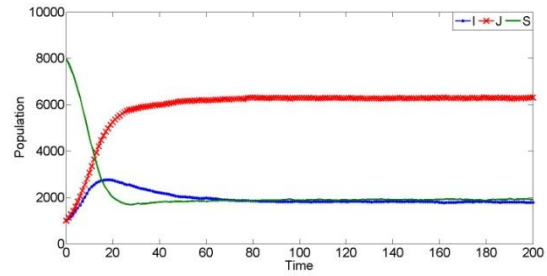


Figure 11: AB Model with Two Serotypes with Parameters: $\omega_1 = \omega_2 = 0.05$; $\gamma_1 = 10$; $\gamma_2 = 15$; $\kappa = 5$

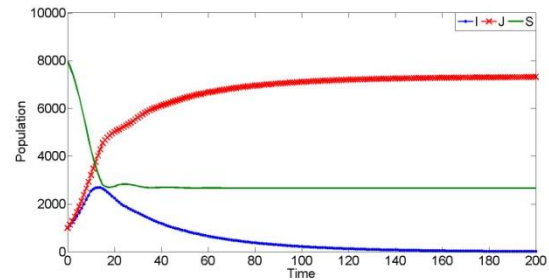


Figure 12: STOCH Model with Two Serotypes with Parameters: $\omega_1 = \omega_2 = 0.05$; $\gamma_1 = 10$; $\gamma_2 = 15$; $\kappa = 5$

In case of two equally strong serotypes, the steady states are different for all three models. In the cases where one serotype dies out in all three models, the steady states of the STOCH model and the ODE model are the same and the steady state of the AB model is very similar, while the adaption phase is similar for the STOCH model and the AB model and different for the ODE model. These results are obvious, because after extinct of one serotype the models are equivalent to the simple SIS models. There are two problems: First, for the ODE model and the STOCH model there is coexistence of both serotypes only in special cases. Second, the steady states of the models are quantitatively different. These problems show that further research for concurrent serotypes is required.

4. CONCLUSIONS

There are many possibilities to model a real problem. In this paper, three modelling approaches with different properties for an SIS epidemic have been compared in equivalent settings and there are some situations in which the three models agree and some in which they do not agree. In the latter case, a special validation of the results is needed to see which model gives the most realistic results. It is not possible to say only by comparison which modelling approach is the best, however it clearly shows qualitative and quantitative differences and similarities of these approaches. It always depends on the given situation which approach is right for the problem. In the cases where all three models agree, the models are considered to be valid.

Cross-validation is an important method in the field of modelling (Bharathy and Silverman 2010). It describes the attempt to solve the question with an alternative model and compare it with the actual model. If the

alternative model provides the same results the actual model might be considered to be more reliable, and consequently more valid.

In this study we show that the three modelling approaches all result in the same steady states for the simple SIS model, even though they have different adaption phases. The adaption phases can be explained by structural differences of the approaches. However, it seems that all three models are suitable to model SIS epidemics.

The situation is more complex for two competitive serotypes. In that case only the agent based model produces a stable coexistence of both serotypes while the ODE model and the STOCH model lead to an unstable behaviour. This means that the models simulate different situations caused by different structures, although their assumptions are all the same. We conclude that the choice of a model for SIS epidemics with competitive serotypes highly depends on the actual situation. It is important to examine structures and values in the real world crucially, compare them with the model and choose the right approach upon this knowledge. However, only the AB model is able to produce stable coexistence of serotypes. Usage of this model will be subject to further validation methods depending on the actual problem. This work might include further examination and revision of the ODE and STOCH model to find out more about consequences of structural differences.

Finally, the comparison of the modelling approaches can help to show in which situations models are valid and to get a better understanding of SIS epidemics.

REFERENCES

- Bauer, A., Pöll, C., Winterer, N., 2011. *Vergleich verschiedener Modellbildungsansätze anhand einer SIS Epidemie*. Thesis (BSc). Vienna University of Technology.
- Bharathy, G.K., Silverman, B.G., 2010. *Validating Agent Based Social Systems Models*. Proceedings of the 2010 Winter Simulation Conference. December 5-8, 2010, Baltimore, MD.
- Kermack, W.O., McKendrick, A.G., 1927. *A Contribution to the Mathematical Theory of Epidemics*. University of Texas at Austin. Available from: <http://www.ma.utexas.edu/users/davis/375/LECTURES/L26/km.pdf> [accessed 13 May 2012]
- Urach, C., 2009. *Modellierung und Simulation von Impfstrategien gegen Pneumokokkenkrankungen*. Thesis (Dipl.-Ing.). Vienna University of Technology.

AUTHOR'S BIOGRAPHIES

BAUER ANDREAS

After finishing high school and his social service, he went on to study technical mathematics at the Vienna University of Technology. In 2011 he finished his Bachelor studies. Now he continues studying for his Master's degree.

PÖLL CARINA

She finished high school in 2007. Then she began to study technical mathematics at the Vienna University of Technology. After she finished her Bachelor studies in 2011, she continues studying for her Master's degree.

WINTERER NINA

In 2007 she finished high school. Afterwards she started to study technical mathematics at the Vienna University of Technology. In 2011 she finished her Bachelor studies. Currently she is studying in the Master program.

A KALMAN FILTERING BASED APPROACH FOR THE MODELING OF THE CARDIOVASCULAR REGULATION SYSTEM

Brett Matsuka^(a), Jesper Mehlsen^(b), Mette S. Olufsen^(c), Hien T. Tran^(d), Nakeya D. Williams^(e)

^(a,c,d,e)Department of Mathematics, North Carolina State University, Raleigh, NC

^(b)Coordinating Research Center, Frederiksberg Hospital, Denmark

^(a)bjmatzuk@ncsu.edu, ^(b)Jesper.Mehlsen@frh.regionh.dk, ^(c)msolfuse@ncsu.edu, ^(d)tran@math.ncsu.edu,
^(e)ndwilli5@ncsu.edu

ABSTRACT

Short-term cardiovascular responses to orthostatic stress involve complex interactions among various mechanisms of short-term cardiovascular and respiratory blood flow and blood pressure control. Clinical procedures such as sit-to-stand and head-up-tilt (HUT) are often used to assess a patient's ability to regulate blood pressure. In this paper, we present a cardiovascular regulation model capable of predicting blood flow, volumes, and pressures in the systemic circulation during HUT. The cardiovascular regulation model adjusts cardiac contractility, vascular resistance and arterial compliance in response to changes in blood pressure. In previous studies, these control variables were modeled using a piecewise linear function parameterization. In this paper, we present an on-line estimation procedure based on the ensemble transform Kalman filter (ETKF) to estimate these dynamic control variables.

Keywords: cardiovascular regulation system, head-up-tilt, ensemble transform Kalman filter, parameter estimation

1. INTRODUCTION

The main role of the cardiovascular system is to maintain a set level of oxygen and nutrients in tissues as well as to ensure continuous removal of carbon dioxide and other metabolites. This is accomplished through tightly regulated control mechanisms. One of the main control systems promoting this regulation is the baroreflex system, which is part of the autonomic nervous system. Quantities being controlled by baroreflex regulation include blood flow and blood pressure. These quantities are kept close to their reference levels by a complex feedback control system regulating heart rate and vascular tone. Failure of this system has clinically significant consequences including dizziness, falls and reflex mediated syncope (Zaqqa and Massuni 2000), in particular for the elderly and for patients with hypertension and diabetes. The underlying pathophysiology lead to regulatory failure, which can be difficult to analyze since the detailed physiology involved with blood flow and pressure control is not well understood, and it is difficult to study the complex regulatory responses experimentally. These facts

suggest that there is a need for development of more advanced methodologies to predict blood flow and pressure regulation.

Sit-to-stand as well as the more commonly used head-up-tilt (HUT) (Lanier, Mote and Clay 2011) tests are often used to assess patient's ability to regulate blood pressure, in particular for patients who suffer from frequent episodes of syncope, lightheadedness, or dizziness (Miller and Kruse 2005). During HUT test, the patient rests on a tilt-table in supine position. After steady values for pressure and heart rate are achieved the table is tilted to angle of 60-70 degrees. Upon tilting, gravity causes pooling of 500-1000 ml of blood in the lower extremities reducing venous return, which induces a reduction in cardiac filling, pressure and volume. As a response blood pressure in the upper body decreases, while blood pressure in the lower body increases. Baroreceptors located in the carotid sinuses sense the drop in blood pressure causing sympathetic activation and parasympathetic withdrawal, which in turn lead to an increase in heart rate, cardiac contractility, vascular resistance, and vessel tone (Robertson, Low and Polinsky 2004).

In this paper we consider the development of a cardiovascular model capable of predicting blood flow, volumes, and pressures in the systemic circulation during HUT. In essence, the mathematical model consists of three principle components: a lumped cardiovascular model predicting dynamics while the subject is resting in supine position; a model predicting dynamic changes in response to HUT; and a model predicting the impact of the baroreflex and other regulating systems on heart rate, cardiac contractility, arterial compliance, and vessel resistance. In this model, heart rate is used as an input, thus the parasympathetic heart rate regulation is implicitly accounted for in the model. For the other three control variables; cardiac contractility, arterial compliance and vascular resistance, we modeled them in our previous studies by a piecewise linear function parameterization approach (Olufsen, et al. 2005). However, since these control variables, in general, do not vary significantly around some baseline values, an alternative and very promising approach to model them is by using the nonlinear Kalman filtering. The Kalman filter, which was introduced in the 1960's, is a recursive algorithm that calculates the optimal state of the system by taking a

weighted average of the probability distribution from the model and the probability distribution from the measurement (Kalman 1960). It is deterministic in nature and characterizes the entire optimal estimate through the propagation of the mean and covariance of the estimate at each step. The filter is very powerful in several aspects: it takes explicitly into account the measurement errors, it takes measurement data into account incrementally, and it is an efficient and simple to implement computational tool. It is recursive in the sense that at each step, the updated estimate is found through the previously estimated state and the observation data at that step. The Kalman filter is useful for estimating both system dynamics and time-varying or constant parameters.

In the Kalman filter, the model dynamics are assumed to be linear and that the errors in both the model and the observations are Gaussian. However, if these restrictive assumptions do not hold, the Kalman filter fails and adjustments have to be made to account for them (Grewal and Andrews 2008). One approach to deal with the nonlinear model dynamics is by using sampling techniques to characterize the probability distributions of the state. In this paper, we consider the ensemble transform Kalman filter (ETKF) (Evensen 2009a, Evensen 2009b), which belongs to a broader category of filters known as particle filters, for both state and parameter estimation.

2. A CARDIOVASCULAR REGULATION MODEL

In essence, the cardiovascular regulation model consists of three principle components: a lumped cardiovascular model predicting dynamics while the subject is resting in supine position; a model predicting dynamic changes in response to HUT; and a model predicting the effects of cardiovascular regulation.

2.1. A lumped cardiovascular model

The basic cardiovascular model comprises 5 compartments (see Fig. 1) representing arteries and veins in the upper and lower body of the systemic circulation, respectively, and the left ventricle. Here, the upper body compartments represent arteries and veins in the head, thorax, and abdomen, while the lower body compartments represent all vessels in the legs. The model mimics an electrical RC-circuit with voltage analogous to pressure, current analogous to flow, charge analogous to volume, compliance analogous to capacitance, while resistance is the same in both formulations. Note that with this model, it can predict blood pressure and flow in the various compartments, while it cannot predict the actual wave-propagation in the arterial network. Therefore, the model is well suited for predicting systolic and diastolic pressure values. Abbreviations (subscripts) in Fig. 1 are given in Table 1.

For each compartment, a pressure-volume relation is defined as

$$V - V_{un} = C(p - p_{ext}), \quad (1)$$

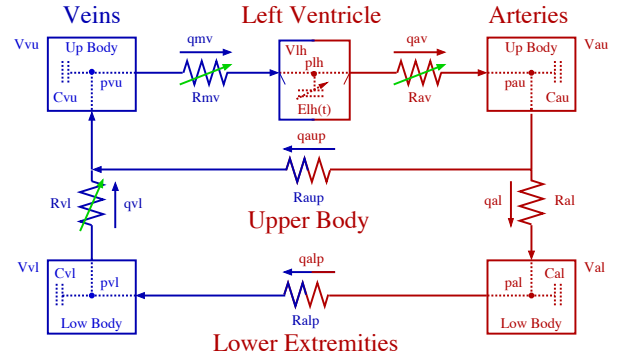


Figure 1: Compartmental model of systemic circulation. The model consists of 5 compartments with resistances in between compartments. The two heart valves, the mitral valve and the aortic valve, are modeled as time-varying resistors R_{mv} and R_{av} , respectively. In addition, the resistance between the lower and upper body veins V_{vl} is also modeled as time-varying function to prevent retrograde flow into the lower-body during the HUT

Abbreviation	Name
<i>av</i>	aortic valve
<i>au</i>	upper body arteries
<i>al</i>	lower body arteries
<i>aup</i>	upper body “peripheral” vascular bed
<i>alp</i>	lower body “peripheral” vascular bed
<i>vu</i>	upper body veins
<i>vl</i>	lower body veins
<i>lh</i>	left ventricle

Table 1: Abbreviations (subscripts) used in the cardiovascular regulation model

where V is the compartment volume, V_{un} is the unstressed volume, C is compartment compliance, p is instantaneous pressure, and p_{ext} is the pressure in the surrounding tissue. Using Ohm’s law, the volumetric flow q and the pressure p are related by the relation

$$q = \frac{p_{in} - p_{out}}{R}, \quad (2)$$

where p_{in} and p_{out} are the pressure in and out of the resistor R , respectively. Differentiating (1) and using Ohm’s law (2) yields the following system of differential equations for the pressure in the four arterial and venous compartments

$$\frac{dp_i}{dt} = \frac{1}{C} \frac{dV}{dt} = \frac{p_{i-1} - p_i}{R_{i-1}} - \frac{p_i - p_{i+1}}{R_i}. \quad (3)$$

For the heart compartment, the change in the volume is given by

$$\frac{dV_{lh}}{dt} = q_{in} - q_{out}. \quad (4)$$

Here, the heart pressure is predicted from the volume using the pressure-volume relation

$$p_{lh} = \frac{1}{C_{lh}}(V_{lh} - V_{un}) = E_{lh}(V_{lh} - V_{un}), \quad (5)$$

where E_{lh} is the elastance (the reciprocal of the compliance). Pumping is modeled by introducing a variable elastance function of the form (Ellwein 2008)

$$E_{lh}(\tilde{t}) = \begin{cases} \frac{E_{\max} - E_{\min}}{2} \left[1 - \cos\left(\frac{\pi\tilde{t}}{T_M}\right) \right] + E_{\min}, \tilde{t} \in [0, T_M] \\ \frac{E_{\max} - E_{\min}}{2} \left[1 + \cos\left(\frac{\pi(\tilde{t} - T_M)}{T_R}\right) \right] + E_{\min}, \tilde{t} \in [T_M, T_M + T_R] \\ E_{\min}, \tilde{t} \in [T_M + T_R, T], \end{cases}$$

where \tilde{t} is the time within a cardiac cycle, E_{\min} and E_{\max} denote the maximum and minimum elastance, respectively. Values for the cardiac cycle T (which is the reciprocal of the heart rate) and T_M are obtained from data, while T_R is a model parameter. Finally, heart valves are modeled using time-varying resistors for which a large resistance R_{cl} represents a closed valve and a small resistance R_{op} represents an open valve. These time-varying resistors have the forms

$$R_v = R_{cl} - \frac{R_{cl} - R_{op}}{1 + e^{-\beta(p_{in} - p_{out})}},$$

where p_{in} and p_{out} denote the pressure in and out of the valve, respectively.

In summary, the five differential equations representing the five compartments are given by

$$\begin{aligned} \frac{dp_{au}}{dt} &= (q_{av} - q_{al} - q_{aup}) / C_{au}, \\ \frac{dp_{al}}{dt} &= (q_{al} - q_{alp}) / C_{al}, \\ \frac{dp_{vl}}{dt} &= (q_{alp} - q_{vl}) / C_{vl}, \\ \frac{dp_{vu}}{dt} &= (q_{aup} + q_{vl} - q_{mv}) / C_{vu}, \\ \frac{dV_{lh}}{dt} &= q_{mv} - q_{av}, \end{aligned}$$

where the volumetric flow q_i is written in terms of the pressure and resistance as in equation (2) (e.g., $q_{av} = (p_{lh} - p_{au}) / R_{av}$).

2.2. Modeling HUT

The response to HUT is modeled by accounting for hydrostatic pressure acting on the affected compartment. During supine position, gravity does not influence the system. Upon HUT, blood is pooled in the lower extremities leading to an increase in pressure in the lower body, while pressure in the upper body decreases. To account for gravity, the pressure at the level of the carotid arteries were used as a reference pressure, so an extra term is added to the flows to (q_{al}) and from (q_{vl}) the lower body compartments. In particular, the modified flow equations are calculated as follows (Olufsen, et al. 2005)

$$q = \frac{\rho g h \sin(\theta(t)) + p_{in} - p_{out}}{R},$$

where

$$\theta(t) = \frac{\pi}{180} \begin{cases} 0, & t < t_{st} \\ v_t(t - t_{st}), & t_{st} \leq t \leq t_{st} + t_{ed} \\ 60, & t > t_{st} + t_{ed}. \end{cases}$$

Here, ρ is the blood density, g is the gravitational constant, h is the distance between the lower and upper compartments, $\theta = 60^\circ$ is the tilt angle, $v_t = 15$ is the tilt speed, and t_{st} and t_{ed} are the starting and ending times of the tilt, respectively.

2.3. Modeling effects of cardiovascular regulation

Upon HUT blood pressure decreases, leading to sympathetic activation and parasympathetic withdrawal. Parasympathetic withdrawal elicits increase in heart rate and cardiac contractility, while the sympathetic response elicits increase in vascular resistance and compliance. The model developed in this paper does not describe in detail sympathetic and parasympathetic afferents, but predicts the impact of the baroreflex and other systems regulating heart rate, cardiac contractility, arterial compliance, and vessel resistance. Heart rate is used as an input, thus the parasympathetic heart rate regulation is implicitly accounted for in the model. Increase of cardiac contractility is modeled by controlling the minimum elastance function of the left heart (E_{\min}), while the decrease of arterial compliance was incorporated in the upper body through regulation of (C_{au}). Finally, regulation of vascular resistance is included in both the upper and lower body. The upper body compartment includes abdominal and intestinal vessels, while the lower body compartment lumps vessels in the lower extremities. Consequently, both R_{aup} and R_{alp} are regulated. However, the compartments representing the upper and lower body arteries appear in parallel; hence, both resistances are not identifiable. In this paper, we controlled R_{aup} directly and let $R_{alp} = kR_{aup}$ with k denotes the ratio of the optimized rest values of R_{aup} and R_{alp} .

As discussed above, three quantities ($E_{\min}, C_{au}, R_{aup}$) are controlled to counteract the effect of the tilt. However, since the development of an accurate physiologically based model for these control variables is a rather nontrivial task, we utilize ensemble transform Kalman filter to estimate these time varying control functions by treating them as unknown parameters in the model.

3. THE ENSEMBLE TRANSFORM KALMAN FILTER

To begin the discussion on the ETKF, we consider a state space model of the form

$$\begin{aligned} x_{i+1} &= f(i, x_i, \theta) + w_i, \\ y_{i+1} &= h(i+1, x_{i+1}, \theta) + v_{i+1} \end{aligned}$$

where f is a nonlinear function of the state $x_{i+1} \in \mathfrak{R}^n$, and h is a nonlinear function relating the observation, $y_{i+1} \in \mathfrak{R}^m$, to the state. w_i and v_{i+1} are independent and identically distributed (i.i.d.) noise processes, and θ are the model parameters. The problem is to find the optimal parameter estimates to give the state the best fit to the data. Classically, parameter estimation is carried out by using an optimization technique to minimize a cost functional, which, in general, is the sum of the squares of the difference between the model observation and the data. Utilizing the filtering technique to estimate parameters, there are multiple methods in which we can achieve our goal. First, and most direct, is to modify the state space representation to accommodate our objective. Since we are assuming the observations are coming from our model and the parameters are our main focus, we can proceed by assuming that the parameters have zero dynamics

$$\begin{aligned}\theta_{i+1} &= \theta_i + r_i \\ y_{i+1} &= h(i+1, x_{i+1}, \theta_{i+1}) + v_{i+1},\end{aligned}$$

where r_i is i.i.d. noise process for the model parameter. However, our goal is to not only estimate the parameters, but also to fit the states as accurately as possible. This methodology is known as dual estimation. There are two methodologies that directly achieve this. First is joint estimation. This is achieved by concatenating the parameters into the state-space as follows

$$\begin{aligned}z &= [x; \theta] \\ z_{i+1} &= f(i, z_i) + w_i \\ \theta_{i+1} &= \theta_i + r_i.\end{aligned}$$

A second approach is called the dual filter. This is done by concurrently running a state filter and parameter filter in parallel. The state filter estimates the state, x_i , using the parameter value from the parameter filter at time, $i-1$. While, the parameter filter estimates the parameter, θ_i , using the state estimate at time, $i-1$. These both propagate forward in time and obtain estimates at each time step.

The main idea around the ensemble Kalman filter is to approximate the error statistics of our estimate by a set of particles sampled from the probability distribution. That is, we calculate the prior and posterior error covariances by the ensemble covariance matrices around the corresponding ensemble mean

$$\begin{aligned}P_{x_{i+1|i}} &= \frac{1}{K-1} U_{i+1|i} U_{i+1|i}^T \\ P_{\theta_{i+1|i+1}} &= \frac{1}{K-1} U_{i+1|i+1} U_{i+1|i+1}^T,\end{aligned}\quad (6)$$

where

$$\begin{aligned}U_{i+1|i} &= [x_{i+1|i}^1 - \hat{x}_{i+1|i}; x_{i+1|i}^2 - \hat{x}_{i+1|i}; \dots; x_{i+1|i}^K - \hat{x}_{i+1|i}] \\ U_{i+1|i+1} &= [x_{i+1|i+1}^1 - \hat{x}_{i+1|i+1}; x_{i+1|i+1}^2 - \hat{x}_{i+1|i+1}; \dots; x_{i+1|i+1}^K - \hat{x}_{i+1|i+1}]\end{aligned}$$

with $x_{i+1|i}^j$ being the j^{th} particle being propagated through the model, and $x_{i+1|i+1}^j$ is the update of each particle. With this, we define K to be the number of particles in which to approximate the distribution and

$$\hat{x} = K^{-1} \sum_{j=1}^K x^j.$$

From now on, we shall define U to be the ensemble perturbation matrix as it gives us the deviation of each particle from the mean. The essence of this methodology is that we just integrate each ensemble member in time through our dynamical model.

To obtain the update equations, we first compute the Kalman gain as follows

$$\begin{aligned}K_{i+1} &= P_{x_{i+1|i}} H^T (H P_{x_{i+1|i}} H^T + R)^{-1} \\ &= (K-1)^{-1} U U^T H^T (H (K-1)^{-1} U U^T H^T + R)^{-1} \\ &= (K-1)^{-1} U (H U)^T ((K-1)^{-1} (H U) (H U)^T + R)^{-1}.\end{aligned}$$

From the above equation, it is noted that anywhere the linear operator H appears, it is coupled to the ensemble perturbation matrix, U . Due to this consequence, given a nonlinear observation function h , we can take advantage by replacing HU with the following approximation

$$V = [h(x_{i+1}^1) - \hat{y}_{i+1}; h(x_{i+1}^2) - \hat{y}_{i+1}; \dots; h(x_{i+1}^K) - \hat{y}_{i+1}],$$

where \hat{y}_{i+1} is the mean of the observation function

$$\hat{y}_{i+1} = \sum_{j=1}^K h(x_{i+1}^j).$$

given each sample, To obtain our final estimates, we apply the classical Kalman filter update equations to each ensemble member

$$x_{i+1|i+1}^j = x_{i+1|i}^j + K_{i+1} (y_{i+1}^k - h(x_{i+1|i}^j)),$$

where K_{i+1} is the Kalman gain defined as above, and the observation, y_{i+1} , is perturbed accordingly

$$y_{i+1}^k = y_{i+1} + \psi_{i+1}^k,$$

where ψ_{i+1}^k is a Gaussian random variable with mean zero and covariance R . This perturbation is a Monte Carlo method applied to the Kalman filter formula that yields an asymptotically correct analysis error covariance estimate for large ensemble sizes (Majda and Harlim 2012). In practice, to keep these perturbations unbiased, we generate these random perturbations by first randomly drawing a $M \times K$ matrix A , where K is the number of particles and M is dimension of the data, and take the singular value decomposition of $(K-1)^{-1} A A^T = F \Sigma F^T$. Therefore, we have unbiased random vectors of y_{i+1}^k , which are just the column vectors of the matrix (Majda and Harlim 2012) $T = ((K-1)R)^{1/2} F \Sigma^{-1/2} F^T A$. Since we are perturbing the measurements, we can define the measurement error covariance matrix to be

$$R = (K - 1)^{-1} E E^T,$$

where $E = [e_1, e_2, \dots, e_K]$ are the ensemble measurement perturbations, with mean zero. Other alternative implementation of the ensemble Kalman filter also exists (Evensen 2009b).

The main idea of the ensemble transform Kalman filter (ETKF) is to take the square root of the covariance matrix, and transform it to a space where it is more robust and well conditioned (Evensen 2009a). There are multiple methods for taking the square root, the one we utilized in this paper is the ETKF as derived in (Bishop 2001). The basic idea is to find a transformation matrix T so that

$$U_{i+1|i+1} U_{i+1|i}^T = U T$$

and

$$U T (U T)^T = (K - 1) R_{i+1|i+1},$$

where $R_{i+1|i+1}$ is the sample posterior covariance matrix as previously defined. The posterior ensemble is then generated by taking the posterior mean and adding each column vector from U . Using the identity

$$A^T (A A^T + R)^{-1} = (I + A^T (R^{-1}) A)^{-1} A^T R^{-1},$$

and letting $V = A$, we apply this to the Kalman gain and arrive at the following expression

$$K_{i+1} = (K - 1)^{-1} U (I + (K - 1)^{-1} V^T R^{-1} V)^{-1} V^T R^{-1}.$$

Substituting the above expression and (6) into the posterior covariance formula, we obtain

$$P_{x_{i+1|i+1}} = \left(I - \frac{U}{K-1} (I + (K-1)^{-1} V^T R^{-1} V)^{-1} V^T R^{-1} H \right) \frac{U U^T}{K-1}.$$

Factoring out $U / (K - 1)$ yields

$$P_{x_{i+1|i+1}} = \frac{U}{K-1} \left(I - \left(I + \frac{V^T R^{-1} V}{K-1} \right)^{-1} \frac{V^T R^{-1} V}{K-1} \right) U^T.$$

Therefore, we finally obtain

$$P_{x_{i+1|i+1}} = \frac{1}{K-1} U (I - (I + B)^{-1} B) U^T,$$

where $B = V^T R^{-1} V / (K - 1)$. Using the identity

$$I - (I + B)^{-1} B = (I + B)^{-1}$$

we obtain

$$\begin{aligned} P_{x_{i+1|i+1}} &= U \left((K - 1) I + V^T R^{-1} V \right)^{-1} U^T \\ &= U \Sigma U^T = U \frac{T T^T}{K - 1} U^T. \end{aligned}$$

Applying this to the above analysis for the ensemble Kalman filter, we obtain the algorithm for the ensemble transform Kalman filter.

4. APPLICATIONS OF ETKF

The performance and feasibility of using the ensemble transform Kalman filtering based approach for modeling the cardiovascular regulation system will be illustrated using HUT data collected from a healthy young male volunteer age 37 who was fit and had no known heart or vascular diseases at the Coordinating Research Center (Frederiksberg Hospital, Copenhagen, Denmark). After resting for 10 minutes in supine position, the subject was tilted to an angle of 60 degrees

at a speed of 15 degrees per second measured by way of an electronic marker. The subject remained tilted for five minutes, and was then returned to supine position at the same tilt speed. For the model based analysis, we extracted a total of 690 seconds of data: including a 180 seconds segment recorded while the subject was resting in supine position and a 180 second segment recorded during HUT.

Literature values and subject specific information were integrated to identify nominal values for all model parameters (resistances, capacitors, heart, and tilt parameters) as well as to predict initial conditions for all state variables. Nominal parameter values were obtained by considering mean values for all pressures, flows, and volumes in the system obtained while the subject was in supine position (before the tilt). Then, using heart rate as an input, we estimate the three unknown model parameters ($E_{\min}, C_{au}, R_{aup}$) from measurements of arterial blood pressure in supine position and during HUT. The parameter estimation is carried out using the ensemble transform Kalman filter methodology discussed in Section 3. It should be noted that the cardiovascular regulation model presented in Section 2 predicts blood pressure and flow as pulsatile quantities, but since the model is analogous to an RC-circuit it does not allow for prediction of wave-propagation, consequently direct comparison of computed and measured values of blood pressure is not valid. To obtain adequate pulsatility, we identify model parameters that allow prediction of systolic and diastolic values of blood pressure. These values can be obtained from computing the maximum and minimum pressure within each cardiac cycle.

In the following, we picked noise covariances w_i and v_i to be $Q = 0.01 I$ and $V = 0.01$, respectively. The noise covariance r_i for the parameter model is $R = 1.0 \times 10^{-8} I$. The estimated blood pressure obtained from the ETKF using 50 ensemble members is depicted in Fig. 2 against the actual data. The figure shows very good agreement between the state estimate and the data.

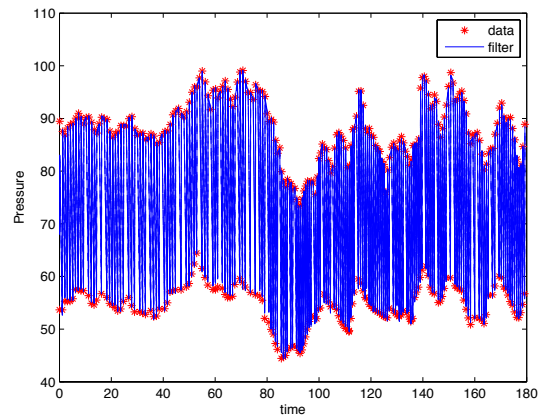


Figure 2: Blood pressure from the state estimates obtained using the ensemble transformed Kalman filter and the actual systolic and diastolic values of blood pressure are shown

Using the joint estimation, the estimates for the regulating parameters ($E_{\min}, C_{au}, R_{aup}$) are depicted in Figures 3, 4 and 5, respectively.

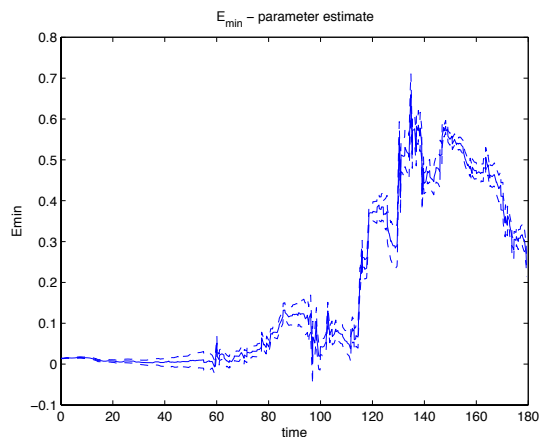


Figure 3: Profile of the time varying parameter E_{\min} obtained from the ETKF with 3 standard deviations (dashed)

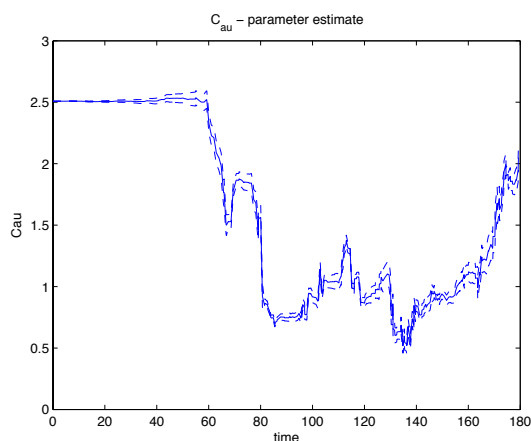


Figure 4: Profile of the time varying parameter C_{au} obtained from the ETKF with 3 standard deviations (dashed)

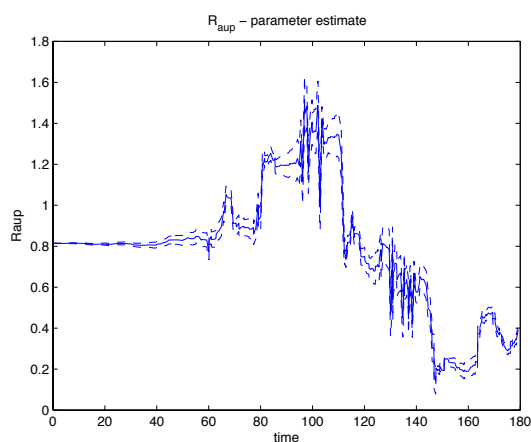


Figure 5: Profile of the time varying parameter R_{aup} obtained from the ETKF with 3 standard deviations (dashed)

5. CONCLUSION

In this paper we illustrate the feasibility of using the ensemble transform Kalman filter algorithm for the dual estimation problem of estimating both the state and unknown time-varying parameters in a cardiovascular regulation model. The ETKF is widely used in weather forecasting applications. Our initial efforts to apply the ETKF to biological applications such as the cardiovascular problem considered in this paper are very promising.

6. ACKNOWLEDGEMENTS

This research was supported in part by NIH-NIAD under grant 9R01AI071915-05, by NIH-NIGMS under grant 1P50-GM094503-01A1 and by NSF-DMS under grant 1022688.

REFERENCES

- Bishop, C.H., Etherton, B.J. and Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129, 420-436.
- Ellwein, L.M., 2008. Cardiovascular and respiratory regulation, modeling and parameter estimation. Thesis (PhD). North Carolina State University.
- Evensen, G., 2009a. *Data assimilation: The ensemble Kalman filter*. 2nd edition, New York:Springer.
- Evensen, G., 2009b. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29, 83-104.
- Grewal, M.S. and Andrews, A.P., 2008. *Kalman filtering: Theory and practice using MATLAB*. 3rd edition, Hoboken:John Wiley & Sons, Inc.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *ASME J. Basic Engineering*, 82, 34-45.
- Lanier, J.B., Mote, M.B. and Clay, E.C., 2011. Evaluation and management of orthostatic hypotension. *Am. Fam. Physician*, 84, 527-536.
- Majda, A. and Harlim, J., 2012. *Filtering complex turbulent systems*. 1st edition, United Kingdom:Cambridge University Press.
- Miller, T.H. and Kruse, J.E., Evaluation of syncope. *Am. Fam. Physician*, 72, 1492-1500.
- Olufsen, M.S., Ottesen, J.T., Tran, H.T., Ellwein, L.M., Lipsitz, L.A. and Novak, V., 2005. Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation. *J. Appl. Physiol.*, 99, 1523-1537.
- Robertson, D.W., Low, P.A. and R.J. Polinsky, R.J., 2004. *Primer on the autonomic nervous system*. 2nd edition, San Diego:Academic Press.
- Zaqa, M. and Massuni, A., 2000. Neurally mediated syncope. *Texas Heart Inst. J.*, 27, 268-272.

IFEDH - SOLVING HEALTH SYSTEM PROBLEMS USING MODELLING AND SIMULATION

Niki Popper^(a), Ingrid Wilbacher^(b), Felix Breitenecker^(c)

^(a)dwh Simulation Services, Vienna, Austria

^(b)Main Association of Austrian Social Security Institutions, Vienna, Austria

^(c)Vienna University of Technology, Institute for Analysis and Scientific Computing, Vienna, Austria

^(a)niki.popper@drahtwarenhandlung.at, ^(b)ingrid.wilbacher@hvb.sozvers.at, ^(c)felix.breitenecker@tuwien.ac.at

ABSTRACT

Health Technology Assessment (HTA) is used to inform decision makers on the level of the health care system or at the population level. The consortium IFEDH (Innovative Framework for Evidence based Decision support in Health care) was formed to invent a new strategy to integrate modelling and simulation into the process of HTA, which deals with limited resources and upcoming new technologies. Dynamic and static modeling is getting more and more essential for this process. Simulation can help decision makers to compare various technologies comparing different goal functions on a basis of evidence. Most important for this is a reliable and reproducible process of modeling and computing of simulations. We will present the developed fundamental framework which connects the most important areas for getting decision bases: Developing the PICO questions, data acquisition and evaluating of data, developing models and interpreting the results. IFEDH is funded by The Austrian Research Promotion Agency (FFG).

PROJECT

Health technology is one of the fastest growing industries and is of main interest for society concerning all layers of the population. Often resources are bounded and long time effects occur, therefore, decisions regarding strategies influence the healthcare policy for decades. Modeling and simulation as a tool for decision-making support is necessary to compute various scenarios. Because of these reasons the idea was to integrate more sophisticated methods of modeling and simulation in the process of HTA.

The main focus of the project lies on the development of these processes, leading an interdisciplinary group of experts in the field of HTA, statistics, modeling, visualization and database analysts through the whole decision support process. Within the last two years the focus was on the application of these methods on infectious diseases simulation and vaccination questions, but the derived methods are directly usable also for all other questions in HTA. Besides the complexity of guiding and controlling the

coupling of different scientific domains to realize a joint overall approach in model based HTA, in all sections new approaches had to be developed and implemented.

As decision support has to become faster, parameter sources and modular reusable model parts have to be developed in advance. But the modeling process and the design of adequate modeling methods are only one (core) part. Besides this, the model and parameter validation as well as verification of the developed and implemented system are in focus. The data quality assessments as well as the outcome visualization and interpretation are set up interdisciplinary, ensuring the necessary project quality and acceptance on the part of policymakers.

Scientific Partners in the project are Main Association of Austrian Social Security Institutions, Ludwig Boltzmann Institute for Health Technology Assessment, Vienna University of Technology (VUT) and VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH, company partners are dwh Simulation Services, E.I.S. Ltd. Florian Endel and FWD GmbH. UMIT- Private University of Health Sciences, Medical Informatics and Technology-GmbH is an additional contributor.

The first step in the project was the analysis of model know-how and structural know-how as well as gathering the state of the art of modeling in HTA in Austria.

Based on this information a specification of requirements regarding model structure and documentation of simulation outputs was set up.

The third step, representing one of the core working tasks of the network was the development of adequate/reusable modeling structures and modeling methods. Therefore an evaluation table of methods in use as well as modeling and simulation strategies from other domains with a potential use in HTA was acquired.

Furthermore modular model parts were developed and tested for their reusability. The analysis of data sources depending on each module as well as the realization of usability tables and interface descriptions finalized this task, ensuring high flexibility and

reusability. Based on this exploratory work was made on:

- different modeling techniques of infectious diseases (Zauner, Popper, Breitenecker (2010)),
- herd immunity effects of population groups using agent based modeling methods (Miksch et al. (2010)) and
- work of IFEDH members on serotype behavior modeling for infectious diseases and vaccination strategies (Zauner et al. (2010))

Based on this research recommendations for good practice were developed.

In act with the evaluation and integration of classical HTA methods, their development for data preparation and analysis with respect to Austrian reimbursement data a "good practice" manual was realized. The elaboration of standardized visualization concepts for

1. model parameters,
2. model structures and
3. the results,

together with research on scenario set up and sensitivity analysis workflow were integrated and tested for practical use, by implementation of three real world HTA questions.

At the beginning of the IFEDH research project stood the evaluation of the status quo acquired by the following tasks:

1. Documentation of standards in modeling and simulation in the field of health technology and health system evaluation
2. Documentation of standards in HTA: the documentation describes the standard process in vaccination program evaluation as well as the used methods and their limitations. This task is realized using expert opinions and a structured questionnaire to get the state of the art in Austria and neighbouring countries.
3. Documentation referring to identified problems with representation of solution pathways: The document lists the open questions, which appeared in earlier projects of partners using modeling and simulation for evaluation of vaccination scopes. For each problem the used solution strategies and the discussion about general use of this strategy are documented.

Based on this first step the second one was defined and realized: Definition of demand profiles regarding modeling and simulation in HTA.

As the whole process and the resulting service is developed necessarily by interdisciplinary partners, who have to work and understand each other, the definition

of a mutual language is a key to success. Therefore a glossary based on international definitions and formulations used by single project partners is realized. To guarantee the actuality, this document is defined as some sort of open document, being expanded during the whole life span of the project by the partners. Inconclusive or "parallel" formulations are discussed by project participants from different domains; the consensus is stringent for all partners.

The compiling of requirements regarding model structure and documentation of simulation outputs are completing the first project phase. The determination of needs helps to ensure an efficient modeling process in which the models subsequently have to be changed only from time to time.

QUESTION

The process of selecting a question for decision making - prioritization - and the balancing of the findings in the broader political context - appraisal - is not covered in this paper in detail. We address how to generate a reasonable question for the modeling process on basis of the PICO question.

To start working on a relevant question usually needs a clarification of the possible decisions, the definition of the population/ condition addressed, the intervention, the comparator and the outcomes of interest. This is the format of the PICO - Population; Intervention; Control; Outcome - question. In this phase of scoping the question has to be formulated with the best possible precision and a first discussion about the feasibility and necessity of modeling has to take place. To understand differences in the process of HTA knowledge of the health care system is necessary. The political decision making process can be visualized in the following way shown in Figure 1.

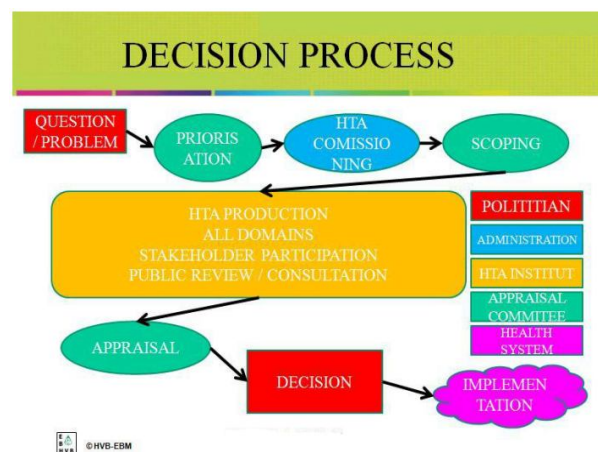


Figure 1: Decision Making Process

Following the PICO question the part of the HTA, which can be subsidized by modeling and simulation has to be specified. The decision which research questions can be answered using models and what is needed for doing this has to be taken. This is done by a

steering group as different aspects are important: The PICO question itself is always the starting point. Based on this data and structure of the problem has to be analyzed. In an iterative process the modeling technique will be developed and – again within the steering group – decided. To be able to handle this process and to guarantee a quality management, a special kind of documentation process was implemented.

DOCUMENTATION

At the end of this process the group has defined an exact definition of “not-model based” HTA questions and the questions which can be handled with methods of modeling and simulation. Based on these questions the modeling process has started and the method decision, as well as the exact specifications of the model were done. At the end the whole parameter and data set, which will be needed to implement the basic simulation as well as all comparative scenarios is defined and communicated within the interdisciplinary team.

On basis of this the process of data acquisition is started. In parallel this process has to be documented reliable and reproducible. IFEDH has developed and described a focused procedure (on basis of the EUnetHTA project and the HTA Core Model) how to describe the process and the status of used data. A web tool – the HTA-Manager - for documentation of source, status and manipulation of data is used

With the HTA-Manager the data needed for all simulation runs (the basic simulation – f.e. the status quo of a given therapy and all possible scenarios - f.e. a new vaccine or new regulatory requirements) can be documented permanently. To make this possible the documentation is done in three steps. The first step is to document the sources for all data sets. There are three different categories of used data: CKAN data, which is one of the world’s leading open-source data portal platforms, other open data sets and private or secret data, f.e. data of the insurance associations. After documentation of these sources as step 2 the modification or the adaption of all data sets is documented. These adaptations are f.e. SQL requests, filters or any other modification of the given raw data sets. Step 3 is the download or storage of the used data set for (1) all parameters , for (2) all simulation runs and scenarios into the HTA-Manager. Step 3 is the most important part of the documentation, as with the given data and a full documentation of the model and implementation the simulation runs for the basic system and all scenarios can be reproduced any time any place. This possibility is a main goal of the IFEDH project, to raise the credibility of modeling and simulation within the HTA community.

DATA

Another methodological goal of IFEDH was the development of standards and methods concerning data preparation and data analysis. The aspect of requirements of models referring to data and statistical

preparations is under treatment. Concerning the generated model structure definition of requirements on data and their statistical preparations are given. A special effort is put on the necessary quality and the granularity (i.e. how detailed data have to be provided). Again the work is realized in interdisciplinary groups benefiting from results that are comprehensible and usable in a general area.

After identifying the granularity and the data sources a concept of data quality assessment is acquired: The concept explains the data quality assessment that has to be performed using health data including theoretical principles, characteristics concerning health data and information referring to the implementation. The following figure (see Figure 2) motivates the connection and interaction of modeling & simulation, parameterization and data quality assessment. The influence of the quality of data input, parameter estimation and modeling structure (simplifications, unsecure assumptions ...) is discussed in an interdisciplinary manner. In an early project phase the goal is the identification of problems regarding parameterization and the sensitivity of diffuse parameters..

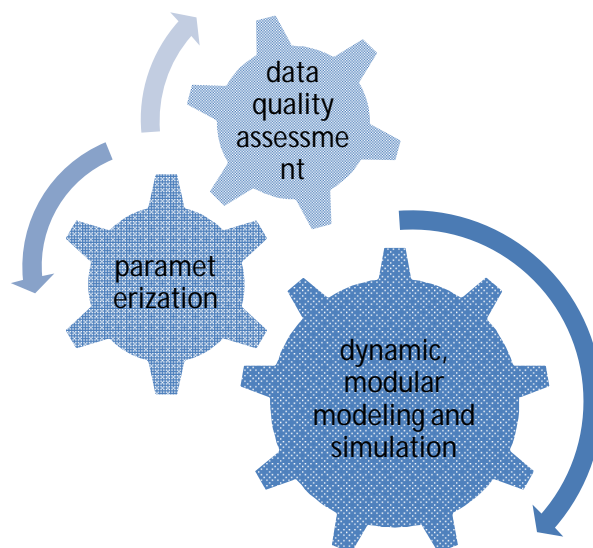


Figure. 2. The connection between dynamic modular models, their parameterization and the data quality assessment is shown.

The influence on the parameterization based on model structure and time intervals simulated is obvious. Furthermore data quality assessment influences the parameterization: Availability of data sources, reliability of parameters and range of confidence interval. Information about missing data quality or even lack of data results are necessary for changes in the model structure or in additional HTA literature work and searching for alternative parameterization attempts. Certainly the model has to fit the real life conditions in an adequate manner. Therefore a consensus in the interdisciplinary group has to be reached in the early project stage. Changes in a later phase generally cause

tremendous additional workload and bring about problems with the decision makers, if changes in the model structure have to be justified.

An important aspect within IFEDH was the usage of routine data in HTA, as gathering data tends to be an expensive and time consuming task. During the IFEDH research project different methods for using routine data on models in HTA were discussed, improved and developed. Connecting a rich dataset from Austrias inpatient sector lacking patient identifiers and (kind of) personalized but sparse records from the outpatient sector provided by different social security institutions is the objective of this project. A detailed description of the setup and usage of the results were presented at the SHIP Conference 2011 in St. Andrews (<http://www.scotship.ac.uk/conference-2011>) and the International Data Linkage Conference 2012 in Perth (<http://www.data-linkage2012.com.au/>).

Documentation of prior processing and information of the provided data were not fully available and also questionable data quality and the presence of possible duplicates result in technical and contextual challenges. After preprocessing, data quality assessment and other preparations, a deterministic record linkage approach was developed using a combination of the open and freely available statistical environment R and PostgreSQL database. Based on dynamically created SQL statements and extensive logging, the linkage process can be enhanced easily if new knowledge about the input data gets available.

The resulting linked dataset provides high quality and immediately available information. Additionally the deterministic linkage process can be examined and understood by its users. Therefore linkage and data errors are identified easily and feedback can be used to enhance the overall result. These experiences also lay the foundation for more advanced linkage methods and further improvements.

After the long and challenging way from the first data import to a functioning data collection, adequate information can now be used in different projects with low costs and users confidence.

MODELING

One of the core working tasks of the network is the development of adequate/reusable modeling structures and modeling methods. Modular model parts are developed and tested for their reusability. The analysis of data source depending to each module as well as the realization of usability tables and interface descriptions finalize this task, ensuring high flexibility and reusability (see Figure 3). Based on exploratory work on different modeling techniques of infectious diseases [1], research on herd immunity effects of population groups using individual based modeling methods [2] and work of IFEDH members on serotype behavior modeling for infectious diseases and vaccination strategies [3] recommendations for new questions are in progress.

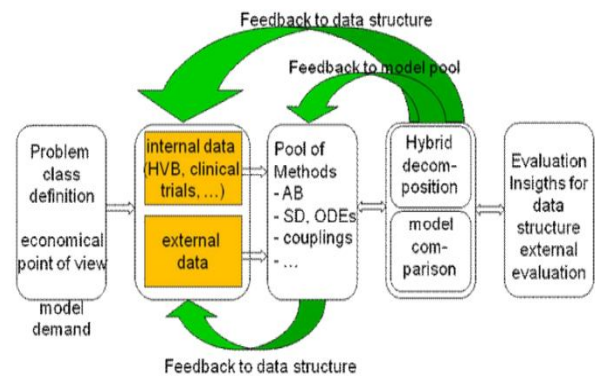


Figure 3: Basic Diagram of model based decision support

First of all, the use of modeling and simulation in HTA is discussed once again to see why this technical approach is used and, furthermore, to identify the raising questions:

- Formulation of the problem:
- What are the questions that shall be answered?
- Concept of the model: Which values are important, which describe the states of the model, which are parameters, which values influence the model in general? Which relations between the values exist?
- Is the model concept useful: Is enough knowledge and data accessible to implement the model? Can the proposed questions be answered using the model if the model assumptions are true?
- Can the model be validated?

Besides the identification of useable modeling methods in answering HTA questions the classification in respect of different points of view is integrated in the description. This classification can help to integrate experts for HTA into the process, even if they are not specialists in modeling and simulation. As classification

- Black-box versus white-box modeling
- Top-down and bottom-up approaches
- Classification representing time

are chosen. Especially the interpretation of differences in top-down and bottom-up modeling techniques as well as explanation of processing of time helps the HTA and data experts to understand what modeling and simulation can perform.

Beyond that the communication regarding how to provide data in usable formats is motivated. This motivation together with an extra task and proof of concept examples strengthen the quality of the developed service by the IFEDH partners.

In case of restrictions of the defined methods that cannot be solved using another method from the first part of the project, combination of different methods has to be discussed. Using this open approach; first defining the HTA questions using a PICO schema and then choosing the modeling methods based on the

research question and information about data structures and expected time behavior is depicted.

The starting point is the definition of the problem/the HTA question. The main important fact in this system is that the data structure analysis is performed before the modeling method is chosen. Nevertheless the modeling process is done iteratively. The hybrid decomposition as well as the comparison of different methods and modular use of pre-developed tasks is part of the new developed concept in the IFEDH research project.

The development of reusable parts of the model and especially the theoretical background is focused. Their advantages and disadvantages, restrictions and potential applications in the field of evaluation of vaccination strategy are discussed. A general framework (see Figure 4) is developed and tested using pre-defined proof of the concept – examples that are processed together by the different partners in IFEDH.

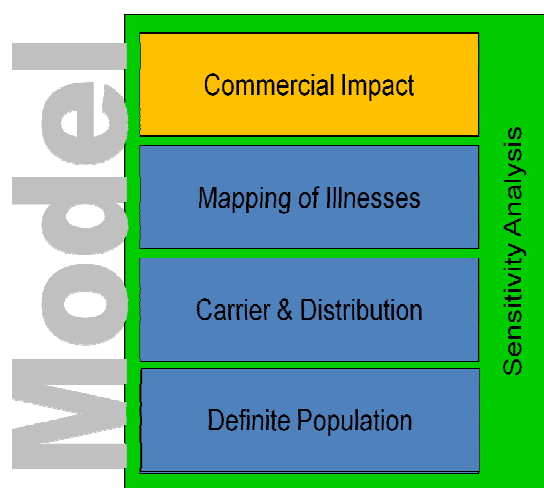


Figure 4. Example for splitting a cost effectiveness. Evaluation of a vaccination program for a given population into modules. The overall modeling concept can be discussed, implemented and parameterized in several reusable blocks.

The necessary question about the context between model and how to get the parameters into the framework is pointed out. To ensure the reliability and the quality of the datasets used following methods are discussed:

- sensitivity analysis – using this method the HTA and modeling experts get knowledge about the overall influence of parameters of interest,
- parameterization and calibration: Calibration is a systematic adjustment of model parameters that are either unknown or uncertain (Taylor, 2007). The strategy is to adjust these questioned parameters in a way so that the model results fit given data sufficiently.
- verification,
- validation, and, finally,

- simulation experiments and scenarios that provide results.

Calibration is a process of setting parameters, running the model, assessing the results, adjusting the parameters based on the results and running the model again. This procedure is performed until the results are satisfying. There are two types of situations that require calibration of a model (Taylor, 2007):

- When data are inadequate or missing to estimate all model inputs.
- When validity of a model is questioned.

The development of hybrid modeling methods and modular model classification in the field of communicable diseases and especially for vaccination questions in HTA together with pointing out the main working tasks for parameterization close this main task of the IFEDH project.

VISUALISATION

The development of standardized visualization methods and representation of results – was integrated in the project to generate additional insight in the model structure, to help explaining the calculated results and also to have a powerful graphical tool for parameterization.

The workload was split into several tasks, whereby the analysis of state of the art methods in HTA was one of the first. Additionally, to think outside the box, adequate visualization methods from other areas were detected and analyzed and also a rating regarding potential use in modeling questions for HTA was done. This rating was directed by the visualization experts together with the modeling group and classical HTA expertise.

The other tasks were mainly based on the data basis for visualization. A guideline for visualization of generally evaluable data basis for vaccination program evaluation and the explorative analysis and visual inquiry of data quality was developed.

For visualization of results, also methods from other application areas were introduced into HTA. Only to name one, parallel sets were integrated as interactive visualization approach for analyzing markov models, as common methods to visualize Markov states over time (e.g., Markov-cycle trees or state probability graphs) do not scale well to many cycles and have limitations concerning the perception of proportions. To overcome these limits new visualization methods of Markov models and their results were investigated, inspired by the “Parallel Coordinates”. An interactive technique called Parallel Sets has been developed for visualizing multidimensional categorical data. The visualization lays out axes in a parallel way where each axis represents one categorical dimension. Within each axis, boxes represent the categories which are scaled according to the respective frequencies. Applied to Markov Models, the categorical dimensions correspond to the various cycles. Joint probabilities of categories from adjacent axes are shown as parallelograms

connecting the respective categories. The parallelograms can be interpreted as the number of patients transiting from one state to another. Depending on the purpose, the color of the parallelograms indicates the categories of a chosen cycle or could refer to additional attributes of the patients like age or sex. (see Figure 5)

State probability and survival curves merely show specific aggregates of the data while classic Markov trace visualizations with for example bubble diagrams do not visualize data in a sense that would facilitate a detection of proportions and trends. Applying Parallel Sets to analyze Markov models provides an interactive visualization technique where changing the reference Markov cycle is as easy as highlighting particular dimensions, thus enabling the exploration of the progress of patient cohorts with certain characteristics through the model. Model development always requires thorough analysis of its structure, behavior and results.

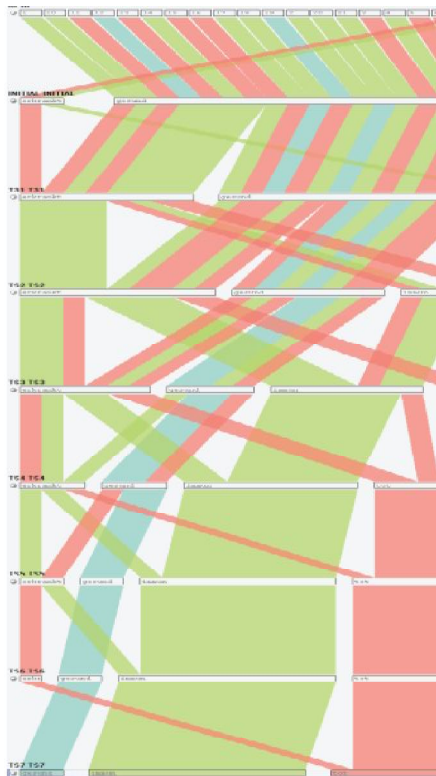


Figure 5: Parallel Sets enable an intuitive and efficient visualization technique for presentation purposes as well as exploratory analysis.

SUMMARY

The research project description presented in the paper shows how different expert domains can be joined to an overall approach in model based HTA of infectious diseases and vaccination strategy evaluation. As decision support has to become faster because of globalized interaction of people and therefore spreading of infectious diseases over borders and continents, a predefined work plan, parameter sources and modular reusable model parts are developed.

The theoretical work in the field of statistical methods, testing of adequate visualization methods and data quality assessment tools as well as a content management tool for the process guarantee that the boundaries regarding time span and quality can be observed.

The development of modern modeling methods and setting up an interdisciplinary process, dealing with the rising questions in HTA especially for infectious diseases and evaluation of vaccination strategy is realized. The description of the research project presented in the paper shows how different scientific domains can be joined to a joint overall approach in model based HTA. As decision support has to become faster, parameter sources and modular reusable model parts have to be developed in advance.

Using the developed knowledge and hierarchically structured framework an adequate system of decision support for model based HTA questions is realized. The closer look on Austrian data and healthcare system provides a fast implementation process for real world tasks.

ACKNOWLEDGEMENT

The Innovative Framework for Evidence based Decision making in Healthcare (IFEDH) research project is founded by the COIN – Cooperation and Networks program of the Austrian Research Promotion Agency (FFG), the national funding institution for applied research and development in Austria. The project is headed by dwh and lasts from October 2010 till September 2012.

REFERENCES

- Zauner G., Popper N., Breitenacker F. Evaluation of Different Modeling Techniques for Simulation of Epidemics. *Proceedings of the 7th Congress on Modeling and Simulation.(2010) ISBN: 978-80-01-04589-3*
- Miksch F., Popper N., Endel G., Schiller-Frühwirth I., Breitenacker F. *Herd Immunity as a Result in Dynamic Agent-based Epidemic Models. J. Wiley; Value in Health / Wiley-Blackwell, Vol. 13/ No. 7 (2010).*
- Zauner G., Popper N., Breitenacker F., Endel G.: Multi Agent Simulation Techniques For Dynamic Simulation of Social Interactions and Spread of Diseases with Different Serotypes. *Public Health Decision Making - Program and Abstracts. U. Siebert, E. Fenwick, J. Pliskin (Ed.); Target Conferences, (2010).*
- Taylor D., Kruzikas D., Pandya R., Iskandar R., Gilmore K., Weinstein M. Methods of Model Calibration: A Comparative Approach. *Value in Health - The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2007, vol. Volume 10, Number 3, p. A7.*

MODULAR MODELING AND HYBRID COMBINATION IN HEALTH TECHNOLOGY ASSESSMENT MODELS – EXAMPLES AND TECHNOLOGY

G. Zauner^(a), P. Einzinger^(b), F. Miksch^(c), I. Zechmeister^(d), G. Endel^(e), F. Breitenecker^(f)

(a) dwh simulation services, Vienna, Austria

(b) dwh simulation services, Vienna, Austria

(c) dwh simulation services, Vienna, Austria

(d) Ludwig Boltzmann Institute for Health Technology Assessment, Vienna, Austria

(e) Main Association of Austrian Social Security Institutions, Vienna, Austria

(f) Vienna University of Technology, Institute for Analysis and Scientific Computing, Austria

^(a) guenther.zauner@drahtwarenhandlung.at, ^(b) patrick.einzinger@drahtwarenhandlung.at,

^(c) florian.miksch@drahtwarenhandlung.at, ^(d) Ingrid.Zechmeister@hta.lbg.ac.at,

^(e) gottfried.endel@hvb.sozvers.at, ^(f) felix.breitenecker@tuwien.ac.at

ABSTRACT

Simulation of communicable diseases is an important field of interest in HTA. As requirements of stakeholders in the field of HTA are getting higher there is a need for complex models using different approaches.

A concept to deal with the increasing effort and complexity is to split up models into single parts, so called modules, which enables easier development, more flexibility and reusability.

In this work we concentrate on two different modular approaches for simulation of epidemics:

The agent based approach, dealing with individual behavior of persons. A model for Influenza with different environments, a model for Streptococcus-Pneumoniae with complex infection rules and another influenza-model with social interaction deal as examples.

The top-down approach starts with the epidemiological part where, epidemiological outcomes are calculated in a more general way, following by the demographical part. It ends with the economical part that calculates costs and evaluates economic aspects. A model for human papillomavirus vaccination (HPV) serves as an example.

Keywords: modular, modeling, agent based, HTA, health technology assessment, epidemics, communicable diseases, sexual transmitted disease, STI.

1. INTRODUCTION

Modelling and simulation of communicable diseases is an important field of interest in HTA. The aims range from understanding epidemic dynamics to predictions of an outbreak and to economical assessments.

There are several ways of modelling the spread of an epidemic. The traditional ways are statistical calculations or Markov models, mathematicians prefer differential equations or System Dynamics. These approaches are well known and examined but due to their structure they have certain limitations. Due to increasing computational power, other methods like agent based modelling or cellular automata became popular in recent years.

However, the closer such models are to reality the more complex they become. They get harder to develop, to handle and to control. For that matter we try to split such models into single parts, also called modules. These parts are easier to develop and to maintain because they are smaller and, if they are split wisely, only refer to a single aspect of the model. Another benefit is that modules might be replaced without touching the rest of the model which increases reusability and even makes it possible to connect different modelling methods.

In this work we are going to provide two different modular approaches that both provide detailed calculations and allow flexible usage: An agent based approach and a top-down approach..

2. THE AGENT BASED APPROACH

An agent based model is a model that consists of single parts that act individually and independently. For an epidemic model it is simple: The agents are the persons. We propose a general structure that consists of four modules (see also Fig. 1).

This structure was further developed and improved within IFEDH (Innovative Framework for Evidence-Based Decision-Making in Healthcare), a project funded by FFG (FFG project number: 827347).

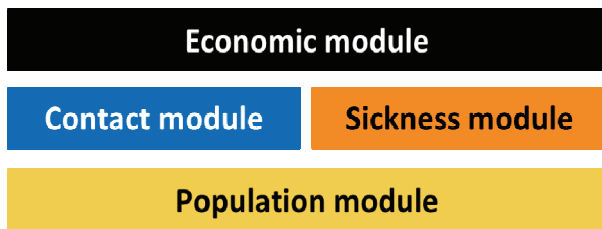


Figure 1. The proposed modules of an agent based epidemic model.

The population module is in charge to construct the persons in the beginning of the simulation and equip them with correctly distributed parameters (i.e. gender, age), monitor them during the simulation and assure their progression (i.e. that they get older).

The contact module creates an environment in which the persons can move and can have contacts and assigns collective or individual rules to the persons concerning moving around in that environment and meeting other persons there.

The sickness module assigns collective or individual rules what happens to a susceptible person when he meets an infectious person.

It assigns rules that control the progression of a sickness after an infection and it performs preventive interventions like vaccinations at certain points in time and on certain persons.

The economic module keeps track of all costs caused by preventive interventions or sicknesses and performs statistical analyses on the population and on the costs.

3. EXAMPLES OF THE AGENT BASED APPROACH

In the following part three examples for modular modeling of infectious diseases using agent based technology are listed. The first example uses also hybrid coupling of another bottom-up technique: cellular automata for modeling of spatial dynamics of the agents.

3.1. Combination of cellular automata and agent based modeling techniques

The first example is a model to simulate an influenza epidemic. The population module constructs a population according to Austria's demographics where people have a gender but they do not have a specific age, they are assigned to an age group. The contact module is more complex. The environment consists of households, schools, workplaces and neighbourhoods that compose as a collection of households. People are assigned to those places according to data from Statistics Austria. They visit these places on a daily basis. Contacts within households are generated stochastically between all household members. All other places are constructed like lattice gas cellular automata (LGCA) while people move around like particles according to LGCA-rules. Contacts are considered whenever two or more persons are together

in one cell. Fig. 2 shows a visualization of the contact module.

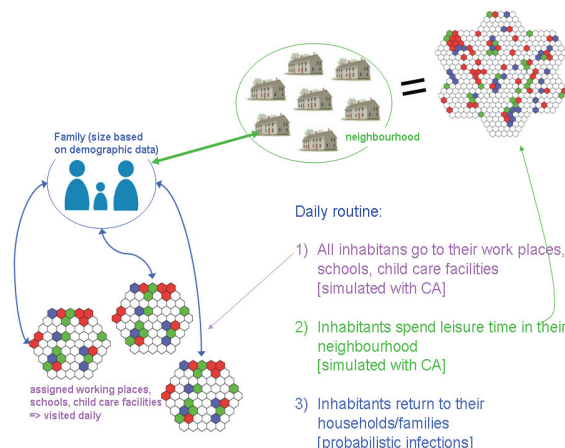


Figure 2. The contact module from the first agent based Influenza model (Emrich et al. 2008).

The sickness module has been modelled (according to available medical information) as follows: Individuals start out in the state susceptible in which they are vulnerable to infection. Upon infection individuals are in the state infected without symptoms in which they can pass on the infection but not yet be detected (e.g. cannot stay at home because they feel sick) eventually their status changes to infected and symptomatic, recovered (and immune) or dead. Alternatively individuals that are not symptomatic can become recovered (and immune) without going through the symptomatic state. These transitions are time-triggered through with a normal distribution around the set time-period. Additionally recovered individuals lose their immunity after a period of 75 days. (Emrich et al. (2008), Zauner et al. (2011)).

3.2. Streptococcus Pneumonia children vaccination model

The second example is a model for Streptococcus Pneumoniae. This bacteria can cause pneumonia, otitis media, meningitis, sepsis. The population module constructs persons according to Austria's demographics with gender and a specific age. Due to a simulation time of 20 years people get older, can die and babies (persons with age 0) are born. The contact module is relatively simple. It assumes a well-mixed population where people meet each other randomly on a weekly basis. The sickness module has to deal with several issues. There are more than 90 types of Streptococcus Pneumoniae pathogens, called serotypes. They are in a concurrent situation since a person can be infected by only one serotype at one time but never by two or more. The vaccination only works against some of these serotypes. The sickness module has to assure that the preconditions are preserved. It does not consider all serotypes in detail, instead it considers two serotypes: One stands for all serotypes that can be repelled by the vaccination and the other one for the rest. It provides these rules (see also Fig. 3):

Infected people recover after a while and can be infected again by any serotype.

Vaccinated persons can be infected only by the second serotype not included in the vaccine, simulating a hundred percent protection rate.

Infected people can either be blocked for or infected with the other serotype. If an infected person gets infected by another serotype he gets rid of the old one. The new serotype replaces the old serotype out of the person.

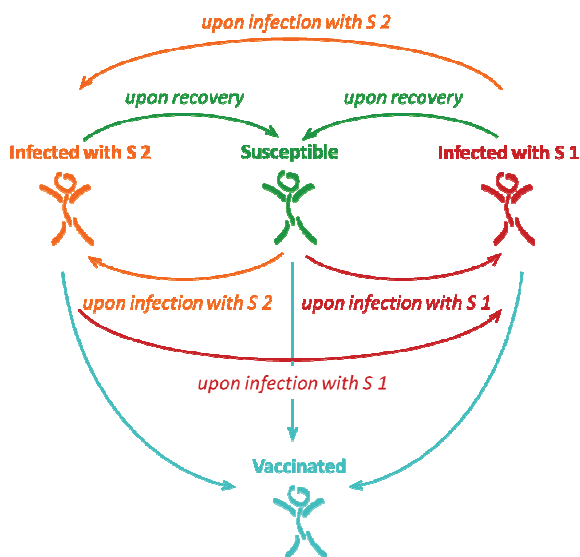


Figure 3. Sickness module of the agent based model for Streptococcus Pneumoniae (Zauner et al. 2010).

The economic module keeps track of all infections by the serotypes. Based on this data it estimates the number of sickness cases and furthermore costs due to treatment and vaccinations. It provides several outputs to analyse the spread of the pathogen as well as sicknesses as well as costs and cost-benefit analyses. (Zauner et al. (2010))

3.3. Modular real world influenza model

The third example is another model for Influenza. The population module is the same as the one used in the Streptococcus-Pneumoniae-model in the second example modelling realistic Austrian population (Brooks-Pollock et al.. (2010)). The main difference is that no aging of the population is implemented due to the fact that an Influenza season lasts only about 15 weeks and therefore no effects of changing demographic structure has to be implemented. The contact module is very detailed and complex. The environment consists of households, schools, workplaces and a leisure place. According to data provided by Statistics Austria people are assigned to these places. This means, households in the model are filled with persons, so that the number of inhabitants and their age distribution correspond with real world data.

The same conditions apply for schools that consider children and teachers classrooms. It assures that

children of the same age are put together in a classroom so that they have more contacts with each other than with children of other age. The number of people in workplaces is also distributed accordingly to the situation in Austria. This sets the size of groups that can have contacts at work.

People know where they live, study and work so they are mainly always surrounded by the same potential contact persons in the different places. The leisure place can be visited by everybody and does not need assignments.

Based on an individual daily schedule people move around between those places and can meet other present people. Parameters are adjusted in a way that the sum of contacts in a certain place corresponds with current social studies that evaluate daily contacts that are relevant for airborne diseases. Fig. 4 shows a visual representation of this module, split into different parts explaining the logic development process.

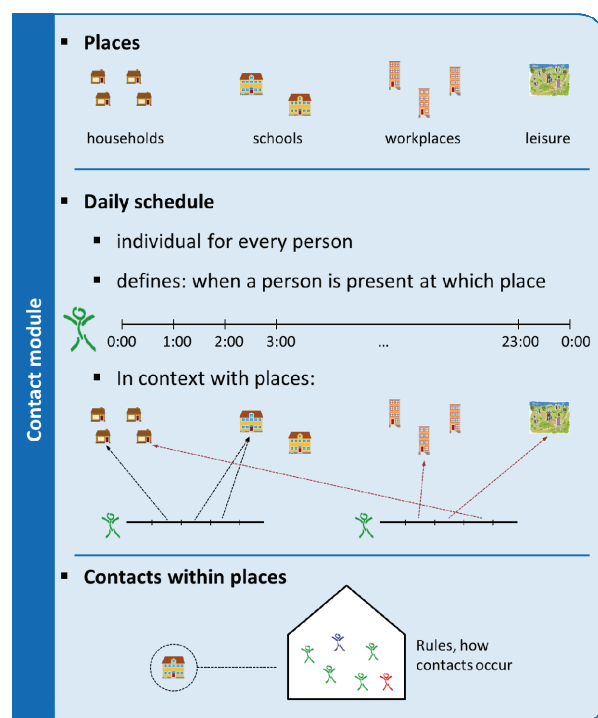


Figure 4. The contact module in the second agent based model for Influenza.

The sickness module is the same as used in the first Influenza model (see Fig. 5). The definition of interfaces between the different modules are defined in an interdisciplinary team, including specialists with health technology assessment (HTA) knowledge, modeler, and physicians. This guarantees the technical practicability as well as reusability and that the standardized HTA output measures can be calculated.

The economic module performs evaluations on the spread of Influenza and analyses in detail how many infections occur in the single places (Miksch et al. (2010)).

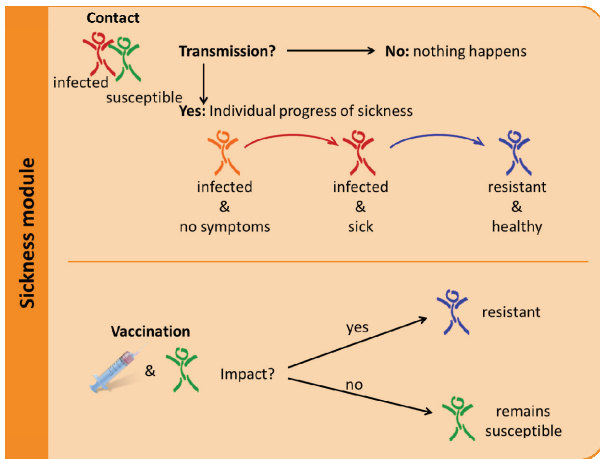


Figure 5. The sickness module of the second agent based model for Influenza.

Cost data are not parameterized in detail yet because of limited reliability of cost data for Influenza cases in Austria. Nevertheless the modular structure and the interfaces are implemented using object oriented programming techniques. The economic evaluation module calculates the costs per life years gained (€/LYG) without discounting the costs, but discounting the lost life years due to superinfections (pneumonia, etc.) caused by influenza Infection (module depiction see Fig. 6). The usage of costs without discounting is based on the fact that the costs occur at the actual time point whereas lost life years occur at a future time.

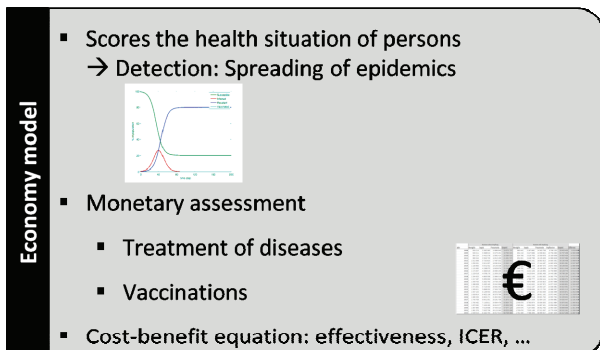


Figure 6. The economy module structure of the advanced agent based model for influenza

4. THE TOP-DOWN APPROACH

Top-down models for epidemics are a quite different approach than agent based models. They do not consider single persons representing a bottom-up approach; instead they calculate the fractions of the population that are in certain states representing the persons of interest and their behaviour using a top-down point of view. Common techniques especially in health technology assessment applications are Markov models – mainly Markov chains, based on quite simple decision trees for each time step not taking into account a longer history - and ordinary differential equations (ODEs) realized using System Dynamics simulation tools without knowledge of the equations.

In the following discussion an ODE approach is the basis for all assumptions and findings. As in case of agent based techniques the model in general has to provide a structure for simulation of a real world population and its demographic change over time, the disease propagation and the resulting costs/benefits from a health care service provider point of view.

One possible approach splits the problem into three modules:

The **disease module** calculates clinical and epidemic outcome parameters, including the number of carriers, sicknesses and deaths for a research population/standardized population with homogeneous cohorts.

The **demographic module** takes the results of the disease module and matches it on the demographics of a given population. This means, it calculates the outcome parameters for a specific population.

The **economic module** calculates the costs and performances for the clinical and outcome parameters from the demographic module using predefined average cost data and statistical distributions.

4.1. ODE modeling example: HPV vaccination

The included example simulates the spread of human papillomavirus (HPV) in the Austrian population and is used for modeling of vaccination effects (Zechmeister et al. 2009).

As the effect of decreased cervix cancer incidence and mortality is highest for women age 50 years and older, however vaccination is done in the age of 12 years, long simulation time span (and thereby unsecure events) have to be focussed. The long term effects achieved by vaccination against the most common known virus strains and the concurrence behaviour changes by vaccination against these strains is not known and can therefore not be found in literature to get secure parameterization. Calibration techniques and comparison of the qualitative behaviour of other diseases have to be realized to get a reliable model structure.

As the HPV modelling question of interest is quite complex in the following bullet list the main boundary conditions are listed in non-graded form:

HPV is the main known reason for cervix cancer and genital warts (causing death due to severe illness) and therefore has to be modelled quite detailed to ensure to fit the model quality requirements.

HPV is a sexually transmitted infection (STI), which means that other distribution pathways are the basis for transmission as in case of airborne diseases (in comparison with the other examples given in this paper)

In case of HPV another prevention method of severe diseases than vaccination is under discussion: screening.

Although screening is not implemented in the realized model in the actual form, interfaces for additional screening application have to be discussed.

The realization of the disease module was implemented in C using a set of coupled ordinary differential equations modeling the sexual contacts of

population groups for a research population. The contact data are based on international studies and expert opinion, beyond that these data sets are identified as one of the most unsecure information integrated in the model. Additionally the model and system behaviour is quite sensitive regarding changes in the contact parameterization. That is why univariate and multivariate sensitivity analysis for sexual contact data is planned.

The ODE structure in this model part is based on simple SIR structures.

Using the calculated distribution of infected women (men can be carrier of the virus strains, but do not get ill), age dependent probabilities are used to estimate the number of cervix cancer cases and genital warts due to HPV infection. Additionally statistical methods are used to assess the number of deaths and lost life years (based on assumed remaining lifetime for each modelled population group) for the research population.

The demographic module matches the gathered results to the real Austrian population. Errors based on the different behaviour of research (simulation) population and real life population are tolerated due to the complexity and the bounded reliability of the used parameter sets. The weakness of the model is mainly based on parameterization questions and other structural boundaries not the conversion of the simulated population to the real life population.

But anyhow the problematic behind this strategy has to be focussed analyzed in each additional model extension.

The economic module as top level of the model – being the part of highest interest for decision maker in health technology assessment and evidence based medicine – calculates the discounted costs and benefits for the vaccination strategy. As costs the vaccination program costs and the treatment are taken into account; additionally the lost life years due to death by cervix cancer are analyzed in discounted form for each strategy realized in the model and the comparing scenario – in this case the state of the art without organized vaccination program.

The benefits are prevented illness cases and therefore reduced morbidity and mortality, leading to a cost reduction for treatment of severe illnesses.

The interface for implementation of screening technologies can be realized in the disease module; the transmission part is not directly influenced and can therefore be used as a basic model block. The resulting virus carrier (men and women) can be understood as population group of main interest, which should be detected by the screening program.

The result of screening as a new layer in the module hierarchy could be placed in between disease module and demographic module.

If model refinement would be done, taking into account that people detected as carrier or with prestage cervix cancer are handled quite special because this population group knows their carrier stage, structural

changes in the ODE system have to be realized. In general STI models are not realized that detailed.

5. CONCLUSIONS

The examples show in different simulation situations that it is possible to split up a disease model into single modules.

The agent based approach dealing with different circumstances fits into the four-module-concept. Even more, it is possible to share some modules while other modules needed to be developed newly to fit the situation accurately. This clearly points out the strength of the modular approach.

The two Influenza models show (in the first example) the possibility of hybrid coupling used in a modular model structure, realizing spatial dynamics by implementing neighbourhood strategies and (second Influenza example) explain the usage of a very detailed contact model for Austrian settings which can be used for a long list of communicable airborne diseases.

The top-down approach also depicts benefits of modular models and reusability: The problem is split into three tasks that can be solved independently. It even allows using and adopting one module that has been developed in Britain instead of developing it in Austria again.

In general we can conclude that modular modeling helps to structure modeling process and that reusability of tested and validated modules makes projects more efficient and reliable.

6. OUTLOOK

The presented modular and hybrid modeling techniques for infectious disease strategy evaluation presented in this paper are applications focused on in the IFEDH framework research program in Austria. Besides the development of standardized reusable modeling techniques, the set up of interdisciplinary processes for simulation model based decision support in HTA questions is realized.

One result of cooperation and immediately one of the future working tasks is the optimization of the model output quality by the following tasks:

- Integration of HTA experts and physicians in the model definition process; discussion of reusability of pre-developed, validated modeling modules;
- Constant cooperation of data quality assessment specialists and data experts during the model set up and parameterization phase;

The integration of these points into the model realization process, together with the consequent refinement and validation of the already realized modules will lead to faster implementation of future issues in HTA and the results will obtain higher acceptance by the decision maker because of higher reliability.

The definition and implementation of additional module interfaces will be one of the most important future tasks. This work will be done iteratively to secure

that external expert knowledge from HTA and statistics can be taken into account.

REFERENCES

- Brooks-Pollock E, Cohen T, Murray M. (2010). The impact of realistic age structure in simple models of tuberculosis transmission. *PLoS One*. 2010;5(1):e8479.
- Emrich S., Breitenecker F., Zauner G., Popper N. (2008). Simulation of Influenza Epidemics with a Hybrid Model - Combining Cellular Automata and Agent Based Features. *Proceedings of the ITI 2008 30th Int. Conf. on Information Technology Interfaces*. p. 709 - 714. SRCE University Computing Centre University of Zagreb. ISBN: 978-953-7138-13-4.
- Miksch, F., Urach C., Popper N., Zauner G., Endel G., Schiller-Frühwirth I., Breitenecker F., (2011). New Insights on the Spread of Influenza Through Agent Based Modeling. *Value In Health*, 14 (2011), 7 *Value in Health - The Journal of the International Society for Pharmacoeconomics and Outcomes Research*. 14(7). J. Wiley. *Value in Health / Wiley-Blackwell*.
- Zauner G., Miksch G., Popper N., Endel G., Schiller-Frühwirth I., Breitenecker F., (2010). Long-Term Effects of Children Pneumococcus Vaccination: An Agent Based Approach. *Value in Health - The Journal of the International Society for Pharmacoeconomics and Outcomes Research*. 13(7). J. Wiley. *Value in Health / Wiley-Blackwell*. ISSN: 1098-3015. p. 383 - 384.
- Zauner G., Einzinger P., Miksch F., Popper N., Breitenecker F., (2011). Hybrid Coupling Techniques in Modular Infectious Disease Modeling. *Book of Abstracts ECCS'11 Vienna*. Löcker, ISBN: 978-3-85409-613-9.
- Zechmeister I, Blasio BF, Garnett G, Neilson AR, Siebert U. (2009). Cost-effectiveness analysis of human papillomavirus-vaccination programs to prevent cervical cancer in Austria. *Vaccine*. 27(37). Epub 2009 Jun 28. 5133-41.

VOLATILE ORGANIC COMPOUNDS IN EXHALED BREATH: REAL-TIME MEASUREMENTS, MODELING, AND BIO-MONITORING APPLICATIONS

Julian King^(a,c), Karl Unterkofler^(b,c), Susanne Teschl^(d), Anton Amann^(c), Gerald Teschl^(a)

^(a)Faculty of Mathematics, University of Vienna, Nordbergstr. 15, A-1090 Wien, Austria

^(b)Vorarlberg University of Applied Sciences, Hochschulstr. 1, A-6850 Dornbirn, Austria

^(c)Breath Research Institute, Austrian Academy of Sciences, Rathausplatz 4, A-6850 Dornbirn, Austria

^(d)University of Applied Sciences Technikum Wien, Höchstädtplatz 5, A-1200 Wien, Austria

^(a)Julian.King@univie.ac.at; Gerald.Teschl@univie.ac.at, ^(b)Karl.Unterkofler@fhv.at,

^(c)Anton.Amann@oeaw.ac.at, ^(d)Susanne.Teschl@technikum-wien.at

ABSTRACT

The present paper reviews some paradigmatic applications of modeling and simulation in the context of exhaled breath analysis, which has emerged as a promising tool for collecting *non-invasive* and *continuous* information on the metabolic and physiological state of an individual. In particular, it is illustrated how real-time breath profiles of volatile organic compounds (as obtained, e.g., by PTR-(TOF)-MS) can be used to extract *in vivo* estimates of endogenous quantities that are difficult to access by conventional analytical approaches based on blood, urine, or tissue. Typical examples include the determination of production and metabolism rates, as well as tracking the tissue accumulation of exogenously administered trace gases (e.g., volatile anesthetics).

Keywords: exhaled breath analysis, volatile organic compounds (VOCs), physiological modeling, pulmonary gas exchange

1. INTRODUCTION

Due to its broad scope and applicability, breath gas analysis holds great promise as a versatile framework for general bio-monitoring applications (Amann and Smith 2005).

Volatile organic compounds (VOCs) in exhaled breath represent a unique biochemical probe in the sense that they can provide both *non-invasive* and *continuous* information on the metabolic and physiological state of an individual. Apart from diagnostics and therapy control, this information might fruitfully be used for dynamic assessments of normal physiological function (e.g., during exercise stress tests, in an intra-operative setting, or in a sleep lab (King *et al.* 2009; Schubert *et al.* 2012; King *et al.* 2012a)), pharmacodynamics (drug testing (Beauchamp *et al.* 2010)), or for quantifying body burden in response to environmental exposure (e.g., in occupational health).

However, quantitative modeling approaches for exploiting this potential remain challenging due to the multifaceted impact of physiological parameters (such as cardiac output or breathing patterns) as well as due to

the sparse and often conflicting data regarding potential biochemical sources or sinks of VOCs in the human body. Within this framework, a primary goal of our investigations is to clarify the behavior of several prototypic breath trace gases by providing phenomenological reference data as well as mechanistic modeling. The present paper reviews some paradigmatic applications in this context, discussing models for the endogenous VOC isoprene (a tentative biomarker for oxidative stress and lipid metabolism disorders) and the exogenously administered volatile anesthetic agent sevoflurane.

2. EXPERIMENTAL METHODS

The range of measurement techniques employed for breath gas analytical investigations is extremely diverse and each method comes with its specific strengths and weaknesses. Here, the main focus is on direct mass spectrometry, particularly proton transfer reaction mass spectrometry (PTR-MS) (Lindinger *et al.* 1998). A major hallmark of PTR-MS is its *real-time* capability, allowing for concentration measurements of VOCs with a sampling frequency of more than 1 Hz (i.e., on a breath-by-breath resolution).

The possibility of generating high frequency data can be viewed as an essential requirement for relating short-term variations of breath VOC concentrations to quick metabolic and physiological changes (e.g., in blood or ventilatory flow). An experimental setup for the parallel recording of real-time PTR-MS trace gas profiles as well as hemodynamic and respiratory parameters is shown in Figure 1 (King *et al.* 2009; King *et al.* 2010b).

This instrumentation provides an excellent opportunity to model/simulate the blood-gas kinetics and systemic distribution of specific VOCs within the human body, and to validate such simulations by adequately designed experimental regimes, e.g., by measuring concentration levels during workload segments on an ergometer (see Figure 1) or by looking at the concentration changes during sleep (with different sleep-related events giving rise to different VOC breath patterns (King *et al.* 2012a)).

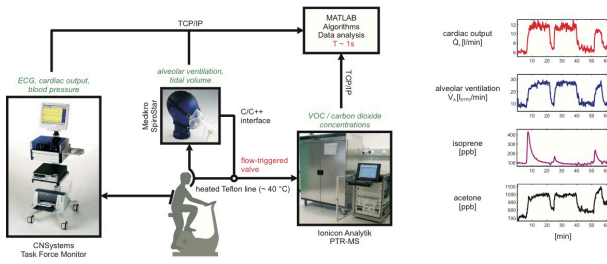


Figure 1: Experimental setup used for obtaining VOC profiles in conjunction with a number of physiological parameters. Items in *italic* correspond to measurable variables. Right panel: typical smoothed profiles of *end-tidal* isoprene and acetone concentrations in response to the following workload regime: 7 min resting, 15 min ergometer challenge at 75 Watts, 3 min resting, 15 min at 75 Watts, 12 min resting, 5 min at 75 Watts, 5 min resting. Data correspond to one single volunteer.

3. MODELING BREATH VOC PROFILES

In classical pulmonary inert gas elimination theory the relationship between the blood concentrations of VOCs and their respective concentrations in the gaseous phase is captured by the well-known equation due to Farhi (1967), *viz.*,

$$C_{\text{measured}} = C_A = \lambda_{\text{b,air}} C_a = \frac{C_v}{\lambda_{\text{b,air}} + \dot{V}_A / \dot{Q}_c} \quad (1)$$

where the subscripts A, a, and v denote alveolar, arterial, and mixed-venous concentrations, respectively, and λ is the substance-specific blood:air partition coefficient describing the diffusion equilibrium at the alveolar-capillary interface according to Henry's law. The influence of respiratory and hemodynamic variables is captured by the ratio between alveolar ventilation \dot{V}_A and cardiac output \dot{Q}_c .

3.1. Isoprene modeling

While the Farhi equation represents the starting point for any quantitative modeling approach in exhaled breath analysis, it turns out that this simple formulation is generally inadequate for explaining the observed dynamic behavior of breath VOCs in response to relatively quick physiological transitions.

Indeed, as can be seen in Figure 1 the profiles of isoprene and acetone considerably depart from the trend predicted by Equation (1). More specifically, at the onset of exercise the ventilation-perfusion ratio increases drastically and hence, other factors being equal, a more or less pronounced drop in the respective breath concentrations could be expected (such a drop can in fact be observed for some endogenous breath VOCs, e.g., butane and methane, see (King *et al.* 2010b) and Figure 6, respectively). As we shall illustrate in the following, in the case of isoprene this discrepancy can be explained by reference to a peripheral (extra-hepatic) production of the compound, which is realized by complementing the basic Farhi

model with two systemic compartments, affecting the time evolution of the mixed-venous concentration according to their fractional perfusion.

To this end, it is instructive to note that exercise bouts at constant workload interrupted by breaks of variable duration lead to markedly different heights of the characteristic breath isoprene concentration peak at the onset of pedaling, despite an almost identical behavior of cardiac output and alveolar ventilation throughout all workload segments (see Figure 1). This appears to exclude pulmonary gas exchange as a primary cause for the peak-shaped dynamics and points towards a substance-specific wash-out from an isoprene buffer tissue within the human body. In order to further investigate this hypothesis we additionally performed a series of *one-legged* ergometer experiments, showing that a switch of the working leg after a short break following an ergometer exercise segment results in an immediate recovery of the initial peak height, whereas continuing the exercise with the same leg leads to a wash-out effect similar to the two-legged case (King *et al.* 2010a).

From the phenomenological findings outlined above we concluded that a major part of isoprene variability during exercise phases can be ascribed to an increased fractional perfusion of the working skeletal muscles, eventually leading to higher isoprene levels in mixed venous blood at the onset of physical activity. This idea has subsequently been incorporated into a physiologically based compartment model (King *et al.* 2010a; Koc *et al.* 2011), describing the distribution of isoprene in various functional units of the organism, see Figure 2. By writing down the mass balance relations for each compartment such model structures can directly be expressed as a set of ordinary differential equations with smooth right hand side \mathbf{g} ,

$$\dot{\mathbf{x}}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (2)$$

where the state vector \mathbf{x} contains the molecular concentrations of the investigated trace gas within the tissue compartments introduced; \mathbf{u} stands for external inputs (blood/respiratory flows, temperature, etc.) that can be modified by experimentation; Θ lumps together an ensemble of constant, unknown model parameters (including, e.g., kinetic constants such as endogenous production or metabolization rates). The latter have to be estimated from the available process data (possibly together with some components of the initial conditions \mathbf{x}_0), which are linked to the system dynamics via the scalar *measurement equation*

$$y(t) = h(\mathbf{x}(t)) \quad (3)$$

Here, h is some smooth function defining the observable model output (i.e., the measurable breath concentration).

The proposed model can be calibrated based on the physiological data derived from the setup in Figure 1.

More specifically, the unknown (and volunteer-specific) model parameters and initial equilibrium concentrations are determined by standard least-squares procedures, implementing a *multiple shooting* routine in Matlab.

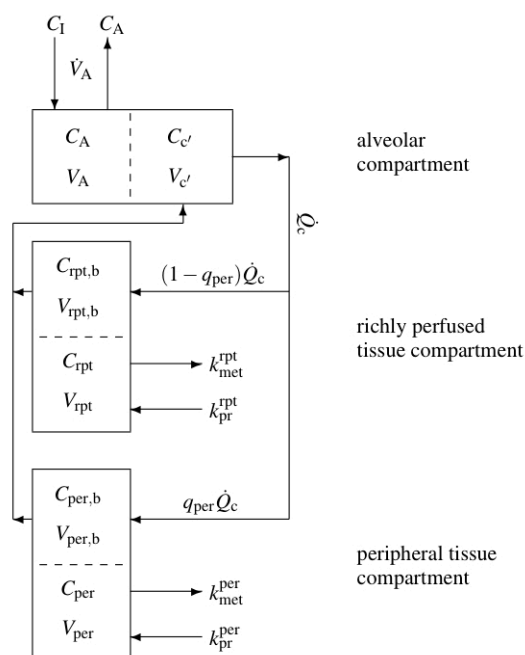


Figure 2: Schematic representation of the model structure used for describing the dynamics of isoprene concentrations in various parts of the human body. The latter is divided into three distinct functional units: alveolar/end-capillary compartment (gas exchange), richly perfused tissue (metabolism and production) and peripheral (muscle) tissue (storage, metabolism and production). Fractional blood flow to the periphery is denoted by q_{per} , while k_{pr} and k_{met} represent constant production and (linear) metabolism rates, respectively. Dashed boundaries indicate a diffusion equilibrium as determined by the respective blood:air and blood:tissue partition coefficients. Subscripts denote as follows: I-inhaled; A-alveolar; c'-end-capillary; rpt-richly perfused tissue; per-peripheral tissue; b-blood.

This iterative method can be seen as a generalization of the Gauss–Newton algorithm, designed to avoid divergence issues of the latter due to large residuals. For further details as well as convergence and stability properties we refer to (Peifer and Timmer 2007) and references therein. Figure 3 summarizes the results of these calculations.

The physiological mechanism revealed by this simulation is as follows: at rest the peripheral (muscle) compartment is characterized by high isoprene concentrations resulting from extra-hepatic production at a constant rate. However, due to the relatively small fractional perfusion of these tissues, the mixed venous concentration is mainly governed by the lower isoprene content of the venous blood returning from the richly perfused tissue group. At the start of exercise, the fractional perfusion q_{per} to the periphery drastically

increases. Consequently, isoprene is washed out from the peripheral compartment and the mixed venous concentration becomes dominated by peripheral venous return. The isoprene concentration peak visible in breath can thus be interpreted as a direct consequence of the associated rise in mixed venous blood concentrations.

Two important findings emerge from this modeling analysis. Firstly, it is suggested that the working skeletal muscles act as an active production site of isoprene in the human body. While such a hypothesis contrasts previous work postulating a purely hepatic origin of this compound, recent investigations observing a pronounced reduction of breath and blood isoprene levels in late stage Duchenne muscle dystrophy patients support this novel viewpoint (King *et al.* 2012b). In particular, the latter study is a good example how mechanistic modeling ultimately allows for designing theory-driven experiments.

Furthermore, the proposed model can serve as a quantitative tool for estimating relevant endogenous parameters governing isoprene dynamics (such as production or metabolism rates) on the basis of exhaled breath, hence allowing for the non-invasive extraction of patient-specific metabolic information. Indeed, the relative standard deviations of the estimated parameters as determined by *residual bootstrapping* usually range below 10%, thus demonstrating that the unknown model quantities are reasonably identifiable within the present experimental scenario (King *et al.* 2010a).

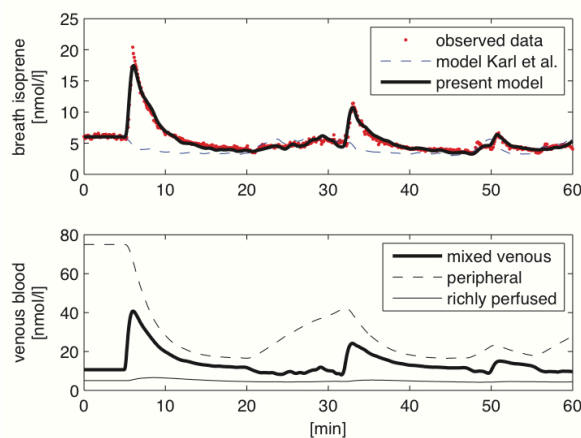


Figure 3: First panel: simulation of individual breath isoprene behavior during exercise conditions, compare Figure 1; second panel: predicted concentrations in mixed venous blood and venous blood returning from the peripheral (muscle) and richly perfused tissue groups.

3.2. Anesthetic monitoring

It has recently been demonstrated that PTR-MS allows for acquiring breath concentration profiles of inhalational (e.g., sevoflurane) as well as intravenous (e.g., propofol) anesthetic agents in real-time. By coupling these profiles with physiological modeling concepts and signal processing tools, non-invasive *on-line* schemes for continuously monitoring certain key quantities during anesthesia might be derived. Examples

include the estimation of agent concentrations in blood or the central nervous system (i.e., the target sites), or tracking major hemodynamic variables such as cardiac output.

For this purpose, a simple compartmental description for capturing the tissue accumulation of sevoflurane during inhalation anesthesia has been developed (King *et al.* 2011b). The model is largely analogous to the one in Figure 2, except that a brain compartment has been added and the peripheral tissue compartment has been replaced by adipose tissue, which represents a large storage volume for lipophilic compounds such as sevoflurane. Additionally, due to the fact that sevoflurane is not produced endogenously and only poorly metabolized, all production and metabolism rates have been set to zero. Using standard literature values for the involved physiological parameters, the resulting model predictions are in good agreement with published *in vivo* concentration profiles, see Figure 4.

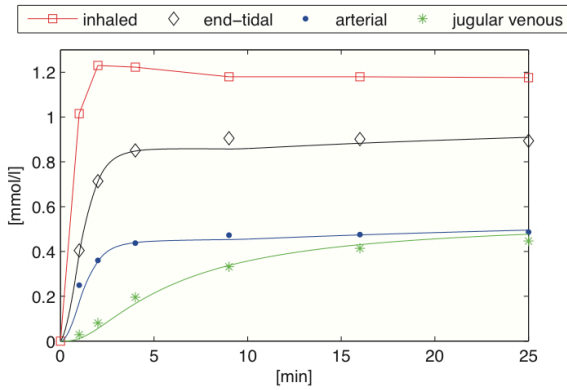


Figure 4: Simulation of sevoflurane profiles in inspired/end-tidal air, arterial blood, and jugular venous blood during administration of approx. 3% sevoflurane over 25 min. Discrete points reflect sample means associated with pooled data from 11 patients as measured by Nakamura *et al.* (1999).

Going beyond the direct problem of simulating the time evolution of breath sevoflurane levels given a number of physiological parameters, it is intriguing to ask whether one can also solve the related inverse problem, e.g., whether one can reconstruct (possibly in real-time) physiological quantities such as cardiac output or tissue (particularly brain) sevoflurane concentrations on the basis of sevoflurane dynamics observed in exhaled breath. Note, that this represents a question of considerable clinical relevance, as the former quantities are reflective of both the patient's hemodynamic status as well as his anesthetic/amnestic response, however, they are not easily accessible during the perioperative period. In the following we will present a proof-of-concept how the problem stated above can be approached by employing nonlinear filtering techniques. The limiting factor here is that due to the lack of adequate *in vivo* data the reconstruction is

based on *simulated* rather than measured breath sevoflurane data.

In a first step, by employing zero-order hold sampling and introducing Gaussian, white, zero-mean noise sequences \mathbf{w}_k , ϑ_k , and v_k the sevoflurane dynamics given by Equations (2) and (3) can be discretized as

$$\begin{aligned}\mathbf{x}_k &= G_{k-1}(\theta_{k-1})\mathbf{x}_{k-1} + \mathbf{f}_{k-1} + \mathbf{w}_{k-1} \\ \theta_k &= \theta_{k-1} + \vartheta_{k-1} \\ y_k &= H_k \mathbf{x}_k + d_k + v_k\end{aligned}$$

Here, the time-varying parameter θ reflects cardiac output and y_k is the measurable breath sevoflurane concentration. The aim is to sequentially reconstruct \mathbf{x} (i.e., the non-accessible compartmental concentrations) and θ from the stacked observations $Y_k = (y_k, \dots, y_1)$, which in this case, as mentioned before, have been generated by simulating the original deterministic system (2) and (3) and adding Gaussian noise with fixed variance).

Sequential state and parameter reconstruction can be achieved by virtue of a marginalized particle filtering scheme as introduced in (Schön *et al.* 2005). Briefly, starting from Gaussian prior densities $p(\mathbf{x}_0)$ and $p(\theta_0)$ encapsulating the available information on the initial compartment concentrations and the initial cardiac output, respectively, this framework allows for the recursive approximation of the posterior probability $p(\mathbf{x}_k, \Theta_k | Y_k)$, which embodies all accessible information on \mathbf{x}_k and the parameter sequence $\Theta_k = (\theta_k, \dots, \theta_0)$ up to time k . For the purpose of computing this density, it is instructive to note that, conditional on Θ_k , the discretized equations above represent a linear Gaussian system. Consequently, we may decompose

$$p(\mathbf{x}_k, \Theta_k | Y_k) = p(\mathbf{x}_k | \Theta_k, Y_k) p(\Theta_k | Y_k) \quad (4)$$

where the first factor can be determined analytically by means of the standard Kalman filter formulae. Simultaneously, particle filtering is employed for producing m samples (particle trajectories) $\Theta_{k,i} = (\theta_{k,i}, \dots, \theta_{0,i})$, $i = 1, \dots, m$ following $p(\Theta_k | Y_k)$, thus allowing for a (point mass) approximation of the second factor in Equation (4).

Combining these two mechanisms we end up with a bank of m Kalman filters running in parallel, each one associated with a single particle trajectory $\Theta_{k,i}$. When proceeding from time $k-1$ to time k , first the time update of each Kalman filter is performed. The resulting *a priori* estimates are subsequently used to update each particle trajectory according to the standard *sampling importance resampling* scheme fundamental to particle filtering (Gordon *et al.* 1993). In a last step, the measurement update for each Kalman filter is performed using the previously updated particle trajectory. Finally, the arithmetic mean of all *a posteriori* estimates (conditional means) derived from

the m separate Kalman filters constitutes a minimum mean square error estimate of the compartment concentrations \mathbf{x}_k , and similarly for θ_k . For further technical details the interested reader is referred to (King *et al.* 2011b).

The applicability of this filtering scheme within the anesthetic monitoring framework introduced above is demonstrated in Figure 5. Here, the mean and standard deviation of the prior distribution $p(\theta_0)$ reflecting the initial knowledge about the current cardiac output is set to 8 L/min and 3 L/min, respectively (corresponding to a poor initial guess with relatively large uncertainty). The resulting state and parameter estimates calculated as described above provide a reasonable reconstruction of their simulated counterparts. In particular, incidences such as abrupt drops in cardiac output are recognized within a delay time that is sufficiently small for enabling intra-operative interventions.

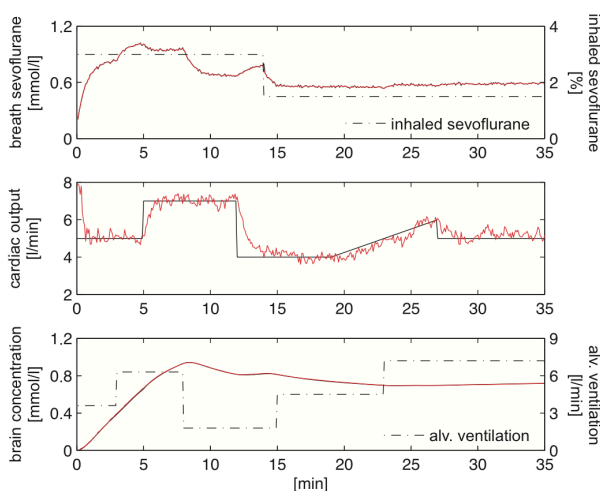


Figure 5: Simulated (solid black lines) and recovered (solid red lines) profiles of the breath sevoflurane concentration, cardiac output and brain concentration, using a marginalized particle filter with size $m = 300$. Dash-dotted lines represent the profiles of the inhaled sevoflurane concentration and alveolar ventilation used for the simulation of breath sevoflurane data.

4. OUTLOOK AND CONCLUSIONS

While the above approaches demonstrate that coupling the high-frequency information obtainable from breath gas analytical techniques with well-established tools from parameter identification and signal processing potentially allows for (continuous) estimation and monitoring of endogenous processes, we are well aware of the fact that several aspects have to be investigated more deeply before such methodologies can become clinically relevant (e.g., as a part of automated anesthesia delivery systems in the case of sevoflurane).

The primary building block for a successful implementation is the availability of reliable physiological models for the endogenous distribution of the volatile compounds under study. Consequently, additional experimental efforts, data gathering and

modeling attempts (accounting for potential substance-specific confounding factors, e.g., time-varying metabolization patterns, compartmental sequestration, etc.) are required in order to extend the validity of simple models such as the ones presented above over a sufficiently wide range of possible dynamics.

Another important contribution towards generalizing such modeling results is the increased availability of analytical techniques allowing for *parallel* real-time measurements of distinct substance classes within the same experimental regime. A prototypic method in this regard time-of-flight mass spectrometry coupled with chemical ionization, as realized for instance in PTR-TOF (Herbig *et al.* 2009). Briefly, such devices obviate the usual trade-off between sampling frequency and the number of trace gases that can be monitored simultaneously and offer very detailed spectral information in every sampling instant. Typical results are shown in Figure 6.

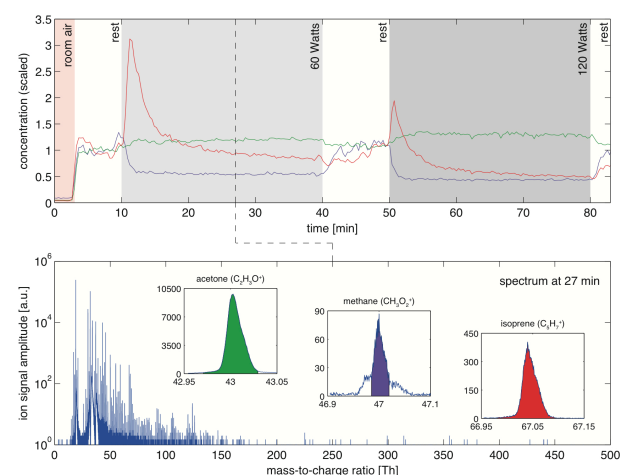


Figure 6: Profiles of breath acetone, methane, isoprene during a similar exercise scenario as in Figure 1. Data correspond to one single volunteer and are obtained by chemical ionization TOF mass spectrometry using O_2^+ primary ions. The second panel shows the TOF spectrum at 27 min. In the ideal (non-overlapping) case, each peak in the spectrum corresponds to one VOC, which may be identified *a priori* using pure gas standards. The concentration at a specific time instant is roughly proportional to the associated area under curve. Note that distinct VOCs can show a rather different behavior in response to the same physiological stimulus. While methane obeys the dynamics predicted by Equation (1), substance-specific factors like peripheral wash-out (isoprene) or airway gas exchange (acetone, cf. (King *et al.* 2012c; King *et al.* 2011a)) may lead to characteristic deviations from that trend.

As a concluding remark, the development of quantitative formulations relating breath concentrations of trace gases to their underlying systemic levels clearly lags behind the enormous analytical progress in exhaled breath analysis. In this sense, the approaches presented in this paper are also intended to further strengthen the

role of mathematical modeling and simulation as core techniques in breath gas analytical investigations.

ACKNOWLEDGMENTS

J.K., K.U., and G.T. gratefully acknowledge support from the Austrian Science Fund (FWF) under Grant No. Y330 and Grant No. P24736-B23. We appreciate the generous support of the government of Vorarlberg, Austria.

REFERENCES

- Amann, A., Smith, D., (eds.), 2005. *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*. Singapore: World Scientific.
- Beauchamp, J., Kirsch, F., Buettner, A., 2010. Real-time breath gas analysis for pharmacokinetics: monitoring exhaled breath by on-line proton-transfer-reaction mass spectrometry after ingestion of eucalyptol-containing capsules. *Journal of Breath Research* 4:026006.
- Farhi, L.E., 1967. Elimination of inert gas by the lung. *Respiration Physiology* 3:1-11.
- Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. A novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F* 140:107-113.
- Herbig, J., Müller, M., Schallhart, S., Titzmann, T., Graus, M., Hansel, A., 2009. On-line breath analysis with PTR-TOF. *Journal of Breath Research* 3:027004.
- King, J., Koc, H., Unterkofler, K., Mochalski, P., Kupferthaler, A., Teschl, G., Teschl, S., Hinterhuber, H., Amann, A., 2010a. Physiological modeling of isoprene dynamics in exhaled breath. *Journal of Theoretical Biology* 267:626-637.
- King, J., Kupferthaler, A., Frauscher, B., Hackner, H., Unterkofler, K., Teschl, G., Hinterhuber, H., Amann, A., Hoegl, B., 2012a. Measurement of endogenous acetone and isoprene in exhaled breath during sleep. *Physiological Measurement* 33:413-428.
- King, J., Kupferthaler, A., Unterkofler, K., Koc, H., Teschl, S., Teschl, G., Miekisch, W., Schubert, J., Hinterhuber, H., Amann, A., 2009. Isoprene and acetone concentration profiles during exercise on an ergometer. *Journal of Breath Research* 3:027006.
- King, J., Mochalski, P., Kupferthaler, A., Unterkofler, K., Koc, H., Filipiak, W., Teschl, S., Hinterhuber, H., Amann, A., 2010b. Dynamic profiles of volatile organic compounds in exhaled breath as determined by a coupled PTR-MS/GC-MS study. *Physiological Measurement* 31:1169-1184.
- King, J., Mochalski, P., Unterkofler, K., Teschl, G., Klieber, M., Stein, M., Amann, A., Baumann, M., 2012b. Breath isoprene: Muscle dystrophy patients support the concept of a pool of isoprene in the periphery of the human body. *Biochemical and Biophysical Research Communications* 423:526-530.
- King, J., Unterkofler, K., Teschl, G., Teschl, S., Mochalski, P., Koc, H., Hinterhuber, H., Amann, A., 2012c. A modeling-based evaluation of isothermal rebreathing for breath gas analysis of highly soluble volatile organic compounds. *Journal of Breath Research* 6:016005.
- King, J., Unterkofler, K., Teschl, G., Teschl, S., Koc, H., Hinterhuber, H., Amann, A., 2011a. A mathematical model for breath gas analysis of volatile organic compounds with special emphasis on acetone. *Journal of Mathematical Biology* 63:959-999.
- King, J., Unterkofler, K., Teschl, S., Amann, A., Teschl, G., 2011b. Breath gas analysis for estimating physiological processes using anesthetic monitoring as a prototypic example. Proceedings of the *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1001-1004, August 30 - September 3, 2011, Boston, USA.
- Koc, H., King, J., Teschl, G., Unterkofler, K., Teschl, S., Mochalski, P., Hinterhuber, H., Amann, A., 2011. The role of mathematical modeling in VOC analysis using isoprene as a prototypic example. *Journal of Breath Research* 5:037102.
- Lindinger, W., Hansel, A., Jordan, A., 1998. Proton-transfer-reaction mass spectrometry (PTR-MS): on-line monitoring of volatile organic compounds at pptv levels. *Chemical Society Reviews* 27:347-354.
- Nakamura, M., Sanjo, Y., Ikeda, K., 1999. Predicted sevoflurane partial pressure in the brain with an uptake and distribution model comparison with the measured value in internal jugular vein blood. *Journal of Clinical Monitoring and Computing* 15:299-305.
- Peifer, M., Timmer, J., 2007. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. *IET Systems Biology* 1:78-88.
- Schön, T., Gustafsson, F., Nordlund, P.J., 2005. Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing* 53:2279-2289.
- Schubert, R., Schwoebel, H., Mau-Moeller, A., Behrens, M., Fuchs, P., Sklorz, M., Schubert, J., Bruhn, S., Miekisch, W., 2012. Metabolic monitoring and assessment of anaerobic threshold by means of breath biomarkers. *Metabolomics*, in print (DOI: 10.1007/s11306-012-0408-6).

INVESTIGATION OF THE EFFECT OF DRUGS ON SOLID TUMOURS WITHIN A SYSTEMS-BASED MATHEMATICAL MODELLING FRAMEWORK

Cong Liu ^(a), J. Krishnan ^(b), Xiao Yun Xu ^(c)

^(a) Department of Chemical Engineering, Imperial College London, South Kensington, London SW7 2AZ, UK

^(b) Department of Chemical Engineering, Imperial College London, South Kensington, London SW7 2AZ, UK

^(c) Department of Chemical Engineering, Imperial College London, South Kensington, London SW7 2AZ, UK

^(a)cong.liu06@imperial.ac.uk, ^(b)j.krishnan@imperial.ac.uk, ^(c)yun.xu@imperial.ac.uk

ABSTRACT

The study develops a skeletal modeling framework to systematically evaluate the effect of anticancer drugs on solid tumors, especially from the perspective of cellular signaling. The modeling framework incorporates interstitial drug transport, intracellular apoptosis signaling and the dynamics of tumor cell density, which are regarded as the essentially minimal elements. The study deliberately starts with coarse grained descriptions of the cellular signaling, which nevertheless are capable of correctly capturing the qualitative dynamics involved. A series of simulations have been performed to provide mechanistic and predictive insights into cellular response towards different forms of drug stimuli. It is found qualitatively different intracellular signaling models can give rise to similar tissue behavior. Within the context, validating the models must be performed with care by considering a variety of drug stimuli.

Keywords: drug transport and effect, intracellular signaling, systems modeling framework, solid tumors

1. INTRODUCTION

It is a highly challenging multi-scale problem to elucidate the effects of anticancer drugs on solid tumors, concerning tremendous complexities associated with transport and cellular response (Minchinton and Tannock 2006; Kreeger and Lauffenburger 2010). Several attempts have been made to mathematically model the anticancer effects (Liu, Krishnan, Stebbing and Xu 2011); however, they failed to offer a transparent systems level description of the sources of these complexities, in particular a dynamical systems basis for the description of the cellular signaling. Considerable amounts of research activity in unraveling cell signaling processes under the umbrella of systems biology are especially relevant in the current context.

This study takes the first steps towards developing a systems-based modeling framework, which is capable of capturing the information flow correctly and in particular, should be modular and transparent to allow for further systematic refinement and augmentation.

This study integrates basic descriptions of essentially minimal elements, namely interstitial drug transport, apoptosis signal transduction and tumor cell density dynamics, in tumor cord geometry (Figure 1). A particular focus is paid on how to appropriately incorporate the complexities of cellular response which is characterized by highly non-linear chemical signal transduction. The efficacy of drugs is examined under simple (persistent and square pulse) drug signal (Liu, Krishnan and Xu 2011) and physiologically realistic signals with exponential decays. The modeling framework intends to yield mechanistic predictions of drug effect and provide clear-cut insight into the interconnections among the constituent elements.

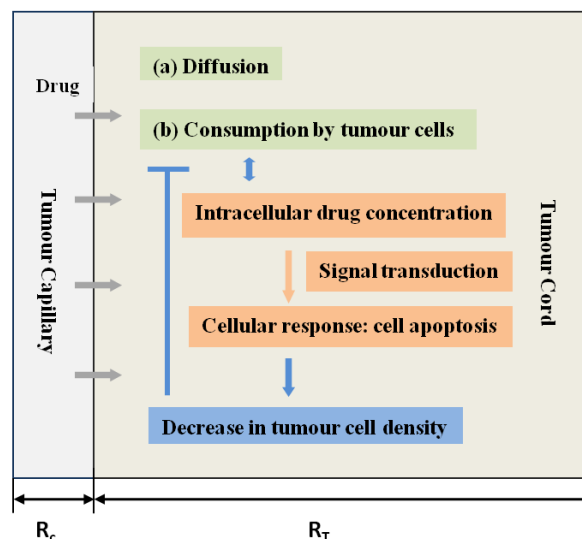


Figure 1: Diagrammatical representation of information flow incorporated in the modeling framework in the tumor cord geometry.

2. METHODS

Key descriptions of the modeling framework are presented as follows. The main assumptions of the model are:

1. Cells are initially alive and uniformly distributed in the tumor cord.
2. Cells are assumed to be stationary in the tumor cord.
3. Homogeneous conditions are assumed inside the tumor cord prior to injection of the drug.
4. Cellular variability and stochasticity are ignored.

Bearing in mind that the study focuses on developing a skeletal systems platform for gradual expansion and incorporation with further cellular/transport complexities, data fitting is not attempted at the current stage.

2.1. Interstitial drug transport

Interstitial drug transport is described by diffusion-reaction equation in 1D tumor cord geometry with both axial and radial symmetry (Eikenberry 2009). The model examines the spatio-temporal distribution of extracellular free, extracellular bound and intracellular drug concentration. Extracellular drugs can bind to, or unbind from proteins, such as albumin; in addition, the free form of extracellular drugs are taken up or released (pumped out) by tumor cells. Their dynamics are governed by:

Extracellular free drug:

$$\frac{\partial c_E}{\partial t} = D_E \nabla^2 c_E + c_t \left(\frac{V_2 c_I}{k_2 + c_I} - \frac{V_1 c_E}{k_1 + c_E} \right) - k_a c_E + k_d c_B \quad (1)$$

Extracellular bound drug:

$$\frac{\partial c_B}{\partial t} = D_B \nabla^2 c_B + k_a c_E - k_d c_B \quad (2)$$

Intracellular drug:

$$\frac{\partial c_I}{\partial t} = \frac{V_2 c_I}{k_2 + c_I} - \frac{V_1 c_E}{k_1 + c_E} \quad (3)$$

2.2. Intracellular apoptosis signaling

In the current study, the main intracellular process of interest is drug induced apoptosis. Concerning the facts that many mechanistic details in the intracellular response are yet to be fully understood, but it is urgently needed to incorporate the key features of cellular effects identified already at an appropriate level, the modeling strategy adopted is to deliberately start with simpler descriptions. Meanwhile the validity of the simplified models is checked by comparing to sample detailed models, especially for the types of drug stimuli encountered by cells.

Two key features must be reflected in any apoptosis model that there must be threshold effect present and apoptosis must be ‘switched’ on in an irreversible manner. Bistable switches which have self-contained threshold behavior and irreversibility have been adopted in modeling cellular signal transduction leading to apoptosis (Eissing, Conzelmann, Gilles, Allgower,

Bullinger and Scheurich 2004; Legewie, Bluthgen and Herzel 2006). However, there is still a debate in the literature of systems biology as to the qualitative nature of signal processing leading to cell killing. The irreversibility exhibited in cell apoptosis might arise from a simple irreversible reaction triggered in very special contexts (Albeck, Burke, Spencer, Lauffenburger and Sorger 2008; Gu, Zhang, Chen and Lei 2011), rather than is necessarily or even reasonably reflected as a steady state. Therefore, two types of apoptosis models are employed here: a bistable apoptosis model and a monostable switch sequentially connected to a downstream irreversible reaction effect.

Both models are essentially minimal but include a downstream intermediate element connecting input to response and a response element responsible for triggering apoptosis.

The study adopts a sample bistable model arising from a positive feedback, which is catalyzed by the upstream signal (the intracellular drug concentration c_I) (Ferrell Jr, Pomerening, Kim, Trunnell, Xiong, Huang and Machleder 2009).

$$\frac{dR}{dt} = \frac{V_f(1-R)}{K_{m1} + (1-R)} + (p + qc_I)k_{fb}R(1-R) - \frac{V_r R}{K_{m2} + R} \quad (4)$$

When the signal c_I crosses a threshold value for sufficient time, a sharp increase in the steady state of R is attained, even in the case of transient signal. The response in turn drives the activation of downstream species R_I from its inactivation form, which results in a different higher steady state of R_I . Cell death is triggered when R_I is across its threshold. When cellular apoptosis is reflected to the cell population density, it is modeled in a reversible manner.

$$\frac{dR_I}{dt} = k_f R(1 - R_I) - k_r R_I \quad (5)$$

The monostable switch in the second model is described by a Hill-type non-linearity, where the intracellular drug concentration acts as the upstream signal.

$$\frac{dR}{dt} = k \left(\frac{c_I^n}{k_h + c_I^n} - R \right) \quad (6)$$

The output drives the same activation of R_I as before. Only when the upstream monostable switch is kept ‘on’ for a substantial period of time, can R_I reach its threshold, which signifies cell death. In this case, such effect is reflected at the population level in an irreversible fashion.

2.3. Tumor cell density

The tumour cell density is governed by logistic equations. As it is described in a continuous form, the triggering of cellular apoptosis is represented as sharp decrease in growth rate at the population level. In this context (when $a < 0$), the only biologically relevant

steady state is the zero steady state, which indicates all the cells will perish eventually.

$$\frac{\partial c_t}{\partial t} = ac_t - bc_t^2 \quad (7)$$

It should be noted that computational implementation of this threshold effect is achieved in a reversible way in the bistable model, while in a unidirectional (irreversible) way in the monostable model.

3. RESULTS

Analysis of numerical results is presented by considering both variants of the intercellular signal transduction models and multiple types of drug stimuli. Here, doxorubicin is chosen as the anticancer drug for parameterization purpose.

3.1. Persistent drug input

Drugs are assumed to be injected directly at the entry to the tumor cord. Upon persistent drug infusion, a homogenous distribution of drug concentration is established throughout the tumor cord; and depending on the input intensity, drug concentration is either above or below the threshold. As a result, homogeneous cell response is observed that tumor cells are either killed or alive as shown in Figure 2. This trend is seen for both variants of intracellular apoptosis models.

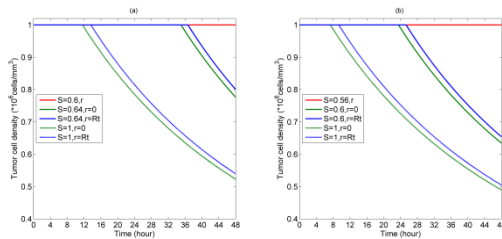


Figure 2: Temporal profiles of cell density at specific spatial locations for persistent infusion for different drug intensities. (a) Bistable switch case, (b) Monostable switch case. Note drug input S in the dimensionless form normalized by $0.001\mu\text{g}/\text{mm}^3$ for convenience.

3.2. Square pulse drug input

A square pulse drug input is assumed to be imposed at the inner capillary wall as well. Temporal profiles of intracellular drug concentration and tumor cell density are presented in Figure 3. A pulse injection generates a transient period of elevated intracellular drug concentration (Figure 3(a)), which is able to activate intracellular apoptosis pathway. This aspect is clearly revealed in Figure 3 (b), where the tumor cell density in the proximal starts to decline and would eventually becomes zero while it is kept as its initial state in the distal of the tumor cord. It suggests that a single pulse can kill tumor cells in a certain region only. Additional simulations show that an increase in the pulse strength/infusion time results in a broader cell death region, even the entire domain and the trend is in a strongly non-linear fashion for both bistable and

monostable apoptosis models (Liu, Krishnan et al. 2011).

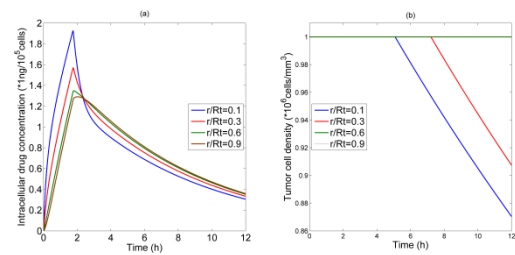


Figure 3: Time course of (a) intracellular drug concentration and (b) tumor cell density at various radial locations under pulse injection $S=4$, $T=1.75\text{h}$, for bistable apoptosis switch.

3.3. Exponentially decaying pulse drug input

The study also examines the case where drug is injected intravenously and the drug concentration input is determined by adopting a pharmacokinetic model (Robert 1982). Herein, the drug input is expressed in the exponentially decaying terms Presented in Figure 4 are temporal profiles of intracellular drug concentration and tumor cell density, which display dramatic similarities to those in Figure 3. All these demonstrate that our model can be used as a platform to systematically investigate the cellular response to various forms of drug injections, and even various forms of drugs (free drug versus drug delivery systems) in the future study.

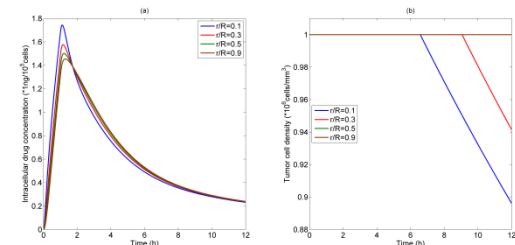


Figure 4 Time course of (a) intracellular drug concentration and (b) tumor cell density at various radial locations following a 1hour infusion of a dosage $300\text{mg}/\text{m}^2$, for bistable apoptosis switch.

3.4. Double square pulse drug inputs

The study also examines the effects of double pulse injections and finds that a second pulse injection does indeed expand the cell death zone as shown in Figure 5. It is also shown that the time interval between pulses is a key parameter in determining the efficacy of the following pulse injection. Doubling the time interval substantially reduces the width of cell death region, which is because the second pulse cannot take advantage of the residual effect from the first pulse. It should be noted that a second bolus has little effect on the time course of cell death in regions which stand already. Similar trends are observed in the case of monostable apoptosis switch.

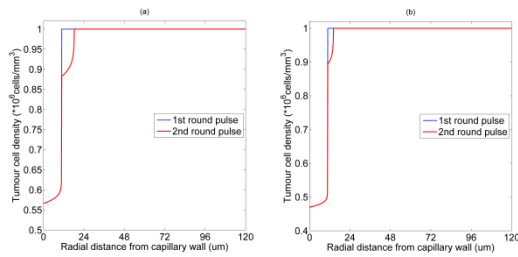


Figure 5: Spatial profiles of cell density for two rounds of pulse injection with the same pulse characteristics, $S=4$, $T=1.5h$ for bistable apoptosis switch, (a) tumor cell density at 36h with 24h interval between pulses, (b) tumor cell density at 60h with 48h interval between pulses.

4. DISCUSSION

The modeling framework incorporates the essential descriptions of interstitial drug transport, intracellular signal processing and tumor cell density dynamics, which are regarded as minimal elements connecting drug input to cell killing.

It should be mentioned that coarse grained descriptions of signaling pathways are adopted to qualitatively capture the input-output characteristics of the signal transduction dynamics while to retain the predictive advantages of mechanics models. The immediate focus of the present study is not on obtaining quantitative values to fit data rather than predictions regarding qualitative trends

Two variants of intracellular apoptosis models (one involving a bistable switch and the other a monostable irreversible switch) are examined. Interestingly, qualitatively different characteristics of the signal transduction notwithstanding, both models predict the essentially similar trends, suggesting that such difference plays a minor role in determining the responses to simple (and typical) stimuli encountered in realistic drug treatment. Difference might arise when more complex temporal signals are applied.

Alongside the simplified apoptosis signaling models, a sample detailed caspase activation model of type II apoptosis (intrinsic pathway) (Legewie, Blüthgen and Herzel 2006) has been examined here at the ODE level. From the perspective of input-output characteristics, the detailed bistable apoptosis model and the simplified one adopted in this study display qualitatively similar trends under simple (and typical) stimuli (discussed in Appendix A). Taken together, it can be speculated that additional biochemical details involved in the signaling network may play a more important role in the quantitative effect on the tumor dynamics, but not in the qualitative trends.

5. CONCLUSION

The study takes the first steps towards creating a skeletal systems-based modeling platform with

integration of essential descriptions of drug transport, cellular signaling and tumor tissue evolution. The modeling framework is modular and transparent to allow for systematic refinement and augmentation that could be incorporated in each module and between modules, and in tumor type- and anticancer agent-specific application. It is expected that the modeling strategy is capable of pinning down the roles that additional layers of complexities play in the system and yielding more comprehensive and predictive insight into the effects of anticancer drugs on solid tumors.

APPENDIX A

Examined first is the threshold effect that arises from the bistability exhibited in both the sample detailed and simplified bistable cases. The steady-state stimulus-response curves shown in Figure A1 depict that the systems remains the 'off' (lower) steady state even with increasing stimuli and are switched to the 'on' (higher) steady state until the stimuli reaches the threshold. It should be noted that the bistable system in the concrete models displays irreversibility; in other words, the system remains in the 'on' state permanently even after the removal of the signal. To capture the irreversibility in the abstract bistable apoptosis model, the upstream signal (c_I) is linearly factored in the form $p + qc_I$, where the basal level p lies in the bistable regime.

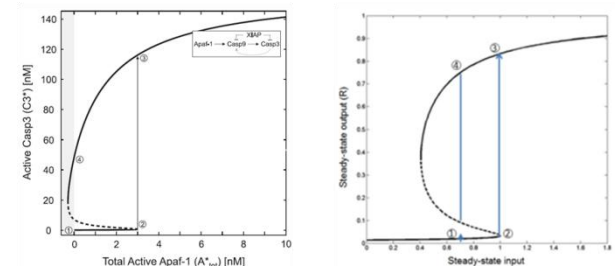


Figure A1: Steady-state stimulus-response curve of (a) caspase cascade in detailed apoptosis model adopted from (Legewie, Blüthgen et al. 2006), (b) the simplified bistable model in this study.

REFERENCES

- Albeck, J. G., J. M. Burke, et al. 2008. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol* 6: 2831-2852.
- Eikenberry, S. 2009. A tumor cord model for doxorubicin delivery and dose optimization in solid tumors. *Theor Biol Med Model* 6: 16.
- Eissing, T., H. Conzelmann, et al. 2004. Bistability analyses of a caspase activation model for receptor-induced apoptosis. *Journal of Biological Chemistry* 279: 36892-36897.
- Ferrell Jr, J. E., J. R. Pomerening, et al. 2009. Simple, realistic models of complex biological processes: Positive feedback and bistability in a cell fate switch and a cell cycle oscillator. *FEBS Letters* 583: 3999-4005.

- Gu, C., J. Zhang, et al. 2011. A trigger model of apoptosis induced by tumor necrosis factor signaling. *BMC Syst Biol* 5 Suppl 1: S13.
- Kreeger, P. K. and D. A. Lauffenburger. 2010. Cancer systems biology: a network modeling perspective. *Carcinogenesis* 31: 2-8.
- Legewie, S., N. Bluthgen, et al. 2006. Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *Plos Computational Biology* 2: 1061-1073.
- Legewie, S., N. Blühgen, et al. 2006. Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Comput Biol* 2: e120.
- Liu, C., J. Krishnan, et al. 2011. Use of mathematical models to understand anticancer drug delivery and its effect on solid tumors. *Pharmacogenomics* 12: 1337-1348.
- Liu, C., J. Krishnan, et al. 2011. A systems-based mathematical modelling framework for investigating the effect of drugs on solid tumours. *Theor Biol Med Model* 8: 45.
- Minchinton, A. I. and I. F. Tannock. 2006. Drug penetration in solid tumours. *Nat Rev Cancer* 6: 583-592.
- Robert, J. 1982. Pharmacokinetics of adriamycin in patients with breast cancer: Correlation between pharmacokinetic parameters and clinical short-term response. *European journal of cancer & clinical oncology* 18: 739-745.

STRUCTURAL CORRELATION METHOD FOR PRACTICAL ESTIMATION OF PATIENT SPECIFIC PARAMETERS IN HEART RATE REGULATION

Johnny T. Ottesen^(a) and Mette S. Olufsen^(b)

^(a) Department of Science, Systems, and Models, Roskilde University,
Universitetsvej 1, 4000 Roskilde, Denmark

^(b) Department of Mathematics, North Carolina State University,
Campus Box 8205, Raleigh, NC 27502, USA

^(a) Johnny@ruc.dk, ^(b) msolufsen@ncsu.edu

ABSTRACT

Numerous mathematical models have been proposed for prediction of baroreflex regulation of heart rate. Most models have been designed to provide qualitative predictions of the phenomena, though some recent models have been developed to predict observed data. In this study we show how sensitivity and correlation analysis can be used for model reduction and for obtaining a set of identifiable parameters that can be estimated reliably given a model and an associated set of data. We show that the model developed by Bugenhagen et al. to predict heart rate dynamics in the Dahl SS rat can be simplified significantly, without loss of its ability to predict measured data.

Keywords: Parameter estimation; Inverse problems; Model reduction; Subset selection; Simulation and modeling; Non-linear heart rate model; Patient specific modelling.

1. INTRODUCTION

Most models (including the one analysed here) have been developed with the aim of estimating dynamics of the system studied. Often models are developed in steps with the aim of including all known properties, rather than with the aim of obtaining the simplest self-contained model. The former is essential for gaining understanding of how various mechanical or physical properties impact the system dynamics, but may not be practical if the objective is to study how model parameters change within and between groups of subjects. For the latter, a simpler self-contained model may be better. This type of model often has fewer states and parameters. Estimation of reliable model parameters requires that sufficient data is available to identify all model parameters. Typically that is not the case, since experimental data often is sparse since they may be difficult and/or expensive to obtain. Consequently, as discussed in previous studies (Pope et al. 2009, Olufsen and Ottesen 2012) only a subset of parameters may be identifiable. In this study we show how sensitivity analysis and correlation analysis can be used for identifying model redundancies, which in turn can be used for model

reduction. Furthermore, (as in previous studies) we show how reliable parameters can be estimated in a reduced model.

Based on (Houk et al. 1966, Scrivivasan-Nudelman 1972, Hasan 1983, Alfrey 1987, Ottesen 1997, Olufsen et al. 2006, Ottesen and Olufsen 2011) Bugenhagen et al. [2010] presented a nonlinear differential equations model developed to predict baroreflex regulation of heart rate as a function of blood pressure for Dahl SS rats. This model contains complex nonlinear dynamics and a large number of parameters. Data for this model is considered sparse since only one output quantity is measured (heart rate), though it is sampled at a high frequency. The model is complex since it contains nonlinear dynamics as well as several time scales including fast inter-beat dynamics and slow dynamics associated with baroreflex regulation. It contains more than 30 parameters characterizing all known properties of the system. The model is developed from first principles describing the underlying mechanisms, with model parameters representing physiological quantities including arterial wall deformation, deformation of the baroreceptor nerve-endings, firing of afferent neurons, prediction of neurotransmitter dynamics (acetylcholine and noradrenaline) and the impact on heart rate. Although, the model was able to predict measured data no attempts were made to simplify the model or to analyse the reliability of the estimated parameters.

In this study we show how to simplify the heart rate model developed by Bugenhagen et al. [2010] by reducing the number of adjustable parameters. Then, for the reduced model, we demonstrate how to identify a subset of its parameters and estimate these so the model is able to predict measured heart rate.

2. MATHEMATICAL MODEL

In this section we first discuss concepts needed for prediction of sensitivities and correlations; second we apply these methods to analyse the heart rate model developed by Bugenhagen et al. [2010].

Sensitivity, correlation analysis, and parameter estimation

To predict sensitivities and pairwise correlations, we assume that the model can be formulated as a system of nonlinear differential equations of the form

$$\frac{dx}{dt} = f(t, x; \theta),$$

where $f: \mathbb{R}^{1+n+q} \rightarrow \mathbb{R}^n$, $t \in \mathbb{R}$ denotes time, $x \in \mathbb{R}^n$ denotes the state vector, and $\theta \in \mathbb{R}^q$ denotes the parameter vector. Associated with the states we assume an output vector $y \in \mathbb{R}^m$ corresponding to the available data (heart rate). We assume that this output can be computed algebraically as a function of the time t , the states x , and the model parameters θ , i.e.

$$y = g(t, x; \theta),$$

where $g: \mathbb{R}^{1+n+q} \rightarrow \mathbb{R}^m$. By construction the model output y is associated with data Y sampled at times t_i , where the sampling rate may vary between output components.

Sensitivities predict how much the model output changes with a change in the parameters. Classically (Frank 1978), the sensitivity matrix S of order $n \times l \times q$ (where l is the sampling cardinality) is defined by

$$S = \frac{\partial y}{\partial \theta}.$$

If model parameters vary significantly in magnitude, it may be advantageous to use relative sensitivities defined by

$$\tilde{S} = \frac{\partial y \theta}{\partial \theta y}, \quad y \neq 0.$$

Rank of the sensitivities (either classical or relative) can be computed as

$$S_R = \|S\|_2.$$

Pairwise parameter correlations can be predicted from the sensitivity matrix using the structured analysis discussed in (Olufsen and Ottesen 2012). As a point of departure, this method uses the model Hessian (a positive definite symmetric matrix sometimes denoted the Fisher information matrix, (Cintron-Arias et al. 2009)), which for problems with constant variance σ^2 , can be defined by $\mathcal{H} = \sigma^{-2} S^T S$ (Yue et al. 2006). Using \mathcal{H} , the correlation matrix c can be computed from the covariance matrix $C = \mathcal{H}^{-1}$ as

$$c_{i,j} = \frac{C_{i,j}}{C_{i,i}C_{j,j}}.$$

Notice that the matrix C only exists if the determinant of \mathcal{H} is non-vanishing. The matrix c is symmetric with elements $|c_{i,j}| \leq 1$. Parameter pairs (i, j) are considered correlated, if $|c_{i,j}| \geq \gamma$ for some value of γ . We denote such pairs as practically correlated parameters.

In this study we used the following structured approach to identify a set of sensitive and uncorrelated parameters. Assuming \mathcal{H} is non-singular,

1. Compute the correlation matrix c and identify all correlated parameter pairs, i.e., identify parameter pairs for which $|c_{i,j}| \geq \gamma$.
2. Sort correlated parameters according to their sensitivity. List all parameters ordered from the least to the most sensitive.
3. Remove the least sensitive correlated parameter from θ and recompute c for the reduced parameter set (this set can easily be done by deleting the corresponding column of S). The parameters removed from θ should be kept fixed at their *a priori* value.
4. Continue from 1 until $|c_{i,j}| < \gamma$ for all (i, j) .

For some models, the Hessian is singular. This follows if two or more parameters are conditionally identifiable (perfectly correlated) giving rise to redundancy. For this case, it is often possible to reduce the model eliminating either a parameter or an equation. Moreover, some equations may be superfluous in the sense that they hardly influence the model output. Thus we may perform an additional model reduction by removing or simplifying such model equations. Model reduction as part of the identification of a set of sensitive and uncorrelated parameters is an iterative process alternating between numerical and analytical considerations.

Once a reduced model and a set of identifiable parameters have been identified, parameter estimation methods can be employed to estimate the identifiable parameters. Assuming that the error between the model output and data are normally distributed, parameters can be estimated via solution of the inverse problem. In this study, parameters are estimated using the Levenberg-Marquart gradient-based method, which estimates parameters that minimize the least squares error between computed and measured values of the model output (Kelley 1999).

Heart rate model

Heart rate is one of the most important quantities controlled by the body to maintain homeostasis. The control of heart rate is mainly achieved by the autonomic nervous system involving a number of subsystems that operate on several time and length scales. One of the major contributors to autonomic regulation of heart rate is the baroreflex system, which operates on a fast time-scale (seconds). Baroreflex control consists of three parts: an affector part, a control center, and an effector part (Ottesen 1997, Ottesen and Olufsen 2011). The firing rate of the afferent baroreceptor neurons is modulated by changes in the viscoelastic stretch of the nerve-endings terminating in the arterial wall of the aorta and carotid sinuses. It is assumed that the deformation of the nerve-endings is modulated relative to the deformation of the arterial wall, which is imposed via changes in arterial pressure. The afferent neurons terminate in the nucleus solitary tract (NTS) within the medulla. The effector part consists of sympathetic and parasympathetic outflows, generated in NTS. The parasympathetic outflow travel along the vagal nerve, whereas sympathetic outflow travel via a network of interconnected neurons.

The main neurotransmitters involved with modulating heart rate regulation are acetylcholine, which is released by the vagal nerves, and noradrenaline released from the postganglionic sympathetic nerves. The model components are summarized in Figure 1.

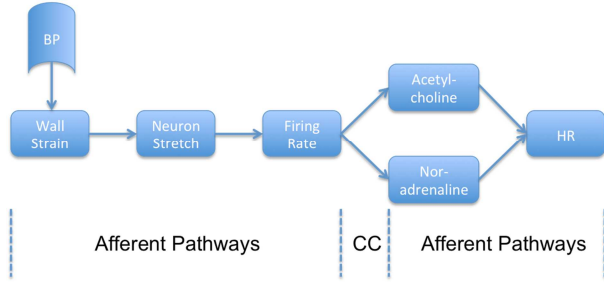


Figure. 1. Model components. CC denotes the control center, the nucleus solitary tract, which integrates all sensory inputs.

It is commonly assumed that the wall strain can be defined as

$$\varepsilon_w = \frac{R - R_0}{R_0},$$

which can be rewritten as

$$A = \pi R_0^2 (\varepsilon_w + 1)^2.$$

The approach suggested by Srinivasen and Nudelman [1972] is adopted assuming that stress is proportional to pressure. The latter is reasonable for a thin-walled tube (Valdez-Jasso, 2011). Bugenhagen et al. [2010] assumed a first order transient toward such static relation predicting the change in area A as

$$\frac{dA}{dt} = \frac{1}{B_w} \left(p - \frac{\sqrt{A/\pi} - R_0}{c_w} \right),$$

where $A_0 = \pi R_0^2$ and $A = \pi R^2$. In contrast to Bugenhagen [2011] these equations may be combined in a single differential equation for that relates wall stretch and pressure

$$\frac{d\varepsilon_w}{dt} = \frac{1}{2\pi R_0^2 B_w} \frac{1}{\varepsilon_w + 1} \left(p - \frac{R_0 \varepsilon_w}{c_w} \right).$$

Note $p(t_i)$ is obtained from the pressure data.

Using the idea originally put forward by Houk et al. [1966] and Hasan [1983] and later used by Alfrey [1987] inspiring Bugenhagen et al. [2010] the stretch of the baroreceptor nerve endings can be predicted assuming that the cells embedded in the wall respond viscoelastically to the vessel strain ε_w . The deformation of the baroreceptor nerve endings is predicted using Fung's (Fung, 1993), mechanical model with three Voigt bodies (see Figure 2) in series, given by

$$\begin{aligned} \frac{d\varepsilon_1}{dt} &= \frac{K_n(\varepsilon_w - \varepsilon_1) - K_1(\varepsilon_1 - \varepsilon_2)}{B_1} - \frac{d\varepsilon_2}{dt} \\ \frac{d\varepsilon_2}{dt} &= \frac{K_1(\varepsilon_1 - \varepsilon_2) - K_2(\varepsilon_2 - \varepsilon_3)}{B_1 + B_2} + B_1 \frac{d\varepsilon_1}{dt} + B_2 \frac{d\varepsilon_3}{dt} \\ \frac{d\varepsilon_3}{dt} &= \frac{K_2(\varepsilon_2 - \varepsilon_3) - K_3\varepsilon_3 + B_2 \frac{d\varepsilon_2}{dt}}{B_2 + B_3}. \end{aligned}$$

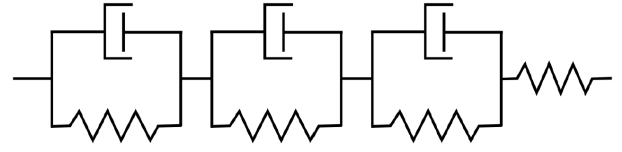


Figure 2: Three Voigt bodies in series. K_i denotes the spring constants and B_i the dashpots characterizing the viscoelastic cell and wall components. Finally, the spring in series with the dashpots is denoted by K_n .

For clarity these equations are rewritten in explicit form as

$$\frac{d\varepsilon_1}{dt} = -(a + \alpha_1 + \alpha_2 + \alpha_3)\varepsilon_1 + (a - b)\varepsilon_2 + (b - c)\varepsilon_3 + (\alpha_1 + \alpha_2 + \alpha_3)\varepsilon_w$$

$$\frac{d\varepsilon_2}{dt} = -(\alpha_2 + \alpha_3)\varepsilon_1 - b\varepsilon_2 + (b - c)\varepsilon_3 + (\alpha_2 + \alpha_3)\varepsilon_w$$

$$\frac{d\varepsilon_3}{dt} = -\alpha_3\varepsilon_1 - c\varepsilon_3 + \alpha_3\varepsilon_w$$

where

$$\begin{aligned} a &= \frac{K_1}{B_1}, & b &= \frac{K_2}{B_2}, & c &= \frac{K_3}{B_3} \\ \alpha_1 &= \frac{K_n}{B_1}, & \alpha_2 &= \frac{K_n}{B_2}, & \alpha_3 &= \frac{K_n}{B_3}. \end{aligned}$$

Notice that the rewritten equation involves six parameters, whereas the original equations in Bugenhagen [2011] involved 7 parameters. By combining two algebraic equations stated in (Bugenhagen et al. 2010) the firing rate model can be written as

$$f = M(\varepsilon_w - \varepsilon_1).$$

The afferent firing rate f is integrated in the NTS, where the sympathetic T_S and the parasympathetic T_P outflows are generated. We emphasize that the representation and interpretation of the rest of the model deviate slightly from what Bugenhagen [2011] did. Assuming saturation, these have been described using Hill functions as (Olufsen et al. 2006, Ottesen and Olufsen 2011)

$$\begin{aligned} T_S &= T_{SM} - (T_{SM} - T_{Sm}) \frac{f^\eta}{f^\eta + f_S^\eta} \\ T_P &= T_{Pm} + (T_{PM} - T_{Pm}) \frac{f^\xi}{f^\xi + f_P^\xi}, \end{aligned}$$

where subscript m refers to the minimum and subscript M refers to the maximum outflows, whereas η and ξ are constants predicting the steepness of the sigmoid. The next step involves prediction of the concentration of neurotransmitters acetylcholine C_A and C_N noradrenaline, which can be obtained from

$$\begin{aligned} \frac{dC_A}{dt} &= -\frac{C_A}{\tau_A} + q_P T_P \\ \frac{dC_N}{dt} &= -\frac{C_N}{\tau_N} + q_S T_S, \end{aligned}$$

where q_i and τ_i are time scales for build-up and decay of the neurotransmitter concentrations. Changes in the concentration of neurotransmitters impact the ionic stimulation of the heart. Acetylcholine binds to muscarinic receptors and noradrenaline binds to β -receptors controlling ion-channels. At least two channels pathways are affected in response to acetylcholine a slow sodium channel pathway and a fast potassium channel pathway, while noradrenaline mainly is associated with slower potassium and calcium channel pathways (Pyetan et al 2003, DiFrancesco 2006, Lyashkov et al 2009, Vinogradova TM and Lakatta 2009). In general about 75% of available acetylcholine stimulates the fast channel pathways while about 25% stimulates the slower channel pathways. For Noradrenaline, the full amount contributes to stimulating slower channel pathways. Common for all neurotransmitters stimulating the system is that the effect saturate at high concentrations. The actual complexity of ion-channels taking place in building up an action potential and herby regulating the heart rate is huge. Thus we simplify this complex mechanism by assuming a quasi-steady state of the occupied muscarinic and β -receptors and lumping all subsequent pathways into three hypothetical substances. For details see appendix A. Consequently, for noradrenaline we have

$$\frac{dC_{NS}}{dt} = \frac{1}{\tau_{NS}} \left(\frac{C_N^2}{C_N^2 + K_N^2} - C_{NS} \right)$$

while for acetylcholine we get

$$\begin{aligned} \frac{dC_{AF}}{dt} &= \frac{1}{\tau_{AF}} \left(\mu \frac{C_A^2}{C_A^2 + K_A^2} - C_{AF} \right) \\ \frac{dC_{AS}}{dt} &= \frac{1}{\tau_{AS}} \left((1 - \mu) \frac{C_{AS}^2}{C_{AS}^2 + K_A^2} - C_{AS} \right), \end{aligned}$$

where τ_i are time-scales, F and S denote a fast and slow response (i.e., $\tau_{NS}, \tau_{AS} \gg \tau_{AF}$), μ is a weighting parameter, and K_i denotes half the max response. Assuming the fast cholinergic process is almost instantaneous, the first equation can be replaced by

$$C_{AF} = \mu \frac{C_A^2}{C_A^2 + K_{AF}^2}.$$

Assuming that fast and slow responses are additive, the overall contribution gives

$$C_{AT} = C_{AS} + C_{AF}, \quad C_{NT} = C_{NS}.$$

Finally, we computed heart rate as

$$\begin{aligned} h &= h_0 + (h_M - h_0)C_{NT} - (h_0 - h_m)C_{AT} \\ &\quad - \frac{1}{h_0}(h_M - h_0)(h_0 - h_m)C_{NT}C_{AT}, \end{aligned}$$

where h_0 is the intrinsic heart rate, h_M and h_m denotes the maximal and minimal heart rate weighting, and the neurotransmitters are defined by C_{NT} and C_{AT} . In terms of the general theory outlined, $n = 8$ denotes the number of differential equations in the model. These are $x = (\varepsilon_w, \varepsilon_1, \varepsilon_2, \varepsilon_3, C_N, C_A, C_{NS}, C_{AS})$, where $m = 1$ denotes

the model output (heart rate), given by $y = h$. The model presented above has 30 parameters

$$\begin{aligned} \theta &= (R_0, c_w, B_w, a, b, c, \alpha_1, \alpha_2, \alpha_3, M, \\ &\eta, \xi, T_{Sm}, T_{SM}, T_{Pm}, T_{PM}, f_S, f_P, \tau_N, \tau_A, q_S, q_P, \\ &\tau_{NS}, \tau_{AS}, K_N, K_A, \mu, h_0, h_m, h_M). \end{aligned}$$

Model reduction and analysis

Sensitivity analysis (as defined by Frank [1978]) reveals that the sensitivity matrix $S = \partial h / \partial \theta$ is singular indicating that the model contains parameters that are perfectly correlated. Analysis of the equations reveals two correlations. First, the equation for ε_w can be simplified as

$$\frac{d\varepsilon_w}{dt} = \frac{p - K_{w1}\varepsilon_w}{K_{w2}(\varepsilon_w + 1)}$$

where $K_{w1} = R_0/c_w$ and $K_{w2} = 2\pi R_0^2 B_w$, i.e. c_w and B_w are conditionally identifiable with respect to R_0 . Second, substituting the expression for f into the expressions for T_S and T_P shows that M is redundant. Thus the gain M may be incorporated into f_S and f_P , i.e. they are conditionally identifiable with respect to M . Hence this equation reduces to

$$f = \varepsilon_w - \varepsilon_1.$$

With these simplifications the model can be formulated using the following 28 parameters:

$$\begin{aligned} \theta &= (K_{w1}, K_{w2}, a, b, c, \alpha_1, \alpha_2, \alpha_3, \\ &\eta, \xi, T_{Sm}, T_{SM}, T_{Pm}, T_{PM}, f_S, f_P, \tau_N, \tau_A, q_S, q_P, \\ &\tau_{NS}, \tau_{AS}, K_N, K_A, \mu, h_0, h_m, h_M). \end{aligned}$$

We emphasize that the resulting reduced model has a non-singular Hessian in contrast to the former model.

3. RESULTS

For the reduced model, ranked sensitivities (see Figure 3) were calculated.

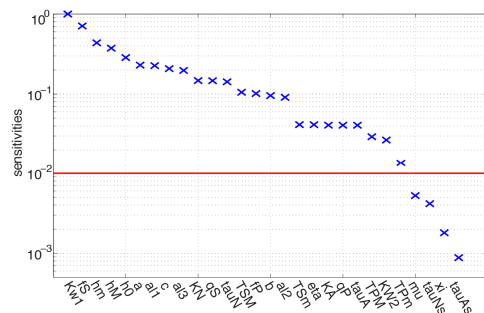


Figure 3: Sensitivity ranking, parameters below the horizontal line are considered insensitive.

Parameters $(\xi, \tau_{NS}, \tau_{AS}, \mu)$ with a sensitivity-norm lower than 0.01 were considered insensitive, and parameters (c, α_3) representing long time-scales compared to available data were kept at *a priori* values. Correlations among the remaining parameters were identified, leaving the following 17 parameters identifiable:

$$\begin{aligned} \theta &= (K_{w1}, K_{w2}, a, b, \alpha_1, \alpha_2, f_S, f_P, \\ &T_{Sm}, T_{SM}, T_{Pm}, T_{PM}, \tau_N, \tau_A, q_A, K_N, h_m). \end{aligned}$$

Table 1 compares parameter values from (Bugenhagen et al. 2010) with those used in the simplified model, and Figure 4 shows measured and optimized blood pressure and heart rate values.

Parameter	Bugenhagen	Simplified model
f_S	0.125*	0.123
f_P	0.138*	0.133
η	15.2	15.2
ξ	23.3	23.3
T_{Sm}	0.487	0.635
T_{SM}	4.20	3.62
T_{Pm}	0.910	1.21
T_{PM}	2.31	2.10
K_{w1}	267 [#]	180
K_{w2}	2.56 [#]	1.89
a	1.5 [#]	1.86
b	0.375 [#]	0.364
c	0.0156[#]	0.0156
α_1	1 [#]	0.710
α_2	0.1 [#]	0.0856
α_3	0.0149[#]	0.0149
τ_N	9.1	4.78
τ_A	0.2	0.137
q_N	0.11	0.110
q_A	5	5.93
K_N	1.12	0.628
K_A	0.65	0.650
τ_{Ns}	2.1	2.10
τ_{As}	2.5	2.50
γ	0.75	0.750
h_0	347	347
h_M	648	1117
h_m	253	253

Table 1: Comparison of parameter values from Bugenhagen et al. [2010] (left) with those obtained by the simplified model (right). Estimated (by optimization) parameters are marked in bold. Values marked by * are predicted to make Bugenhagen's formulation for T_S, T_P match our formulation. Values marked by # are calculated to convert the equations in Bugenhagen to formulation used in the simplified model.

4. CONCLUSION

In this study we used sensitivity and structural correlation analysis to simplify an existing model for heart rate regulation developed by Bugenhagen et al. [2010]. We also showed how to identify a subset of parameters that can be estimated given a model and a given set of experimental data. Results showed that the model contains several parameters that are not identifiable given the heart-rate data. If the objective is to estimate some of these parameters additional data from the same rat is needed.

APPENDIX A

Acetylcholine binds to muscarinic receptors and noradrenaline binds to β_1 -receptors at the sinus node controlling the ion-channels of the cell membranes. The actual complexity of ion-channels, taking into account the building up of an action potential and subsequently

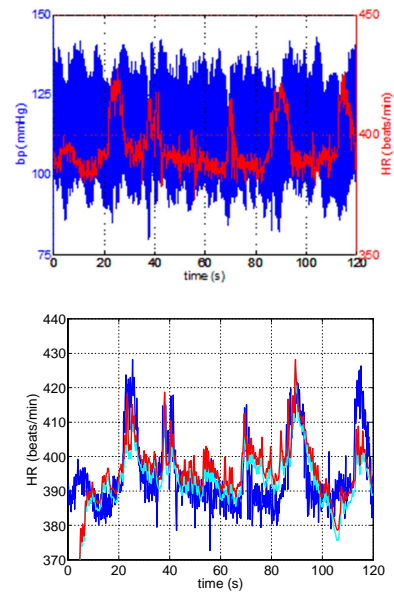


Figure 4. Top: Measured and predicted blood pressure (blue) and heart rate (red). Bottom: Measured (blue) and predicted (red and cyan) heart rate. The red trace shows results with the reduced model and cyan trace shows results from (Bugenhagen et al. 2010).

regulating heart rate is immense: At least six ion-channels are important for the generation of the action potential and hence for the heart rate, f-channels (sodium channels I_f), cholinergic calcium channels ($I_{K;ACh}$), non-transmitter dependent potassium channels (I_K), L-type (long lasting) calcium channels ($I_{Ca;L}$) T-type (transient) calcium channels ($I_{Ca;T}$), and calcium-sodium exchange channels (I_{NCX}) as illustrated in Figure 5. All of these ion-channels are regulated through (either inhibitory or stimulating) G-proteins: Acetylcholine binds to muscarinic receptors, down regulating cyclic AMP (cAMP) and phosphokinase A (PKA), which upregulates Na^+ (I_f) and down regulates K^+ ($I_{K;ACh}$) and Ca^{++} ($I_{Ca;L}$) whereas noradrenaline binds to β_1 -receptors upregulating cAMP and PKA (Pyetan et al 2003, DiFrancesco 2006, Lyashkov et al 2009, Vinogradova TM and Lakatta 2009).

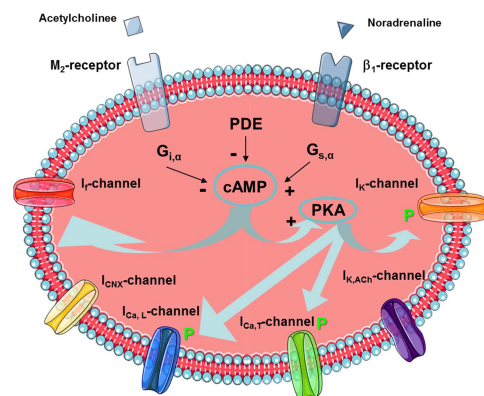
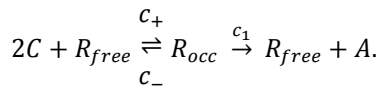


Figure 5. Ion-channels and pathways of greatest importance for the generation of the action potential and thus for the heart

rate: f-channel (sodium channels I_f), cholinergic calcium channel ($I_{K,ACH}$), non-transmitter dependent potassium channel (I_K), L-type (long lasting) calcium channel ($I_{Ca,L}$) T-type (transient) calcium channel ($I_{Ca,T}$), calcium-sodium exchange channel (I_{NCX}), inhibitory G-protein (G_i), stimulating G-protein (G_s), muscarinic receptor (M_2), β_1 -receptor, cyclic AMP (cAMP), phosphokinase A (PKA), and phosphodiesterase (PDE).

Common for all neurotransmitters stimulating such systems is that the effect saturate at high concentrations, which may be explained by simplifying the complex mechanism underlying it, e.g., by assuming a quasi-steady state for the occupied receptors. For simplicity we imagine that two molecules of the neural transmitter (C) binds to an receptor (R) forming an occupied receptor complex in a reversible reaction with rates c_+ and c_- and that the occupied receptor in addition may undergo another transition to the free state producing some substance (A) with rate c_1 ,



We also assume conservation of the receptor type, i.e.,

$$R_{free} + R_{occ} = R_0.$$

Hence the number of occupied receptor is govern by

$$\frac{dR_{occ}}{dt} = c_+ C^2 R_0 - (c_1 + c_- + c_+ C^2) R_{occ}.$$

In the quasi-steady state approximation this gives

$$R_{occ} = R_0 \frac{C^2}{C^2 + k_R^2},$$

where $k_R^2 = (c_1 + c_-)/c_+$. It is further assumed that the occupied receptors control the states of the ion-channels and thus the relevant intercellular pathways in building up the action potential. To omit this complexity we simply lump these path-ways into one or two substrates for each of the neurotransmitters, i.e., two for acetylcholine (C_{AF} or C_{AS}) and one for noradrenaline (C_{NS}), two channels are affected in response to acetylcholine a slow sodium channel and a fast potassium channel, while noradrenaline mainly is associated with slower channels. We assume that 75% of the occupied receptors stimulate the synthesis of the substance C_{AF} corresponding to the fast K^+ -channels while about 25% stimulates the synthesis of the substance C_{AS} corresponding to the slower Na^+ -channel. For noradrenaline, the full amount stimulates the synthesis of the substance C_{NT} corresponding to slow Na^+ and Ca^{++} -channels. Hence, for all three substances a fraction of the amount of occupied receptors serves as a production rate while the elimination are assumed proportional to the amount of the substance itself, thus the elimination rates are assumed constant (r),

$$\frac{dC_{xy}}{dt} = R_0 \left(\mu_{xy} \frac{C_x^2}{C_x^2 + k_R^2} - \frac{r}{R_0} C_{xy} \right),$$

where we use index $x = A, N$ (denoting acetylcholine and noradrenaline) and $y = F, S$ (denoting fast and slow

if necessary). Normalizing C_{xy} by r/R_0 and substituting $\tau_{xy} = 1/r$ gives

$$\frac{dC_{xy}}{dt} = \frac{1}{\tau_{xy}} \left(\mu_{xy} \frac{C_x^2}{C_x^2 + k_R^2} - C_{xy} \right).$$

ACKNOWLEDGMENTS

We would like to thank Scott Bugenhagen and Prof. Dan Beard, Department of Physiology, Medical College Wisconsin for providing data and computer code for their model analysis in this study. Olufsen and Ottesen were supported in part by NSF-DMS #1022688 and by Snedkermester Sophus Jacobsen and wife Astrid Jacobsen's foundation, and Olufsen was supported in part by NIH-NIGMS #1P50GM094503-01A0 subaward to NC-SU.

REFERENCES

- Alfrey KD, 1997. A Model of the Aortic Baroreceptor in Rat. MS Thesis, Rice University, Houston, TX.
- Bugenhagen SM, Cowley AW, Beard D, 2010. Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the Dahl SS rat. *Physiol Genomics*, 42:23-41, 2010.
- Cintron-Arias A, Banks HT, Capaldi A, Lloyd AL, 2009. A sensitivity matrix based methodology for inverse problem formulation. *J Inv Ill-Posed Problems* 17:545-564.
- DiFrancesco D, 2006. Funny channels in the control of cardiac rhythm and mode of action of selective blockers. *Pharmacological Research* 53: 399-406.
- Frank P, 1978. Introduction to sensitivity theory. Academic Press, New York, NY.
- Y.C. Fung, 1993. Biomechanics: Mechanical Properties of Living Tissues. Springer Verlag, New York, NY.
- Hasan Z, 1983. A model of spindle afferent response to muscle stretch. *J Neurophysiol*, 49: 989-1006.
- Houk J, Cornew RW, Stark L, 1966. A model of adaptation in amphibian spindle receptors. *J Theor Biol*, 12: 196-215.
- Kelley CT, 1999. Iterative Methods for Optimization. SIAM, Philadelphia, PA.
- Lyashkov AE, Vinogradova TM, Zahanich I, Li Y, Younes A, Nuss HB, Spurgeon HA, Maltsev VA, and Lakatta EG, 2009. Cholinergic receptor signaling modulates spontaneous firing of sinoatrial nodal cells via integrated effects on PKA-dependent Ca^{2+} cycling and IK_{ACH} . *Am J Physiol Heart Circ Physiol* 297: H949-H959.
- Olufsen MS, Tran HT, Ottesen JT, Lipsitz LA, Novak V, 2006. Modeling baroreflex regulation of heart rate during orthostatic stress. *Am J Physiol*, 291:R1355-R1368.
- Olufsen MS and Ottesen JT, 2012. A practical approach to parameter estimation applied to model predicting heart rate regulation. *J Math Biol*, in press.
- Ottesen JT, 1997. Nonlinearity of baroreceptor nerves. *Surv Math Ind* 7:187-201.

- Ottesen JT and Olufsen MS, 2011. Functionality of the baroreceptor nerves in heart rate regulation. *Comput Methods Prog Biomed*, 101:208–219.
- Pope SR, Ellwein LM, Zapata CL, Novak V, Kelley CT, Olufsen MS, 2009. Estimation and identification of parameters in a lumped cerebrovascular model. *Math Biosci Eng*, 6:93-115.
- Pyetan E, Toledo E, Zoran O, Akselrod S, 2003. Parametric description of cardiac vagal control *Autonomic Neuroscience: Basic and Clinical* 109:42– 52
- Srinivasen R, Nudelman HB, 1972. Modeling the carotid sinus baroreceptor. *Biophys J*, 12: 1171-1182.
- Yue H, Brown M, Knowles J, Wang H, Broomhead DS, Kell DB, 2006. Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NFkB signalling pathway. *Mol BioSyst* 2:640–649.
- Valdez-Jasso D, Bia D, Zocalo Y, Armentano RL, Haider MA, Olufsen MS, 2011. Linear and nonlinear viscoelastic modeling of aorta and carotid pressure-area. *Ann Biomed Eng*, 39: 1438-1456.
- Vinogradova TM and Lakatta EG, 2009. Regulation of basal and reserve cardiac pacemaker function by interactions of cAMP mediated PKA-dependent Ca²⁺ cycling with surface membrane channels. *J Mol Cell Cardiol*. 47(4): 456–474.

PRE-TENDER HOSPITAL SIMULATION USING NAIVE DIAGRAMS AS MODELS

Gabriel Wurzer^(a), Wolfgang E. Lorenz^(a), Manfred Pferzinger^(b)

^(a)Vienna University of Technology

^(b)UMIT, Hall in Tirol

^(a) {gabriel.wurzer|wolfgang.lorenz}@tuwien.ac.at, ^(b) manfred.pferzinger@umit.at

ABSTRACT

Hospital simulation has so-far concentrated on late phases of architectural planning, in which the design is already fully formulated and undergoes optimization. This paper moves into the exactly opposite direction - it tries to embed simulation into the earliest phase imaginable, which, interestingly, is well before any architectural planning occurs: The pre-tender work that is done by or on behalf of the client, using naïve diagrams based on interviews with the medical staff as models.

Keywords: early design process, hospital planning, diagrammatic tools, pre-tender simulation

1. INTRODUCTION

The preparation of a tender represents the earliest work done in hospital planning: Projected medical demands of a care region are mapped to either the establishment of a new hospital, or, more often, to an extension, adaptation or refurbishment of an existing one. Classically, the tool of choice for 'simulating' and 'calculating with' future demands has been spreadsheet software. Apart from statistical data, diagrammatic information (e.g. envisioned work processes and spatial arrangement of departments, obtained by interviewing the medical staff) is also generated. However, and in contrast to spreadsheets, there is a great lack of methods for interacting with the so-encoded data - especially when it comes to animation/simulation of (patient and material) flow.

Contribution: During the past half year, we have thus been working on **reverse-engineering diagrammatic representations into simulation models**, dealing with needs-orientated flow descriptions (patient flow), preliminary floor plans (schemata) and functional decompositions (bubble diagrams), in order to obtain a simulation targeted at the pre-tendering phase. To discuss the techniques used in this context and give a wider perspective on other possible applications is the main topic of this paper (see Section 4, "Contribution Details").

Establishing simulation in the earliest phase of planning enables clients to get a better overview of the project they are writing a tender for. Furthermore, diagrammatic data produced as result of the subsequent

competition phase can thus be evaluated and compared, which is advantageous for a wide audience:

- **Planners.** The spatially simulated flows can inform the design and enable a cross-check of requirements for the planned building. Generally, the flow given by the client represents a preliminary concept ("schema"), in which the spaces are not fully formulated. However, the processes depicted therein stay essentially the same, even after the architect has designed the final form ("floor plan"). By adaptation of the schema into the final floor plan, the architect can simulate the flow in his presented concept, while still relating to the client's vision.
- **Staff.** The hospital staff can be trained using the very same simulation, in which context also an acceptance check can take place. As the initial flow concept is typically generated by the same persons that will see the simulated final design, this enables to re-use the previous knowledge for evaluation.
- **Client.** Verification of requirements, as mentioned, can help in the decision process during the competition phase. Moreover, as the hospital goes into operation, realtime data concerning the actual patient flows can be exported from the Enterprise Resource Planning (ERP) system, in order to input and visualize these in the sense of a "management dashboard". To some extent, this also facilitates a verification of the building's operational concept.

2. RELATED WORK

Diagrams (White 1986; Seonwook and Miyoung 2012) are used as representations for the envisioned organization, i.e. space allocation (zones and their adjacency relations, circulation), hierarchy and responsibilities (organizational chart), temporal or causal order (processes, flow). The sources for this information vary; however, one common point is that planning is never done in isolation, but by drawing on pre-existing knowledge of several health professionals, such as physicians, nurses and therapists, administrative personnel and (most importantly) the patients

themselves. Tool support for employing diagrams as interview technique, used to let workshop participants record daily work routines in a game-like manner, was presented by (Wurzer, Fioravanti, Loffreda and Trento 2010). Simulation in the same setting, as “assessment tool”, has so far not been conducted. What exists in early-stage spatial simulation are approaches exploring space utilization (Tabak 2008; De Vries, Jessurun and Dijkstra 2002), verification of a functional program (Wurzer 2010), usage simulation and pathway visualization (Wurzer 2011) and early process simulation (Wurzer 2012).

3. BACKGROUND

As stated earlier, work that leads to the definition of early diagrams (most importantly: the architectural schema) is conducted collaboratively between health professionals and patients. The overall goal is to define requirements that enable effective workflows *according to customer needs* and expectancies. Because objectivity during such an elaboration process is of vital importance, work typically proceeds in workshops targeted at a *specific problem domain*.

Outputs are at first centered on process knowledge - i.e. daily work routines, clinical procedures and practices, such that one might produce business processes for outpatient, inpatient and day hospital treatment, according to the different areas of the hospital as defined in (DIN 13080 2003) as result. One may also specify procedures in case of special situations - e.g. mass accidents, which is especially important for trauma clinics. With these processes in mind, one may look at the intended **spatial configuration**, which is the topic we are focusing on in our work. In more detail, we employ three specific types of diagrams produced within pre-planning:

1. The **architectural schema** (see Figure 1a) as a hierarchy of “spaces within spaces”. A *space* is a bounded (usually rectangular) region that is denoted by a name. It contains a set of *functions*, i.e. names of activities that may be used in that space. Each function is a resource of limited *capacity*, of which the *usage* is computed during the subsequent simulation (see Simulation). It is crucial to note that the schema is fundamentally different from a floor plan showing the form of each space: It rather gives the approximate area and location of each space in two-dimensional arrangement, which is then detailed into the fully-formulated three-dimensional building layout during the competition.
2. **Flows** (see Figure 1b) depicted as arrows on top of the schema, giving a simple yet effective way to express paths of building users (Lohfert 2005). Usually, these are color-coded to distinguish different kinds of traffic (patients, staff, visitors).

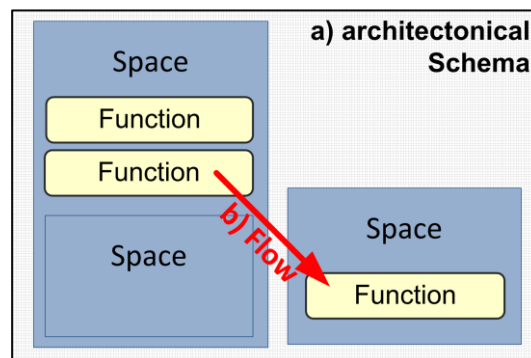


Figure 1: (a) Architectural schema as arrangement of nested spaces with functions. (b) Superimposed flows.

3. **Bubble diagrams** depicting the adjacency relations between functions in a purely abstract manner; a function is visualized as circle (“bubble”), the adjacency relation to another function as connecting line. Adjacency can either be “close” (denoted e.g. by a green line) or “separate” (denoted e.g. by a red line), “not given” or “not applicable” (no line or gray line). Additionally, it is common to depict this relation also by arranging close functions into clusters.

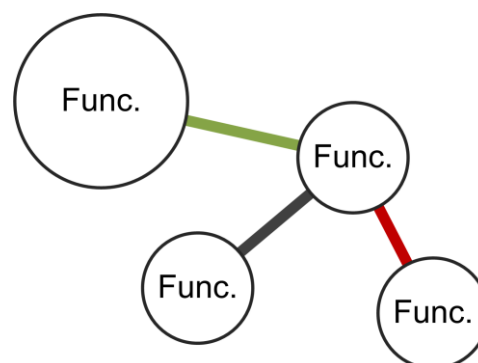


Figure 2: Bubbles diagram showing bubbles (representing functions). The size of a bubble corresponds to its assumed area consumption, the relationships between functions are given as colour-coded lines.

When using diagrams as source of information, one can exploit hidden semantics given by the topology of shapes present, in order to infer semantic relationships. More specifically (refer to Figure 3):

- *Nesting* of spaces (Figure 3a) can be inferred from topological inclusion.
- *Conceptual containment* (Figure 3b), in contrast to ‘real’ nesting, hints at a grouping of spaces using an abstract boundary, signified by a dashed line.
- *Using and being used by* (Figure 3c) according to the z-order - the lower element is using the higher one, transitively.

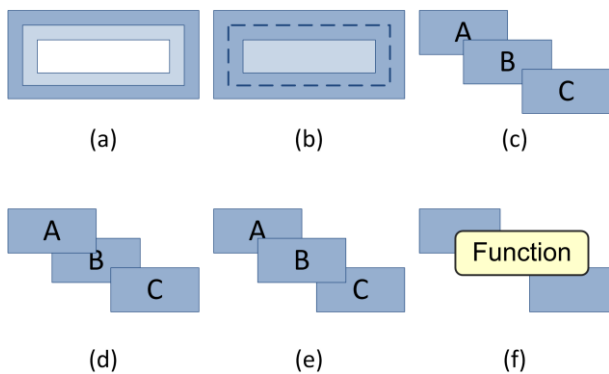


Figure 3: Hidden logic in Diagrams: (a) nesting (b) conceptual containment (c) A using B, B using C - transitivity (d) A and C being used by B (e) B being shared by A and C (f) function sharing

- *Subordinate space usage* (Figure 3d, space B is using A and C) an sharing (Figure 3e, B shared between A and B).
- *Sharing* can also be seen for functions (Figure 3f), however, the opposite case (function using two spaces) is not possible: Functions are always subordinate to spaces.

The actual use of these properties is presented in Section 4.1 “Rectangle analysis”.

4. CONTRIBUTION DETAILS

Our work first derives a spatial model from a schema (see “Rectangle analysis”, 4.1). The result is a semantically rich model of a hierarchy of spaces, together with the functions they contain. This is then used in the second phase, where patient flow is computed using an Agent-Based Simulation (ABM) that computes the progression of patients through the spaces, utilizing a sequence of functions as resources. Arrivals of patients are given as spreadsheets, based on either (hypothetical) arrival times and functional sequences, or, more commonly, using real data obtained through either the Hospital Information System (HIS) or the underlying Enterprise Resource Planning (ERP) System. As functions are the resources within the simulation, their usage is recorded and visualized in the third phase as Bubble diagram: For each function, a bubble with circle size according to capacity is drawn. The circle is colored according to functional utilization compared with capacity (under-utilized, over-utilized or well-utilized). Relationship lines between function are also depicted, based on the simulated flow between functions. Conceptually, this work is similar to e.g. (De Vries, Jessurun and Dijkstra 2002) - being a design and decision support tool based on employing simulation in the architectural workflow. Our focus on earliest stages (i.e. pre-tender phase) is, however, unprecedented in hospital planning, requiring careful thought over what data already exists, as will be shown in the following elaborations.

4.1. Rectangle analysis

Schema diagrams are drawn intuitively, as rectangles-within-rectangles and possible overlaps. In this section, a quick run-through of the analyzed features is made (also refer to Fig. 4):

- **Hierarchy buildup.** The pair-wise analysis of topological relationship between each two rectangles gives either “separated”, “completely included in/completely including”, “intersecting” or “touching”. In the first (trivial) case, the rectangles are completely separate and thus modeled as own entities. In the second case “included in/including”, the rectangles form a parent-child relationship. When “intersecting”, the rectangles also form a parent-child relationship; however, it is still unclear which is the parent and which the child rectangle. Z-order of the rectangles can be used as tie-breaking mechanism - the child element being the one that is arranged ‘on top’. Intersections are commonly used spaces that are used by more than one parent - e.g. a central operation theatre being used by many connecting areas. Accordingly, the hierarchy we build up offers, for each rectangle, the ability to have more than one parent - thus forming a semi-lattice rather than a tree structure (Alexander 1965). The last case, “touching” rectangles, is ignored - we assume each of the both rectangles are children of a common parent structure, and the touching relation being there to depict adjacency, not hierarchy.

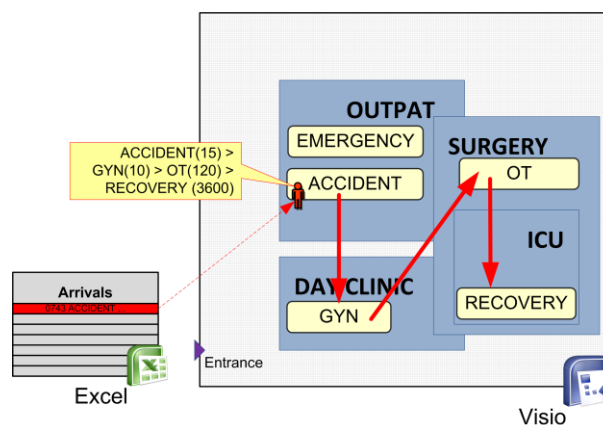


Figure 4: Depiction of the first part of the workflow used in the pre-tender simulation: A schema given as diagram is analyzed into a hierarchy of spaces on which a multi-agent simulation acts, using an underlying arrival list that states, for every agent, the list of functions to visit. The according patient flows are visualized as arrows. Furthermore, functional usage is recorded for later visualization.

- **Space and function attribution.** Up to this point, it is not clear whether the analyzed rectangles stand for spaces or functions, since both are nested rectangular elements. By use of manual attribution, we can infer that a given rectangle represents the latter structure - in which case it must also contain a capacity (integer). Additionally, in order to facilitate the later use of the function as resource, we also allocate a list currently occupying and queued agents.

4.2. Flow Simulation

Flow is given by arrivals in spreadsheet format, where each line represents one single patient (see Figure 4). Also on this line are 1.) the arrival time and 2.) the sequence of functions to visit, including stay time per function. Then, the simulation scheduler progresses in discrete time steps, in which:

1. **Arrivals** for the current instance are instantiated, producing agents with a fixed protocol of functions to visit. The first of these functions is immediately removed from the list, and the agent location is set to the space containing the function.
2. **Movement** of agents is simulated for all agents that are crossing to the next function. This happens with reference to the underlying schema, which is interpreted as circulative network.
3. **Occupation** of functions is simulated by employing the active and queue lists of the function under consideration; the total time spent per function is available in the agent itself, reflecting the amount of time taken for individual treatment. Simulation constructs such as passivation/activation can happen in this step, or, trivially, the agent waits for that time-span.
4. **Goal selection** happens for those agents that have finished using their function, by removing again the first element of the list of functions still to be visited. In case there is no such function, the agent is removed from the simulation.

The simulation runs as long as there are elements in the arrival list or there are active agents on the schema. In case the simulation ends, the visualization is prepared.

4.3. Visualization

In architectural workflow, Bubble diagrams are used to depict relations among functions; however, these are only intended relationships (i.e. intended by the architect, from close collaboration to dislocated). In our simulation, the results are used to build up such a diagram, based on the actual usage of functions and flows between them (also refer to Fig. 5a). For each function (depicted as a circle), we depict the usage as its radius, the interpretation of this usage compared with the capacity as color (under-utilized=blue, over-

utilized=red, well-utilized=green, not enough data=gray). The flows between functions hint at their level of cooperation. By thresholding, we can get a measure of closely cooperating function pairs, depicted as lines. However, in contrast to Bubble diagrams that have been specified manually, we cannot state which functions should not cooperate at all, and should thus be dislocated (e.g. for means of privacy, hygienic aspects, etc.) - our approach is inherently positivistic in this aspect. The resulting Bubble diagram can nevertheless be compared to a manually-made one, as validation.

As visualization, we can record the usage of functions over time (see Figure

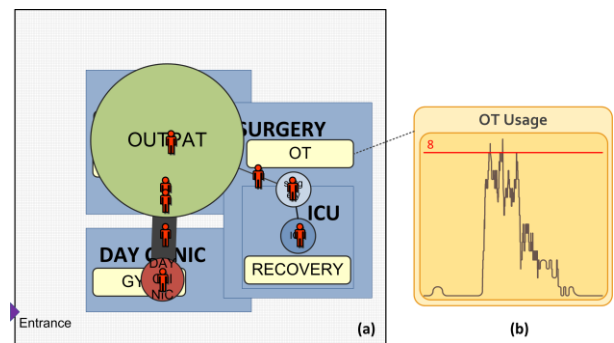


Figure 5: (a) Visualization of functional relationships obtained by using the flow simulation. Circles depict functions, links between them cooperation in the (simulated) work process. The size of each circle corresponds to the usage of the given function, the color is a comparison between usage and capacity (green: well utilized, red: over-utilized, blue: under-utilized, gray: too little data). (b) Depiction of usage of each function (red line indicates the capacity having been set by manual attribution).

5. DISCUSSION AND RESULTS

The stated approach was implemented and tested with anonymized patient trails (inpatients and outpatients, 1900 individuals) exported from a hospital's ERP system (SAP) and OT management system (60 individuals) for the course of one day which we see as being quite average - a Wednesday in non-vacation time. Technically, we received a sequence of time-stamps together with the service point (e.g. 10:05 Radiology, 11:39 Pediatrics) for each anonymized patient. From this, we could build a trail "Radiology > Pediatrics" and inferred durations from the time delta. The resulting duration is, arguably, not correct - but it is not wrong either:

- **Shortcoming 1.** The given timestamps are recorded either at the start or at the end of the service duration, which is bad. But, even worse than that, we can safely assume that both cases are present in the same dataset (e.g. OT times measured at start, radiology at the end of the service duration). Thus, the data is necessarily fuzzy, and must not be interpreted

quantitatively but rather qualitatively and with a grain of salt.

- **Shortcoming 2.** We cannot compute durations from a single time stamp (or, as in the example mentioned earlier, we do not know how long the patient was in the Pediatrics unit). To counter this problem, average service times were used where available.
- **Shortcoming 3.** The duration computed has not necessarily got to do anything with the real service time. For example, the time delta between 11:39 Pediatrics and 10:05 Radiology would be 1 hour 34 minutes, which is a rather long time for, say, an x-ray. Again using average service times (e.g. x-ray: 4 minutes), one can dispatch the patient to the next function and use the remaining time as waiting time (i.e. Radiology(4m)>Pediatrics(1h 30m)).

In contrast to the mentioned problems, we also saw a large benefit: We were able to transfer knowledge about a hospital to a new design, using arrivals and current trails as input for a new design. For this to be possible, a transformation and mapping step was incorporated into the interpretation of the timestamps: Each current service point name was replaced by the name of the future planned function (e.g. ‘Pediatrics’ becomes ‘Pediatric Centre’). In the same instance, we also distinguished trails by the type of patient (inpatient/outpatient) and functions present in the trail, in order to colour-code them (see Figure 6a). In this fashion, an outpatient with the trail Radiology(4m)>Pediatric Centre(1h 30m) would be tagged as child. By the assumed type of person, we also choose an entrance space to complete the arrival, i.e. arrival at 10:05: Main Entrance(0m) > Radiology(4m) > Pediatric Centre(1h 30m) > Main Entrance(0m). The passage times between functions are disregarded.

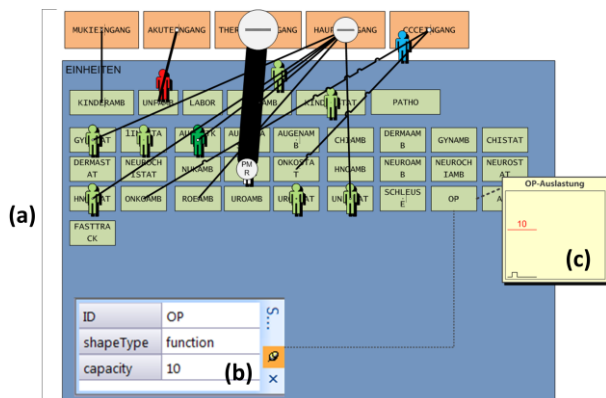


Figure 6: Implementation in Microsoft Visio. (a) Graph editor used for inputting schema and, subsequently, for depicting the flow of agents across a set of functions, which is visualized as bubble-diagram. (b) Manual attribution of shapes and setting of capacity for functions. (c) Depiction of function usage over time.

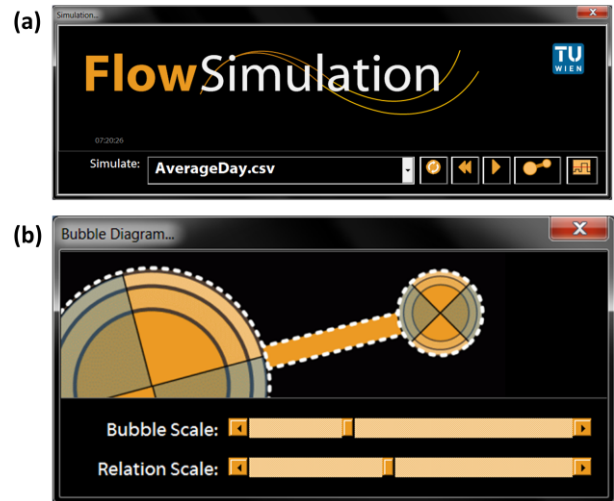


Figure 7: Screens used for the simulation. (a) Main screen with choice of spreadsheet and buttons to start rectangle analysis, reset simulation, run simulation, scale bubbles and compute usage. (b) Scale bubbles screen, giving the choice to scale usage (bubbles) and throughput (relations).

6. IMPLEMENTATION DETAILS

Our implementation uses Microsoft Visio for graph drawing (see Figure 6a). Under the hood, we are employing the bundled Visual Basic for Applications scripting language as means for enabling simulation, i.e. topology analysis and animation/simulation. The former point deserves some more detail: Visio is an excellent tool for doing topological analysis; it comes with a built-in support for finding the spatial relation between pairs of shapes, resulting in either *containing*, *contained in*, *overlapping*, *touching* or *none*. Likewise, the support for animation and automated diagram drawing (which we needed for the bubbles) is excellent, and lead to a total development time of less than a month. Also, the possibility for attribution of shapes via *shape data* (see Figure 6b) proved a valuable tool with which the users were already acquainted. Likewise, the ability to define custom shapes made it possible to put usage monitors (see Figure 6c) into a palette, with users being able to drag them onto the drawing sheet and anchor them to the function to be monitored.

Deployment of the simulation is also fairly easy - the user needs to open a diagram containing the scripts in parallel to the diagram to be simulated. By this, we are able to reach a wide audience that is not tech-savvy, or even (one might say) reluctant to install additional simulation software. For the end-user, the simulation is invoked via an additional menu of Microsoft Visio, which brings up the user interface (refer to Figure 7):

- In the main screen (Figure 7a), the user selects an arrival spreadsheet and can then start the rectangle analysis and simulation.
- During the simulation, the user can scale bubbles and relations (see dialog in Figure 7b). The usage is computed in regular intervals,

which the user can override by pressing a “compute usage” button.

CONCLUSIONS

We have presented a novel approach that helps clients simulate a very preliminary schematical diagram, targeted at the pre-tender phases of a hospital project, where requirements have to be elaborated in close collaboration between medical staff and patients. The output of the approach is suitable as input for the competition, since it speaks the language architects understand (i.e. bubble diagrams, functions and spaces).

REFERENCES

- Alexander, C.A., 1965. A City Is Not A Tree (Part 1), *Architectural Forum*, 122(1), 58-62.
- De Vries, B., Jessurun, A.J., Dijkstra, J., 2002. Conformance Checking by Capturing and Simulating Human Behaviour in the Built Environment, *Proceedings of the 6th Design and Decision Support Systems in Architecture and Urban Planning*, pp. 378-391.
- DIN 13080, 2003. *Gliederung des Krankenhauses in Funktionsbereiche und Funktionsstellen*, Beuth Verlag, Berlin.
- Lohfert, P., 2005. *Methodik der Krankenhausplanung*, Lohfert & Lohfert AS, Copenhagen.
- Seonwook, K., Miyoung, P., 2012, *Architectural and Program Diagrams 1 - Construction and Design Manual*, DOM Publishers, Berlin.
- Tabak, V., 2008. *User Simulation of Space Utilisation*, Ph. D. Thesis, TU Eindhoven.
- White, E.T., 1986. *Space Adjacency Analysis: Diagramming information for architectural design*, Architectural Media, Tucson.
- Wurzer, G., Fioravanti, A., Loffreda, G., Trento, A., 2010. Function & Action: Verifying a functional program in a game-oriented environment, *Proceedings of the 28th eCAADe*, pp. 389-394.
- Wurzer, G., 2010. Schematic Systems – Constraining Functions Through Processes (and Vice Versa), *International Journal of Architectural Computing*, 8(2), pp. 197–213.
- Wurzer, G., 2011. *Prozessvisualisierung in der Krankenhausplanung*, Ph. D. Thesis, Vienna University of Technology.
- Wurzer, G., 2012. In-Process Agent Simulation for Early Stages of Hospital Planning, *Preprints of Mathmod 2012 Vienna*, ARGESIM, Report no. S38 (2012), Paper ID 190.

AUTHORS BIOGRAPHY

Gabriel Wurzer earned his Ph. D. degree in *Process Visualization and Simulation for Hospital Planning* from Vienna University of Technology in 2011. His research in architectural sciences focuses on tool support for early-stage planning of complex buildings, with regular contributions to both Pedestrian and Evacuation Dynamics conference (PED) and the Education and Research in Computer Aided Architectural Design in Europe conference (eCAADe), from which he was awarded the Ivan Petrovic Prize in 2009. He is also an active researcher in archaeological simulation, together with the Natural History Museum Vienna.

Wolfgang Lorenz has a degree in Architecture from Vienna University of Technology and is currently working on his doctoral thesis on *Fractal Geometry and Quality in Architecture*, in which he investigates the concept of applying fractal geometry to architecture. Apart from his research work, he also gives lectures on “programming for architects” at the department of computer aided planning and architecture.

Manfred Pferzinger got his Ph. D. degree from the *Private University for Health Sciences (UMIT)* in 2012, focusing his research on public health and health care quality, in which subject he is author of numerous articles and books. Apart from being associated scientist at UMIT, he is lecturer for the University of Applied Sciences in Upper Austria (Steyr, Austria) and the University of Applied Sciences in Tyrol. He was founding member of the Austrian Competence Circle for Clinical Pathways (A3CP) and currently works as managing director of medipro consulting, a company that focuses on health care efficiency.

TOWARDS LEANER HEALTHCARE FACILITY: APPLICATION OF SIMULATION MODELLING AND VALUE STREAM MAPPING

Waleed Abo Hamad, John Crowe, Amr Arisha

3S Group, College of Business,
Dublin Institute of Technology (DIT),
Aungier Street, Dublin2, Ireland

ABSTRACT

Recently, the application of lean thinking in healthcare has grown significantly in response to rising demand caused by population growth, ageing and high expectations of service quality. However, insufficient justifications and lack of quantifiable evidence are the main obstacles to convince healthcare executives to adopt lean. Therefore, this paper presents a methodology that integrates lean tools with simulation to enhance the quality of patient care in healthcare facilities. This enables healthcare organisations to dedicate more time and effort to patient care without extra cost to the organisation or to the patient. Value stream mapping is used to identify value-added and non-value-added activities. Then, a comprehensive simulation model is developed to account for the variability and complexity of healthcare processes and to assess the gains of proposed improvement strategies. An extensive analysis of results is provided and presented to managers to illustrate the potential benefits of adapting lean practices.

Keywords: lean thinking, modelling and simulation, value stream mapping, healthcare management

1. INTRODUCTION

Science and technology are advancing at a rapid pace; however, the healthcare delivery systems world-wide struggle to cope, especially in their ability to provide high-quality service levels consistently. Healthcare systems are in need of fundamental changes in the way care services are managed. Many patients, doctors, nurses, and healthcare providers are concerned that the care delivered is not, essentially, the care that should be received. The frustration levels of both patients and clinicians have probably never been higher. Policy makers, healthcare providers and managers should provide the quality of care that meets people's needs while improving the efficiency of their business processes based on the best scientific knowledge available. Yet there is strong evidence that this frequently is not the case. Large numbers of disciplined review bodies have reported the scale and gravity in healthcare problems world-wide. More systematic and sophisticated approaches are needed to analyse and manage healthcare processes and to support decision makers and healthcare managers in the provision of

informed decisions and strategies for delivering safe and effective care.

Accordingly, many healthcare organisations have recently adopted Lean management as the performance improvement approach for their systems (Poksinska, 2010). The Lean approach seeks improvements within the existing processes of an organisation without substantial reorganisation requiring high investments (Bahensky et al., 2005). Lean thinking as a philosophy, rather than a stand-alone practice, aims to create a streamlined, high quality system that can achieve a greater level of customer service with minimum cost with little or no waste. Originating in the manufacturing industry from the Toyota Production System (TPS), lean thinking has become one of the most effective management concepts in the world (Iyer et al., 2009). In recent years, many lean academics and experts expanded their lean research beyond manufacturing to include lean supply chain management, lean logistics and lean services and healthcare (Mahfouz et al., 2011).

The main step in Lean healthcare thinking is to put the patient in the foreground and include time and comfort as key performance measures of the system (Womack and Jones, 2003). Lean strategies eliminate process steps that do not add value for patient care, while enhancing those that are valuable and essential. As a result, staff members feel empowered to improve care processes (Spear, 2005). However, insufficient justifications and lack of quantifiable evidence are the main obstacles to convince healthcare executives to adopt lean management. Therefore, this paper presents a framework that integrates lean techniques with simulation to enhance the quality of patient care in healthcare facilities.

2. PROPOSED FRAMEWORK

The proposed framework is a simulation-based lean decision support model for healthcare application. There are three distinct phases to the framework; (1) identification, (2) development and (3) assessment, as illustrated in Figure 1. The framework phases will be discussed in detail in Section 3. The remainder of this section will introduce the core domains that are the foundations of the framework. They are; value stream mapping (VSM), Simulation-based VSM and Healthcare Application.

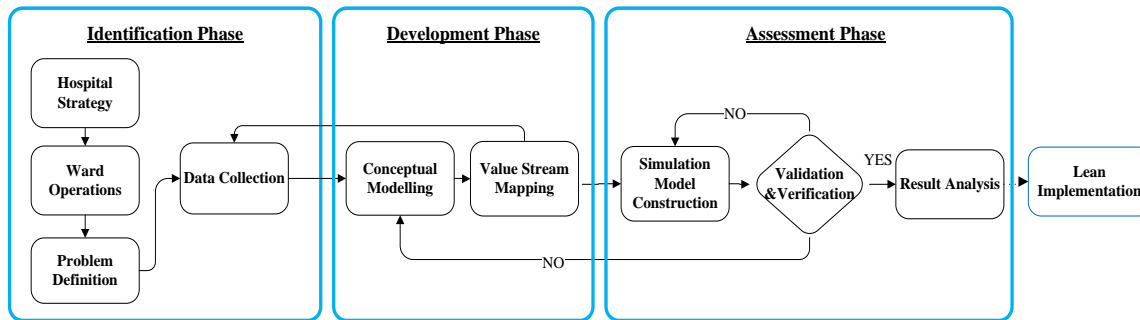


Figure 1. Framework Overview

2.1. Value Stream Mapping

The logic behind lean thinking is pursuing the optimization of value streams from the consumption point of view by eliminating waste and non-value added activities. In order to identify the sources of waste, non-value added activities and opportunities of improvement, value added activities have to be mapped using systematic tools and techniques – VSM technique (Rother and Shook, 1998). A value stream can be defined as the collection of activities (value added and non-value added) that are operated to produce a product or service or a combination of both to a customer (Singh et al., 2006). These actions consider both information and materials flow within the overall system (Abdulmalek and Rajgopal, 2007). The VSM technique demonstrates the material and information flow, maps out value-added and non-value-added activities and provides information about time-based performance. This VSM technique is based on generating a current state map that shows the current performance and conditions of the studied systems and a future state map which serves as the target of improvement actions. VSM has been effectively integrated into several applications, manufacturing and non-manufacturing, due to its simplicity and effectiveness (Tapping and Shuker, 2003). Although the lean concept originated in the automobile industry, the increased application of lean practices in healthcare, has seen growth in the popularity of modelling tool such as VSM (Kim et al., 2006). VSM, as a lean implementation tool, has been successfully utilized in many different healthcare systems, from small physician clinics (Lummus et al., 2006) to more complex systems such as emergency departments (Cookson et al., 2011).

VSM is very effective in presenting system parameters such as operations' cycle time and resources capacity and availability; however it does not have the ability to analyse the system settings impact on performance. Similarly, it is also difficult to know if the best future state regarding to the desired level of system performance is achieved. Moreover, value stream maps do not include information regarding variability (i.e. system variations and uncertainty) (Standridge and Marvel, 2006). Hence, it is required to integrate VSM with another technique that can handle system variation,

show dynamics between system components and validate the future state before the real implementation of the improvement steps. Modelling and simulation capabilities can fulfil this requirement.

2.2. Simulation-based VSM

Simulation can be used to master new business concepts such as agile and lean management (van der Zee and Slomp, 2005). The benefits of using simulation as part of lean and six sigma projects was emphasized by (Ferrin et al., 2005). Unlike VSM, simulation offers more thorough analysis of a system's data including the examination of variability, the determination as to whether the data is homogenous, and the estimation of the probability distribution that fits the data patterns. This kind of in-depth analysis of data enables simulation to be used to support continuous improvement (Adams et al., 1999) and to model systems' future state map showing the ideal state that the system can pursued over time. The advantage of utilizing simulation approach in a lean context is not limited to the phase of developing a future state map but is extended to selecting the best alternative to the current system status. To address its limitations, Shararah et al. (2011) effectively integrated VSM with the more analytically powerful discrete event simulation software and noted that combing both techniques created useful synergies.

The simulation model in this paper is developed over two main phases; (1) creating a conceptual model for the ward drug round and (2) developing a discrete event simulation model that mimics the real-life activities of the drug round. These phases will be discussed in more detail in Section 3.

2.3. Hospital Ward - A Case Study

From an overview perspective, the hospital ward is the location to which patients from the Emergency Department (ED), inter and intra hospital transfer are transported to. Many wards, classified as inpatient wards, also have outpatient facilities. This is due to the location of the specialty medical service provided by the hospital. The arrival rates, patient types and dependency levels of patients are varied and stochastic. This complexity impacts upon operational functionality and exacerbates and exposes inherent flaws within the work

systems. Consequences of operational issues at ward level can have detrimental affects upon the patient, staff, ward and the hospital as a going concern. In short, operational failures, aside from clinical and medical issues, can impact seriously upon a hospitals ability to function. It is estimated that on average operating costs of a hospital account for up to 70% of the overall cost of a hospital (Chaudhury, 2005).

The hospital ward is seldom used as a unit of measure or analysis from the perspective of operations research (Ancarani et al., 2009). More common is the use of the hospital as the unit of measurement with respect to efficiency. This can be explained by the heterogeneous nature of the production processes among hospital wards, the complexity of the information required and the related difficulty in obtaining it (Ancarani et al., 2009). There are a number of benefits to improvement of operational processes within the hospital ward with the goal of providing more time to care for patients. A reduced length of stay (LOS) for the patient may consequently lead to a reduction in excess bed cost per patient, a reduction in staff sicknesses and absences and stock reduction (Improvement, 2011). As outlined in the NHS's Productive Ward series, the aim of the process improvement programs is to release more time for medical staff to have with patients. More direct care is of importance in a number of ways. Patient focused care or more direct care for patients can bring about a reduction in staff costs without compromising care quality or patient satisfaction (Chaudhury, 2005). So there is now an emphasis to improve operational processes at ward level to attempt to free up more time for direct care for patients.

The majority of wards that are currently operational in Ireland are multi occupancy wards. Aside from the medical/nursing advantage of such a design, operational issues exist with this design. Patient transfers in multi

occupancy wards occur on average six to nine times daily and include both internal and external transfers. This represents a significantly higher cost in added paper work, housekeeping, patient transport and medication transport (Bobrow and Thomas, 1994), (Gallant and Lanning, 2001).

The university hospital partner in this research is one of the leading university hospitals in the Republic of Ireland. This 570-bed hospital provides primary, specialised, and tertiary healthcare services. It was agreed with the hospital executive managers to carry out a pilot over a 10 week period on a 30 bedded acute surgical ward to implement the proposed framework. The purpose of this pilot is to highlight non-value added activities with the end result of reducing operational costs, increase staff utilization and most importantly improve patient care within the ward. The detailed implementation of the framework is explained in the next section.

3. FRAMEWORK IMPLEMENTATION

There are three main phases (Figure 1); Identification Phase; Development Phase; and Assessment Phase.

3.1. Identification Phase

The identification phase focuses on system understanding and process analysis. Qualitative research was used in the analysis stage through the collection of primary data of the studied ward. Several field visits, interviews and process analysis sessions are conducted with ward managers in order to frame an understanding about the main parameters and generate an understanding of the ward processes where 25 processes have been identified. A detailed flowchart for each process has been developed and validated with the nursing staff to reflect all the activities within each process. For example, a detailed flowchart of all the activities within the drug round process is given in Figure 2.

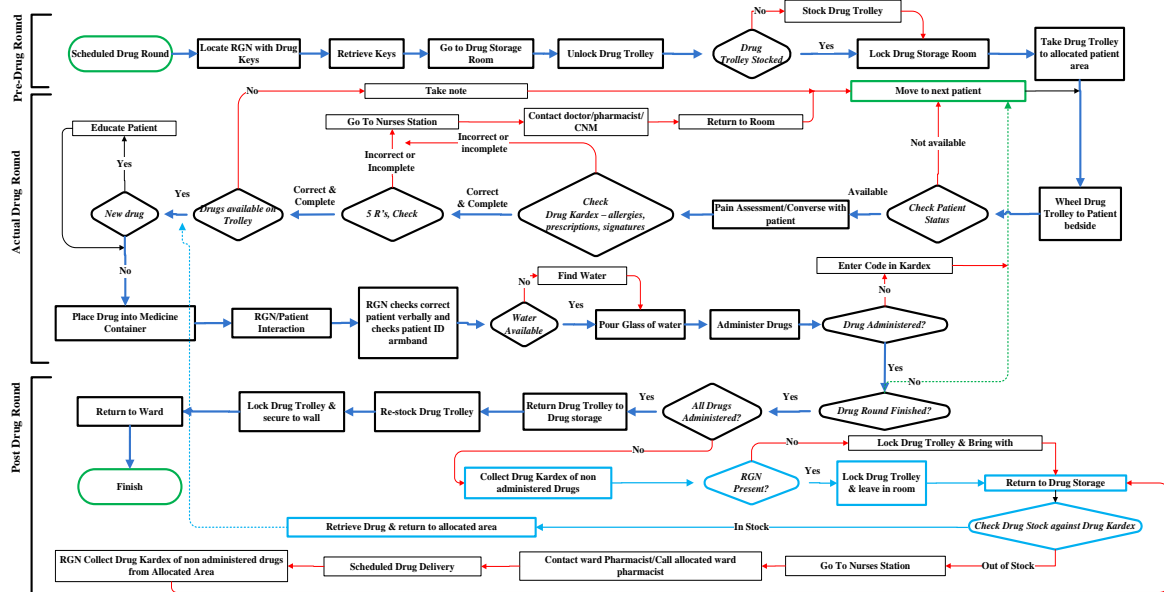


Figure 2. A detailed flowchart of all the activities of the drug round process

3.2. Development Phase

In the development phase, using data and charts developed in the first phase, a VSM was built from the descriptive data for each process in the ward. As discussed earlier, a VSM is based on generating a current state map that shows the current performance and conditions of the processes of the studied ward and a future state map which serves as the performance improvement target. Identifying the value adding activities was the initial step towards creating the current state map. The ward staff current awareness of lean concepts and practices was also a key topic in the discussions and interviews. Meetings were also held with a number of staff nurses with the aim of determining the essential value added and non-value added activities.

Figure 3 shows the current state VSM of the drug round process. A senior nurse manager in the studied ward was interviewed to gather general information about drug round activities and to verify the current state of the system identified in the flowchart (Figure 2). 14 primary processes are mapped along with the cycle time accumulation through each process during a 12 patient drugs round. The cycle timeline is separated into two categories; value added time (green) and non-value added time (red). The VSM is based on value and non-value added time with respect to staff. Time spent in value-added activities direct to the patient are underlined and only represent 23% overall cycle time at 2700 seconds. The activity times circled in red are the parameters chosen for the assessment phase of the framework.

3.3. Assessment Phase

A quantitative model-driven methodology is applied to generate models representing the causal relationships between the lean practices and systems' performance. Based on the detailed flowchart, the current state map and the empirical data analysis, a comprehensive simulation model was developed. Accordingly, the top-level of the simulation model defined the overall process structure, and sub-level blocks comprised additional modules with more details. The main entities for the simulation were patients and nursing staff, where each patient is assigned a set of attributes that represent their location and dependency level. A database was used to save the measured time for all the activities within each process after each simulation run (i.e., replicate), followed by exporting these measures in a tabular form for analysis and validation.

To reduce the model development cycle time and to increase the confidence in the simulation model results, verification and validation were carried out all the way through the development phase to confirm the model represents the actual patient flow. After each model development phase, the model was verified and validated with respect to other previously completed phases. For the verification process, the model logic is verified to ensure that the work flow is mapped correctly. This was achieved by visual tracking of patients and staff using animation and by checking intermediate output values. A warm-up period of one week was found to mitigate any bias introduced by the initial conditions of the simulation model.

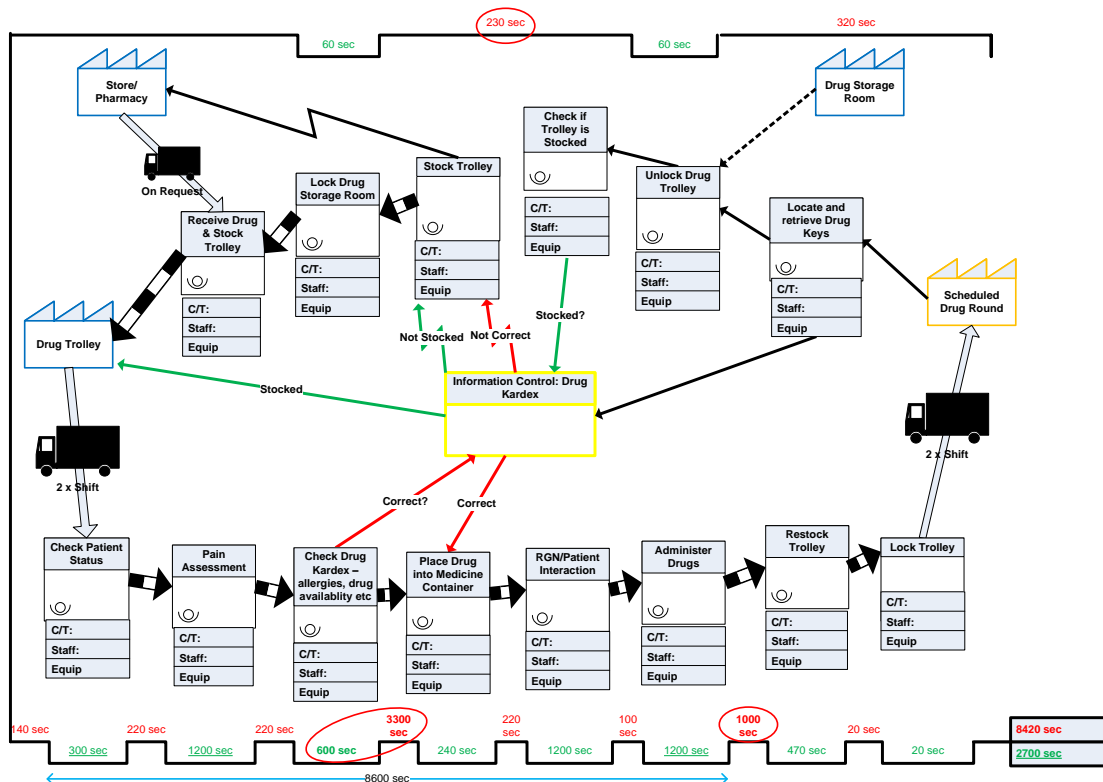


Figure 3. Current State VSM for Drugs Round Process

4. RESULT AND DISCUSSION

In order to show the benefits of the proposed methodology, the discussion in this section is only limited to one process, the drug round process, which was selected by the ward managers to be fully investigated. Regarding the drug round process, the initial findings from the VSM led to 19 standard operations and up to 15 alternative operations when variation occurred (e.g. drug not available, patient allergies etc.). The drug round process was then broken down to three main stages: pre-drug round, actual drug round, and post drug round (Figure 2). The observed timing information highlighted that the overall process takes on average 2.8hrs in order to complete a drug round by a nursing staff for two rooms, each with 6 patients.

As shown in Table 1, 78% of the time is spent in non-value added activities for patients. This highlights the need for improvement in order to increase the time to care for patients by reducing the workload of nursing staff.

Table 1 Analysis of the current state map of the Ward

	Current State (mins)
Pre-Drug Round	10.60
Drug Round per Patient	13.75
Post-Drug Round	09.18
Overall (12 Patient)	173.2
value added	7.5 (22%)
non-value Added	26.03(78%)

Consequently, the ward managers suggested three lean practices (i.e. scenarios) to be tested using the ward simulation model. The first scenario is assuming that the drug trolley is always stocked in the pre-drug round stage. The second scenario is that all the drugs had been successfully administered in the actual drug round (i.e. no rework needed in the post-drug round). Finally, scenario three is to consider no variances in the actual drug round (i.e. all prescriptions are correct and all the required drugs are in the drug trolley). The simulation results for all the scenarios against the baseline (i.e. current state) are shown in Table 2.

Table 2 Simulation results of all the scenarios against the baseline

	Base Line	Scenario 1	Scenario 2	Scenario 3
Pre-Drug Round	010.60	8.5 (-20%)	010.60	010.60
Drug Round per Patient	013.75	013.75	013.75	10.45 (-24%)
Post-Drug Round	009.18	009.18	002.8 (-69%)	02.27 (-75%)
Overall (12 Patients)	173.20	170.4 (-02%)	165.6 (-04%)	136.6 (-21%)

The simulation results shows that scenario three has resulted in a significant reduction in overall cycle time

(i.e. time needed to complete the process) by 21% while scenario's 1 and 2 have a limited impact with a decrease in the cycle time by 2% and 4% respectively. Additionally, the value added activities have been increased from 22% to 32% by adapting more lean practices (Figure 2).

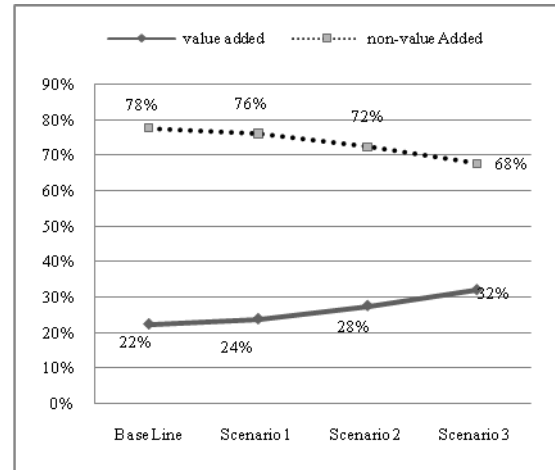


Figure 3 The impact of all the scenarios on the value-added and non-value-added activities

Accordingly, the integrated methodology has shown the ward managers and nursing director the potential benefits of applying lean thinking. This has resulted in initiating a lean training across all departments in the hospital where the framework will be used to assess the impact of proposed practices prior to their implementation.

CONCLUSION

Lean thinking is regarded as a systematic approach used to identify and eliminate waste or non-value added activities in a process through continuous improvement. Lean thinking is a philosophy that can be applied to a variety of organisations since its focus is on improving process performance. Most service organisations are made up of a series of processes, or a set of activities or steps intended to create value for those who are dependent on them – customers. Lean is now being successfully applied to healthcare applications as it reduced the slack on service time. This enables healthcare facilities to reduce operation costs and improve service quality. In order to identify the sources of waste, non-value added activities and opportunities of improvement, value added activities have to be mapped using systematic tools and techniques such as value stream mapping (VSM) technique. VSM is very effective in presenting system parameters such as operations' cycle time and resources capacity and availability; however it does not have the ability to analyse the system settings impact on performance. Unlike VSM, simulation modelling can handle system variation, show dynamics between system components and validate the proposed improvement steps before the real implementation. Therefore, an integrated simulation-based VSM methodology was developed

and applied to an acute surgical ward in a teaching hospital in Dublin. The developed framework allowed senior decision makers in the hospital to plan and visualise how new ward layouts will operate (i.e. single-bedded rooms). Moreover, they were able to foresee the consequences of different scenarios and strategies on ward performance before implementation takes place.

REFERENCES

- ABDULMALEK, F. A. & RAJGOPAL, J. 2007. Analyzing the benefits of lean manufacturing and value stream mapping via simulation: A process sector case study. *International Journal of Production Economics*, 107, 223-236.
- ADAMS, M., COMONATON, P., CZARNECKI, H. & SCHROER, B. J. Simulation as a tool for continuous process improvement. 1999. IEEE, 766-773 vol. 1.
- ANCARANI, A., DI MAURO, C. & GIAMMANCO, M. D. 2009. The impact of managerial and organizational aspects on hospital wards, efficiency: Evidence from a case study. *European Journal of Operational Research*, 194, 280-293.
- BAHENSKY, J. A., ROE, J. & BOLTON, R. 2005. Lean sigma—will it work for healthcare. *J Healthc Inf Manag*, 19, 39-44.
- BOBROW, M. & THOMAS, J. 1994. Hospitals' prosperity should be by design. *Mod Healthc*, 24, 54.
- CHAUDHURY, H. 2005. Advantages and Disadvantages of Single-Versus Multiple-Occupancy Rooms in Acute Care Environments: A Review and Analysis of the Literature. *Environment and Behavior*, 37, 760-786.
- COOKSON, D., READ, C., MUKHERJEE, P. & COOKE, M. 2011. Improving the quality of Emergency Department care by removing waste using Lean Value Stream mapping. *International Journal of Clinical Leadership*, 17, 25-30.
- FERRIN, D. M., MILLER, M. J. & MUTHLER, D. Lean sigma and simulation, so what's the correlation?: V2. 2005. Winter Simulation Conference, 2011-2015.
- GALLANT, D. & LANNING, K. 2001. Streamlining patient care processes through flexible room and equipment design. *Crit Care Nurs Q*, 24, 59-76.
- IMPROVEMENT, N. I. F. I. A. 2011. Rapid Impact Assessment Executive Summary FINAL. NHS Institute for Innovation and Improvement.
- IYER, A. V., SESHADRI, S. & VASHER, R. 2009. *Toyota Supply Chain Management*, New York, McGraw-Hill.
- KIM, C. S., SPAHLINGER, D. A., KIN, J. M. & BILLI, J. E. 2006. Lean health care: What can hospitals learn from a world-class automaker? *Journal of Hospital Medicine*, 1, 191-199.
- LUMMUS, R. R., VOKURKA, R. J. & RODEGHIRO, B. 2006. Improving quality through value stream mapping: A case study of a physician's clinic. *Total Quality Management*, 17, 1063-1075.
- MAHFOUZ, A., CROWE, J. & ARISHA, A. Integrating Current State and Future State Value Stream Mapping with Discrete Event Simulation: A Lean Distribution Case Study. The Third International Conference on Advances in System Simulation (SIMUL 2011), 2011 Barcelona. 169-176.
- POKSINSKA, B. 2010. The current state of Lean implementation in health care: Literature review. *Quality Management in Healthcare*, 19, 319.
- ROTHER, M. & SHOOK, J. 1998. Learning to See: Value Stream Mapping to Create Value and Eliminate Muda. v. 1.1. Oct., *The Lean Enterprise Inst., Brookline, Mass.*
- SHARARAH, M. A., EL-KILANY, K. S. & EL-SAYED, A. E. 2011. Value Stream Map Simulator Using ExtendSim. *Lecture Notes in Engineering and Computer Science*, 2190.
- SINGH, R., KUMAR, S., CHOUDHURY, A. & TIWARI, M. 2006. Lean tool selection in a die casting unit: a fuzzy-based decision support heuristic. *International journal of production research*, 44, 1399-1429.
- SPEAR, S. J. 2005. Fixing health care from the inside, today. *Harvard Business Review*, 83, 78.
- STANDRIDGE, C. R. & MARVEL, J. H. Why lean needs simulation. 2006. Winter Simulation Conference, 1907-1913.
- TAPPING, D. & SHUKER, T. 2003. *Value stream management for the lean office*, New York, Productivity Press.
- VAN DER ZEE, D. J. & SLOMP, J. Simulation and gaming as a support tool for lean manufacturing systems: a case example from industry. 2005. Winter Simulation Conference, 2304-2313.
- WOMACK, J. P. & JONES, D. T. 2003. *Lean thinking: banish waste and create wealth in your corporation*, Simon and Schuster.

AUTHORS BIOGRAPHY

WALEED ABO-HAMAD, Ph.D. is a Senior researcher in 3S Group (A research unit in Dublin Institute of Technology (DIT) specialized in complex systems simulation and optimization). He joined the 3S group in 2008 having spent four years as a researcher in Cairo University-Egypt where he received his B.Sc. and M.Sc. degrees in Computer Science. He received his Ph.D in process optimization from DIT and is currently a project manager at DIT. His research interests include

Modelling and Simulation, Optimization, Computational Intelligence, Machine Learning and Cooperative Intelligent Systems. His email address is Waleed.Abohamad@dit.ie.

JOHN CROWE is a researcher and assistant lecturer in Dublin Institute of Technology (DIT). He joined the 3S group in 2009 with five years industry experience in logistics and inventory analysis. He received his B.Sc. degree specialized in Supply Chain Management from DIT with first class honors. Mr. Crowe won several outstanding prizes during his study time in DIT. In 2009, John successfully won a three year research scholarship with the Irish Research Council for Science, Engineering and Technology (IRCSET). He is currently working towards completing his PhD in Dublin Institute of Technology. John has published several articles in innovation in education and simulation international conferences. His research interests include supply chain management, simulation modelling application in business process analysis, system dynamics and simulation-based learning and education. His email address is john.crowe@dit.ie.

AMR ARISHA, Ph.D. is the director of the 3S Group and the head of international business department in college of business in Dublin Institute of Technology. He Joined DIT in 2005, he received his PhD Industrial Engineering from Dublin City University (DCU). Intel-Ireland has sponsored his research from 2000 - 2005. Dr. Arisha has published many journal and international conference articles in the area of manufacturing systems simulation and scheduling. His research interests include analysis and optimization of complex dynamic systems in manufacturing and service industries using stochastic simulation modelling, and optimisation techniques. His research also includes Supply Chain Management, Operations Excellence, Project Management, Virtual Reality applications in Business Process Analysis, and Optimization for Process Improvement. He is a member in IIE, IMECH, IEI, ESE, ORS, IEEE and ASME. His e-mail address is amr.arisha@dit.ie.

APPLICATION OF DISCRETE SYSTEMS SIMULATION TO REDUCE WAITING TIME IN THE OUTPATIENT SERVICE OF A HOSPITAL IN THE CITY OF SÃO PAULO, BRAZIL

Alexandre Augusto Massote^(a), Domenico Caruso^(b), João Batista Gonçalves Sousa^(c)

^(a)Centro Universitário da FEI

^(b)Centro Universitário da FEI

^(c)Centro Universitário da FEI

^(a)massote@fei.edu.br ^(b)domenico-caruso@hotmail.com ^(c)jbg.sousa@terra.com.br

ABSTRACT

This paper aims to analyze failures in a service at the arrhythmology outpatient units, using as support the simulation of discrete systems, the sizing of an outpatient service unit in terms of the nursing staff required, the room occupation and the waiting time. In particular, the cardiology outpatient service in a hospital located in São Paulo, Brazil, is investigated. The simulation model developed showed consistent results and proved to be an excellent tool to support the decision-making and to enhance the quality of care.

Keywords: simulation, health management, decision support system

1. INTRODUCTION

Hospitals and outpatient services around the world are seeking ways to increase efficiency in outpatient care by reducing waiting time and its variability. The objective of this study was to evaluate the flaws in an answering service in the arrhythmology outpatient units and generate alternative scenarios that could provide improved customer service. For this purpose, we developed a discrete event simulation model which allowed the assessment of nursing staff requirements, the room occupation and the waiting time. The unit is composed of offices dedicated to treatment in the field of electrophysiology and rooms dedicated to tests, such as installation procedures Hölters, ABPM (Ambulatory Blood Pressure), ECG (Electrocardiogram), SAECG (High Resolution Electrocardiogram), Tilt-test (Test of Slope) and CETE (Transesophageal Cardiac Stimulation.)

The sector of arrhythmia occupies the 5th floor of a complex of buildings of a hospital in São Paulo. The total area occupied by the sector is 240 m² and attends an average of 106 patients per day in consultations with medical specialists in arrhythmology. Currently, the service has a waiting time for the patient which exceeds the legally acceptable 30 minutes. Additionally, the incidence of overtime of scheduled working hours for the nursing staff is high. The waiting room remains crowded over the day and it is not uncommon that the closing time has to be delayed by as much as two hours.

According to Jun (et al., 1999) and Dittus (et al., 1996), the simulation has been used extensively to plan health services. Katsaliaki and Mustafee (2009) conclude in their article that the simulation allows stakeholders to make appropriate decisions based on statistical parameters that represent the reality of the simulated system, providing the same experiment with different strategies for resource allocation. Quality improvement in decision making leads to better utilization of available resources, which improves the quality of health services provided to citizens. According to Brailsford (2005), the simulation has been studied in universities for 66 years and 46 are used in health care. The health system in the UK has used simulation models since 1962. From these early experiences, the work of simulation in health care has generated excellent results, supported clinical decision making, facilitated the planning, allocated resources for evaluating and redesigning processes.

Several studies have proven that the simulation of discrete systems is useful in improving care services for outpatients. Ashton et al. (2005) used simulation to identify problems in a call center "walk-in" and suggest improvements in the waiting room and triage processes. Stahl et al. (2004) used simulations to analyze alternative configurations of a staff of anesthesiologists to increase efficiency in performing surgery. Stafford and Aggarwal (1979) developed a model of discrete event simulation for outpatient care of a university clinic. The model was used to explore the effects of different levels of staff and the aggregation of two or more service units within the clinic. Chwen et al. (2003) reported that consumers are demanding more nowadays than ever before. In an intensely competitive world, the pressure, the expectations, and the need to perform more tasks in less time are increasing rapidly.

However, some studies also show the difficulty of simulation models implementation in health institutions. Watt (1977) and Wilson (1981) describe the factors that may be barriers to the implementation of simulation projects in hospitals: the culture of health institutions, where medical professionals and nursing are highly resistant to change; the high cost of simulation software and training of professionals specialized in this area; the

low quality of existing data systems in hospital management; and the high cost of specialized consulting in this area. This factor takes the health institutions to join the universities participating in theoretical studies, which conflicts with the need for hospital administrators, who want a simple and rapid implementation. There is also the difficulty of generic models application in hospitals, since each institution sees its internal processes as totally different from other hospitals, generating the need to create a multitude of different models.

The research carried out by Wilson (1981) shows that from the 200 articles analyzed on simulation projects in health, only 16 were successful. The common characteristics of these articles were: at least one of the authors should be part of the hospital staff and have an interest in solving the problem, the problem must be of high priority for the institution, the funding for the project must be outside and there must be a detailed description of existing data for analysis. Fortunately, all these assumptions cited characterized the development of this study.

As in most health facilities in the country, the application of discrete simulation systems to improve hospital processes is new. Traditionally, empirical methods are used to quantify the nursing staff, and the dimensions and capacity of the reception and waiting room. These measures are determined based solely on experience and observation of similar services within the institution or elsewhere.

The simulation results are consistent with the perception of health professionals and proved to be an excellent tool to support the decision-making. The results provided important information for managerial decision making in relation to the sizing of staff and enhancing the quality of care.

The approach presented in this paper can also bring an important contribution to the design and management of the Brazilian health system, because these services receive a greater number of patients every day with increasing levels of demand, requiring improvements in efficiency and quality of care. Since Brazil is an emerging country, the resources available for investment in the area do not follow the increasing demand, which implies an absolute necessity to invest in the best way possible. Hospitals and outpatient services from around the world are seeking ways to increase efficiency in outpatient care, reducing waiting time and its variability. A simulation model can be a great advantage, providing an overview of the care process and allowing the adjustment of several variables, such as layout, staffing and patient flow.

2. MODEL DEVELOPMENT

Site visits were conducted to monitor the care process and measurement time for each procedure performed in the service. The frequency of patient arrival was obtained through the records of entry into the building's concierge and reception service.

Two nurses accompanied these visits: the manager

of the building where the service works and the nurse coordinator, who in addition to other areas, manages the work in the sector of arrhythmology.

2.1. Collected Observations and Information

Visits were made in locus to monitor the process of care and perform measurements of service time for each procedure performed in the service unit. Figure 1 shows the simplified layout of the simulated system.

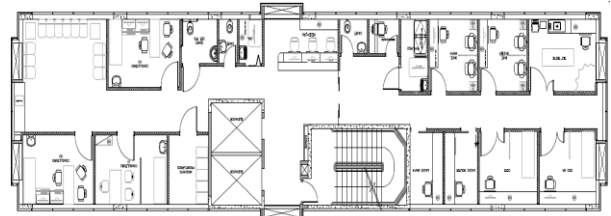


Figure 1: Layout of the simulated system

2.1.1. Structural Information

The Arrhythmia Service is open Monday through Friday, from 7 am to 7 pm. Following are the features of the simulated system before the exams, which are the main focus of the study:

- Board of nursing staff: Two nursing technicians per shift of six hours.
- Reception Service: Capability to care for two patients at a time.
- Waiting Room: Capacity for 12 patients.

2.1.2. Flow of Patients and Time Collected

Figure 2 shows the flow of patients by type of examination performed in the field of arrhythmia, with the average time measured during visits to the sector.

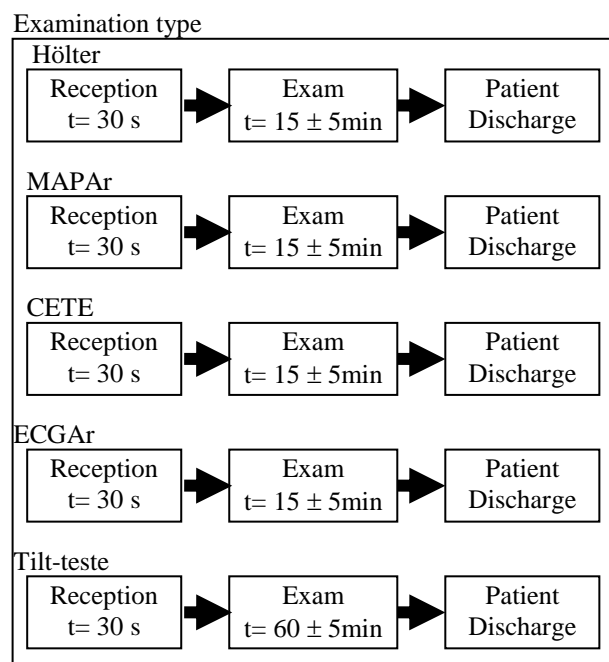


Figure 2: Flow of patients and time for each examination type

The collected data were statistically analyzed and the result is the probability distribution of the time variable in all types of test, following a uniform probability distribution.

Figure 3 shows the flow and the time for queries.

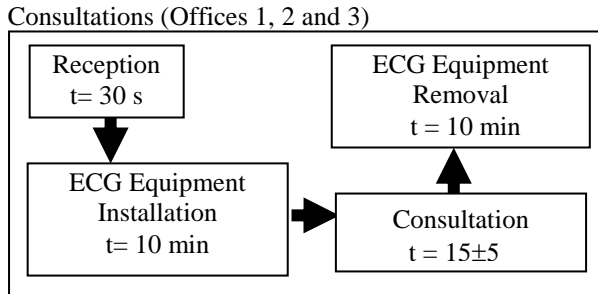


Figure 3: Flow of patients and time for queries

2.1.3. Information about the demands

We analyzed the agenda of patients who used the services sector of arrhythmia during a month. The average daily attendance is 106 patients. We chose to stratify patients by type of care performed by creating percentages of the total number of daily visits.

2.1.4. Development of Simulation Model

From the information collected was used Promodel simulation software, version 7.0, to implement the model, considering the typical work and for generating the various scenarios analyzed.

At first, it was not allowed by the direction of the hospital to consider scenarios involving possible changes to the layout to improve flow of movement, as well as changes in the provisions of the various rooms and offices. Therefore, the only variable that could be modified in various scenarios was analyzed for the quantity of nursing staff accompanying examinations and consultations. Appendix A shows the developed model.

3. RESULTS

Table 1 shows the average utilization of the nursing resources in percentage according to the dimensioning of the nursing workforce by scenario.

Table 1: Nurses utilization

Scenario	Number of Nurses	Quantity Utilized	Utilization %
1	2	96	92.06
2	3	132	87.99
3	4	132	77.68

Table 2 presents the number of patients who were treated within the prescribed period between 7 am and 7 pm. One can assess the demand is met within the service in the simulations involving three four nursing technicians per shift.

Table 2: Demand Achievement

Scenario	Number of Nurses	Demand	Patients Treated
1	2	106	78
2	3	106	106
3	4	106	106

Table 3 presents information on the occupation of the waiting room, according to the design of the framework for nursing. The waiting room was kept fully occupied in 84.04% of the time, with two nursing technicians. As shown in table 4, the average waiting time of patients enrolled in the waiting room was 28.41 minutes for the same situation. This reality is not compatible with the quality of care being provided by the service in question. According to the laws of the country, the maximum waiting time for care in a service is 30 minutes. In the configuration shown, the highest average waiting time is very close to the limit accepted. Certainly, there are patients waiting longer than prescribed by law.

In the configuration that uses three practical nurses per shift, the waiting room remains full for 38.7% of the time, but does not cause failures in patient care. That means the maximum waiting is 31.11 minutes. The configuration with four nursing technicians also presents acceptable values of the waiting room occupancy and maximum average for care.

Table 3: Waiting room occupation

Scenario	Number of Nurses	Fully Occupied (%)	Partially Occupied (%)
1	2	84.04	13.71
2	3	38.07	56.50
3	4	0.00	91.06

Table 4: Average waiting time per patient

Scenario	Number of Nurses	Average Waiting Time (min)
1	2	28.41
2	3	11.31
3	4	3.03

The simulation model was developed with a specific objective of evaluating the staff necessary to meet the demand for an outpatient service. It should be noted that the need for new employees can be met by idle professionals from other sectors, or by improving the efficiency of service processes. Currently, the overtime necessary to meet the demand of the service accounts for 12% of the cost of hiring, wages, and benefits of the employees. The annual cost of a nursing technician comes to \$48,800.00.

Defining the two most appropriate scenarios, and taking into account the premise that the design of two practical nurses per shift does not meet the needs of service demand, the manager responsible for the sector may make the decision to hire new staff or not, with

reference parameters that allow a decision about quality with high reliability and within the policies established by the institution and the costs involved in hiring new workers. Appendix B shows a typical Promodel report.

4. CONCLUSIONS

Brazil has today a strong demand for health services. According to Ceres et al. (2008), the private insurance market is significant in the Brazilian health system. In December 2006, involving 44.7 million bonds of beneficiaries, of whom 82.7% to health care plans and 173% exclusive dental plans, 2,070 operating companies, more than 20,000 plans and thousands of service providers. The population using health services is becoming more demanding about quality.

The simulation showed results consistent with the perception of industry professionals. It has been shown by several researchers that the simulation can clearly represent a real situation, with details that support a logical decision. In this particular case, the largest benchmark was used in the waiting time simulation, which was very similar to the real time. This article presents a model that provides important management information for decision making in relation to the scale of staff and enhancing the quality of care. In a health service, the great need to reduce costs and increase process efficiency creates an environment conducive to the use of simulation models of discrete systems.

APPENDIX A - THE DEVELOPED MODEL

Locations

Nome	Cap	Unidade	Estatísticas
Recepção	inf	1	Série de Tempo
Sala_de_espera	12	1	Série de Tempo
Tilt_teste	1	1	Série de Tempo
Mapa	1	1	Série de Tempo
Posto_de_enfer.	1	1	Série de Tempo
Holter	1	1	Série de Tempo
ECG_AR	1	1	Série de Tempo
Consultório_2	1	1	Série de Tempo
Consultório_1	1	1	Série de Tempo
Consultório_3	1	1	Série de Tempo
CETE	1	1	Série de Tempo

Entities

Nome	Velocidade (mpm)	Estatísticas
Paciente_exame	50	Série de Tempo
Paciente_tilt	50	Série de Tempo
Paciente_mapa	50	Série de Tempo
paciente_holter	50	Série de Tempo
Paciente_ECG_AR	50	Série de Tempo
Paciente_cete	50	Série de Tempo
Paciente_cons1	50	Série de Tempo
Paciente_cons2	50	Série de Tempo
Paciente_cons3	50	Série de Tempo

Path networks

Nome	Tipo	T/S	De	Para	BI	String
Red1	Não-Pass	Veloc.& Dist.	R1	R2	Bi	17.76

Recursos

Nome	Unid	Estatísticas	Pesquisar
Enfermeira	4	Por Unidade	Mais Próx./Velha
Caminho		Movimentação	
Red1		Vazio: 50 mpm	Cheio: 50 mpm

Process

Operation

Entidade	Local	Operação
Paciente_exame	Recepção	wait 0.5

Rote

BI	Saída	Destino	Regra	Lógica de Mov.
1	Paciente_exame	Sala_de_espera	first	1

Operation

Entidade	Local	Operação
Paciente_exame	Sala_de_espera	

Rote

BI	Saída	Destino	Regra	Lógica de Mov
1	Paciente_tilt	Tilt_teste	0.030000	1 move with Enfermeira
	Paciente_mapa	Mapa	0.240000	move with Enfermeira
	Paciente_holter	Holter	0.250000	move with Enfermeira
	Paciente_ECG_AR	ECG_AR	0.19	move with Enfermeira
	Paciente_cete	CETE	0.020000	move with Enfermeira
	Paciente_cons1	Consultorio_1	0.09	move with Enfermeira
	Paciente_cons2	Consultório_2	0.09	move with Enfermeira
	Paciente_cons3	Consultorio_3	0.09	move with Enfermeira

Operation

Entidade	Local	Operação
Paciente_tilt	Tilt_teste	wait u(60,5) Free Enfermeira

Rote

BI	Saída	Destino	Regra	Lógica de Mov
1	Paciente_tilt	EXIT	first	1

Operation

Entidade	Local	Operação
Paciente_mapa	Mapa	wait u(15,5) Free Enfermeira

Rote

BI	Saída	Destino	Regra	Lógica de Mov
1	Paciente_mapa	EXIT	first	1

Operation

Entidade	Local	Operação
paciente_holter	Holter	wait u(15,5) Free Enfermeira

Rote

BI	Saída	Destino	Regra	Lógica de Mov
1	paciente_holter	EXIT	first	1

Operation

Entidade Local **Operação**
 Paciente_ECG_AR ECG_AR wait u(15,5)
 Free Enfermeira

Rote

Bl Saída Destino Regra Lógica de Mov
 1 Paciente_ECG_AR EXIT first 1

Operation

Entidade Local **Operação**
 Paciente_cete CETE wait 15
 Free Enfermeira

Rote

Bl Saída Destino Regra Lógica de Mov
 1 Paciente_cete EXIT first 1

Operation

Entidade Local **Operação**
 Paciente_cons1 Consultorio_1 wait 10
 Free Enfermeira
 wait 15
 get Enfermeira
 wait 10
 Free Enfermeira

Rote

Bl Saída Destino Regra Lógica de Mov
 1 Paciente_cons1 EXIT first 1

Operation

Entidade Local **Operação**
 Paciente_cons2 Consultório_2 wait 10
 Free Enfermeira
 wait 15
 get Enfermeira
 wait 10
 Free Enfermeira

Rote

Bl Saída Destino Regra Lógica de Mov
 1 Paciente_cons2 EXIT first 1

Operation

Entidade Local **Operação**
 Paciente_cons3 Consultorio_3 wait 10
 Free Enfermeira
 wait 15
 get Enfermeira
 wait 10
 Free Enfermeira

Rote

Bl Saída Destino Regra Lógica de Mov
 1 Paciente_cons3 EXIT first 1

Chegadas

Entidade Local **Quantidade** **Primeira Vez**
 Paciente_exame Recepção 1
Ocorrências **Frequência** **Lógica**
 106 U(5,1)

APPENDIX B – PROMODEL REPORT

The screenshot displays a Promodel simulation report with multiple tables. The top table, titled 'General Report (Execução Normal - Rep. 1)', lists entities like Recepção, Sala de espera, TÁ haste, Mapa, and various Consultorio and CETE resources, along with their scheduled times, capacities, and utilization percentages. Below this are several smaller tables, each representing a different model (e.g., 'model1 MOD (Execução Normal - Rep. 1)'), which provide detailed metrics such as 'Number Times Used', 'Avg Time Per Usage (MIN)', and 'Avg Time Travel To Use (MIN)' for specific resources like Enfermeira1 through Enfermeira4.

REFERENCES

Albuquerque, C., et al., 2008. A situação atual do mercado da saúde suplementar no Brasil e apontamentos para o future; *Ciências e Saúde Coletiva*, 13(5), Brasil.

Ashton, R., Hague, L., Brandreth., M., Worthington, D., Cropper, S., 2005. A simulation-based study of a NHS walk-in centre. *Journal of Operational Research Society*, 56, 153–161.

Dittus, R.S., Klein, R.W., Debrot, D.J., Dame, M.A., Fitzgerald, J.F., 1996. “Medical resident work schedules: design and evaluation by simulation modelling”. *Management Science*, 1996, 42(6), 891-906.

Joint Comission Resources, Gerenciando o Fluxo de pacientes – Estratégias e Soluções para Lidar com a Superlotação Hospitalar; *Editora Bookman*, 1ª Edição, Brasil.

Jun J., Jacobson, S., Swisher, J., 1999. Application of discrete-event simulation in health care clinics: a survey. *Journal of Operational Research Society*, 50, 109–123.

Katsaliaki, K., Mustafee, N., 2010. Improving decision making in healthcare services through the use of existing simulation modelling tools and new technologie. *Transforming Government. People, Process and Policy*, 4,(2), 158-171.

Nerenz et al, 2011. A simulation model approach to analysis of the business case for eliminating health

- care disparities. *BMC Medical Research Methodology*, 11-31.
- Robinson, S., 2005. Discrete-event simulation: from the pioneers to the present, what next??. *Journal of the Operational Research Society*, 56,(6), pp. 619-29.
- Rohleder, T.R., Bischak, D.P., Baskin, L.B., 2007. Modeling Patient Services Centers with Simulation and System Dynamics. *Health Care Management Science*, 10, 1–12.
- Ruquet, M.E., 2000. Simulation Modeling Improves Service. *National Underwriter. Life & Health*, 104- 37.
- Sheu, C., Mchaney, R., Babbar, S., 2003. Service Process Design Flexibility and Customer Waiting Time. *International Journal of Operations & Production Management*, 23, 7-8, Abi/Inform Global, pg. 901.
- Smith, D.J., 1994. Computer Simulation Applications in Service Operations: A Case Study From the Leisure Industry. *The Service Industries Journal*, 14-3, Abi Inform Global, 395.
- Stafford E.F.Jr, Aggarwal, S.C., 1979. Managerial analysis and decision-making in outpatient health clinics. *Journal of Operational Research Society*, 30, 905–915.
- Stahl, J.E., Rattner, D., Wiklund, R., Lester, J., Beinfeld, M., Gazelle, G.S., 2004. Reorganizing the system of care surrounding laparoscopic surgery a cost-effectiveness analysis using discrete-event simulation. *Medicine Decision Making*, 24, 461–471.
- Watt, K., 1977. Why won't anyone believe us? *Simulation*, 28, 1-3.
- Wilson, JCT, 1981. Implementation of computer-simulation projects in healthcare. *Journal of Operational Research Society*, 32: 825-832.

AUTHORS BIOGRAPHY

Massote, A.A.: Graduate in Mechanical Engineering by Centro Universitário da FEI (1979), Master in Industrial Engineering by COPPE/UFRJ (1991) and Ph.d. in Industrial Engineering by the Polytechnic School of the University of São Paulo (2001). Has experience in operational research, economic viability and production control. Research efforts are focused on the logistics, supply chain management and simulation. Currently he teaches simulation undergraduate degree in Industrial Engineering and master's program in Industrial Engineering, is also Head of Industrial Engineering Department of Centro Universitário da FEI, São Bernardo do Campo, Brasil

UNCERTAINTY QUANTIFICATION FOR CEREBRAL PERFUSION

Rachael Gordon-Wright^(a), Pierre A. Gremaud^(b), Esther G.H.J. Martens^(c), Vera Novak^(d)

^(a)Department of Mathematics, North Carolina State University, Raleigh, USA

^(b)Department of Mathematics, North Carolina State University, Raleigh, USA and Statistical and Applied Mathematical Sciences Institute (SAMSI), RTP, USA

^(c)Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

^(d)Division of Gerontology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA

^(a)rkagordon@ncsu.edu, ^(b)gremaud@ncsu.edu, ^(c)e.g.h.j.martens@tue.nl, ^(d)vnovak@bidmc.harvard.edu

ABSTRACT

There is a pressing need for noninvasive and continuous measurements of cerebral blood flow (CBF) in several areas of medicine. Transcranial Doppler (TCD) technology is clinically used for measurements of blood flow velocities (BFV). It is assumed that perfusion and vasoreactivity in a vascular territory can be inferred from BFV measurements in the corresponding stem artery. However, only very modest correlations have been found between TCD-based and magnetic resonance imaging (MRI)-based CBF measurements. Several factors, whose values are uncertain, such as vessel diameter, hematocrit and insonation angle, affect the BFV - CBF relationship. Their influence on CBF and vasoreactivity estimates has not been rigorously analyzed but cannot be ignored. We present initial work toward a subject specific computational and experimental model to both quantify and reduce the uncertainty attached to CBF and vasoreactivity estimates based on Doppler ultrasound.

Keywords: uncertainty quantification, perfusion, transcranial Doppler, mathematical modeling

1. INTRODUCTION

Clinical researchers often have to rely on TCD to address questions about rapid CBF regulation and vasoreactivity. However, in spite of long term use, obvious needs and recent advances, further expansion of the TCD methodology in clinical care is impeded by uncertainties inherent to the approach: insonation angles, vessel diameter, skull thickness, blood rheology or even topology of the main cerebral arteries are only approximately known. While the impact of these uncertainties has yet to be fully analyzed, our preliminary studies show the link between measured BFV in main arteries and perfusion or vasoreactivity to be only tenuous.

Magnetic resonance based methods such as Continuous Arterial Spin Labeling (CASL) allow for noninvasive CBF measurements and assessments of regional vascular reactivity; anatomical measurements can also be co-registered. Unlike TCD, they have significant practical limitations: (i) they cannot be used

for continuous CBF monitoring at bedside, (ii) recordings need to be averaged thus limiting time resolution, (iii) they cannot be used for studies involving postural challenges, (iv) they are costly and (v) have their own limitations i.e. normalization procedures to a standard space, T1 effects and blood rheology effects and need for contrast agents to increase signal to noise ratio. Therefore, bypassing the above bottleneck for the TCD methodology is of fundamental practical importance.

Computational modeling has been successful at predicting local blood flow (for instance in the presence of aneurysms) for “generic” patients, see e.g. Castro et al. (2009), Greenberg et al. (2009), Kim et al. (2009), Kleinstreuer et al. (2007), Watton et al. (2009). Due to their complexity, these approaches do not easily lend themselves to patient specific predictions or to statistical studies which may require large numbers of runs. In other words, it is difficult in that framework to make judicious use of relevant data. Our approach is different and is based on computationally inexpensive one-dimensional (partial differential equations) and zero-dimensional (algebraic and differential equations) models, see Devault et al. (2008). Two main challenges usually absent in hemodynamics are central here: (i) how to generate patient specific simulations and (ii) how to track and reduce uncertainties affecting the data, the model, the parameters and the geometry. The first question pertains to data assimilation and the second to uncertainty quantification.

2. DATA ANALYSIS

The acquired database includes 167 older adults (ages between 50 and 85 years old). A complete dataset for each subject includes: demographic data, laboratory values, TCD-based BFV during a 10 minute supine rest, CASL-based perfusion, blood pressure CO₂ at baseline, during hypercapnia and hypocapnia and cognitive challenge; anatomical high resolution T1- and T2-weighted MR imaging and time of flight, MR angiography (TOEF MRA) at 3 Tesla to characterize brain tissue volumes and the Circle of Willis (CoW) and plasma hematocrit. Subjects participated in the protocols approved by the Institutional Review Board at

the Beth Israel Deaconess Medical Center. The demographic characteristics are: 83 women, mean age 65.7 ± 8.4 , 84 hypertensive, 83 normotensive, 62 diabetic. TCD waveform were reviewed and only good quality recordings were accepted for the analysis. We consider in the present study only measurements for the Middle Cerebral Artery (MCA).

2.1. Model and assumptions

The calculations from TCD measured BFV to CBF involves several assumptions. TCD measures Doppler shifts, see for instance Fodale et al (2007). More precisely, if f_0 is the ultrasound frequency at the source (usually around 2MHz), elementary physics implies

$$\text{Doppler shift} = 2 f_0 v/c \tag{1}$$

where v is the apparent velocity the moving erythrocytes in the vessel under consideration and c is the speed of sound in soft tissues (usually taken as about 1540 m/s). If the angle of incidence is zero, i.e., if the source is perfectly lined up with the vessel then v is equal to the velocity of the red blood cells in the vessel. Otherwise, if θ is the angle between the probe and the vessel, i.e., the insonation angle, then

$$v = u \cos\theta \tag{2}$$

where u is the actual velocity. The BFV is assumed to have the following profile

$$u(x,r,t) = (k+2)/k (1-(r/R(x,t))^k) U(x,t) \tag{3}$$

where $U(x,t)$ is the cross-sectional average of the velocity at point x along the vessel and time t , r is the distance from the axial line and $R(x,t)$ is the radius of the vessel. The flow profile parameter k in the expression above is thus essentially unknown. It is usually assumed that $k=2$ which is the classic Poiseuille flow (parabolic profile); higher values of k correspond to flatter profiles.

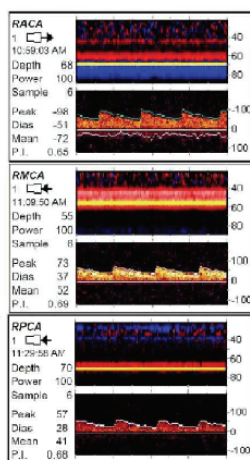


Figure 1: Typical TCD BFV results.

All erythrocytes do not travel at the same speed and the TCD analysis has the great advantage of gathering all spectral info corresponding to the flow, see Figure 1. However in most studies (including the present one) only the values corresponding to the upper envelope of the velocity are kept and averaged over time (leading to v).

Assuming the above profile, the average (in time) maximum (in space) velocity v can be used to express the overall average velocity as $k/(k+2) v$. Consequently, CBF values can be obtained from BFV data through

$$\text{CBF} = k/(k+2) \pi R^2 v/(M \cos\theta) \tag{4}$$

where M the MCA territory mass.

2.2. Lack of correlation across methods

Our results bring to the fore several surprising issues. There is no notable correlation between TCD BFV measurements and MRI (CASL) CBF measurements, as shown in Figure 2.

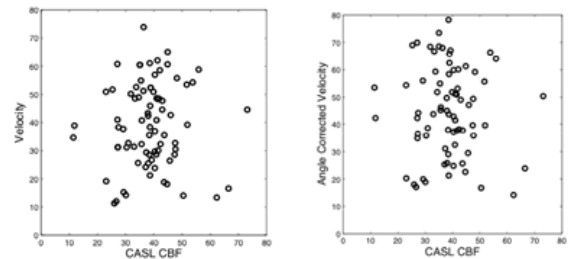


Figure 2: TCD BFV measurements vs. CASL CBF.

Remarkably, correlation does not notably increase if the TCD velocity measurements are corrected by the geometric factor (cosine) linked to the measured insonation angles, see Figure 2, right.

A direct comparison between CASL CBF and TCD CBF is presented in Figure 3.

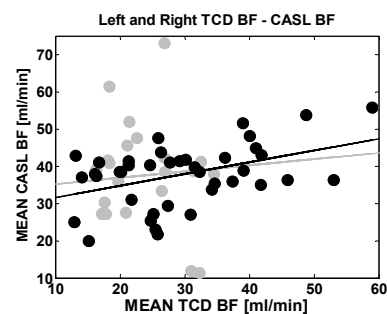


Figure 3: CBF measurement comparison

Correlations are again very low. Further analysis reveals that by splitting left and right side measurements and working only with high quality measurements, correlations between CASL CBF and TCD CBF are found in the .4-.5 range (with $p < .05$), see Figure 3.

Similar conclusions can be reached using Bland-Altman plots.

3. ANALYSIS OF THE SOURCES OF UNCERTAINTY

These low correlations stem from the following possible factors: (i) variability of the factors in the TCD CBF formula, (ii) variability in the TCD BFV measurements themselves, (iii) variability in the CASL CBF measurements. It is worth noting that comparisons of other types of methods lead to similar conclusions see for instance Bookers et al. (2010) for a comparison between ASL MRI and position emission tomography (PET), and systematic overestimating of CBF by ASL. We discuss these aspects below.

3.1. Insonation angle

The insonation angle is not available during routine TCD exam and thus often neglected. Based on 3D-TOEF images, that display both vessels and skull, and with approximate insonation site, these angles can be a posteriori estimated using elementary geometry and insonation depth. Figure 4 displays the values of insonation angle as a function of the measuring depth for a representative group of patients.

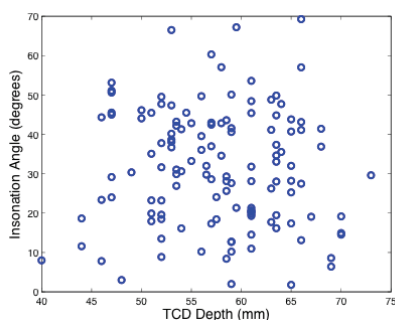


Figure 4: insonation angles vs. measurement depth

The insonation depth is chosen by the operator to optimize signal quality. Two conclusions can be drawn from Figure 4. First, there is no across patient correlation between insonation angle and measurement depth. Second, with an average insonation angle above 30 deg., the geometric factor linked to the insonation angle cannot be ignored.

3.2. Territory mass

The MCA territory mass M is the amount of brain tissue (in units of 100g) that is supplied blood by the two MCAs. The proportion of the brain corresponding to M is unknown and varies from person to person.

We assume the MCA territory mass to be between 35 and 50% of the total brain mass. For our calculations, we split the difference and consider the MCA territory to be 42.5% of the whole brain. Furthermore, we assume that both the left and right MCAs supply equal portions of the brain with blood, so that each artery is responsible for 21.25% of the blood supply to the brain.

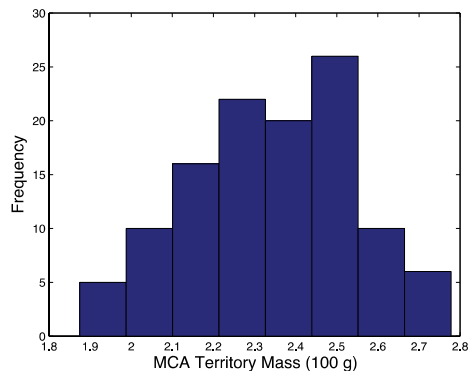


Figure 5: histogram of MCA territory mass

The histogram from Figure 5 suggests that the MCA territory mass roughly follows a normal distribution, with sample mean 2.34 and standard deviation 0.21.

3.3. Uncertainties in velocity data

The value of the velocity profile parameter k is unknown. The TCD protocol assumes that at all times the spatial maximum value of the velocity is obtained. For the assumed velocity profile, the standard deviation in the radial direction is found to be $U/\sqrt{k+1}$.

The measured value v is also assumed to be the temporal average of the velocity. This operation is typically carried internally through proprietary software from the manufacturer of the TCD apparatus. The resulting errors are therefore difficult to assess.

3.4. Resulting uncertainties in TCD CBF

The multiplying factor to the velocity v in the TCD CBF formula, i.e. $k/(k+2) \pi R^2/(M \cos\theta)$, can be estimated. Based on the above assumptions, and for a profile factor $k = 2$, the corresponding distribution is displayed in Figure 6. This gives a first representation of the uncertainties attached to CBF predictions based on TCD.

The large spread of values is a clear indication that, in addition to the velocity data, supplementary patient specific information is needed to guarantee acceptable accuracy.

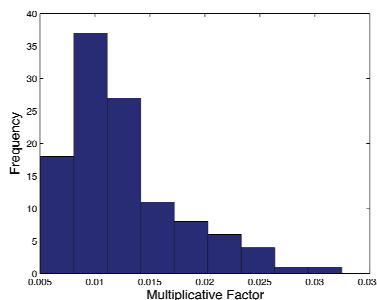


Figure 6: Statistical data for the multiplicative factor of the velocity v in the TCD CBF formula of Section 2.1 (mean is .0124).

3.5. Dependence on hematocrit

The CASL data are also subject to uncertainties.

Blood viscosity, clinically measured as hematocrit, affects not only red cell velocity but also tissue perfusion, labeled blood transfer into the tissue, and iron-induced T1 shortening of magnetization relaxation time (spin-lattice) (Silvoinnen et al., 2003).

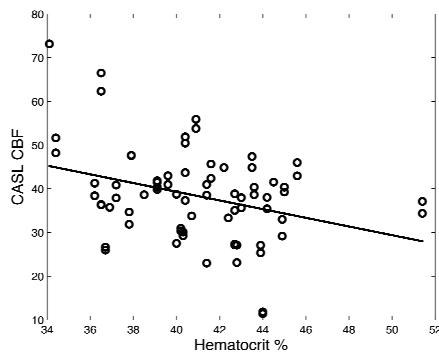


Figure 7: CASL CBF vs. hematocrit %, $r = -0.33$, $p = 0.0034$

As shown in Figure 7, hematocrit is strongly and negatively correlated with CASL estimates and therefore people with lower hematocrit may have higher perfusion estimates. This factor may account for example for gender differences in perfusion and overestimate perfusion in the elderly (Zhao et al., 2009). The longer transit time, as seen with low perfusion states e.g. hypocapnia or small vessel disease in aging, may in contrast lower perfusion estimates, and contribute to differences among vascular territories due to the short decay time of the CASL label (~ 1 s). Therefore, post-labeling delay in conjunction with hematocrit effects may induce under- and over estimation of perfusion values (Thomas 2009).

CASL, fMRI-derived hemodynamics were concordant with $H_2^{15}O$ PET, which is considered a gold standard for CBF imaging (Carroll, et al., 2002),

4. CONCLUSION

TCD and CASL CBF values are only weakly correlated. This is explained by the range of uncertainties attached to parameters entering both protocols: vessel radii, territory mass, insonation angle, velocity profile for TCD and perfusion territory and hematocrit dependence for CASL. Some of the TCD uncertainties can be mitigated. For instance, the analysis of the full spectral information rather than just the envelope may give information about the velocity profile; likewise, insonation angles can be estimated for each patient.

REFERENCES

Bookers, R.P.H., Bremmer, J.P., van Berckel, B.N.M., Lammertsma, A.A., Hendrikse, J., Pluim, J.P.W., Kappelle, L.J., Boellard, R. and Klijn, C.J.M. 2010, Arterial spin labeling perfusion MRI at multiple delay times: a correlative study with $H_2^{15}O$ positron emission tomography in

patients with symptomatic carotid artery occlusion, *J. of Cerebral Blood Flow and Metabolism*, 30, 222-229.

- Carroll, T.J., Jobin, M., Treyer, V., Hany, T.F., Burger, C., Teneggi, V. and Buck A. 2002. Absolute Quantification of Cerebral Blood Flow with MR, Reproducibility of the Method and Comparison with $H_2^{15}O$ PET. *J Cereb Blood Flow Metab*, 22, 1149-1156.
- Castro, M.A., Putman, C.M., Sheridan, M.J. and Cebal, J.R., 2009. Hemodynamics Patterns of Anterior Communicating Artery Aneurysms: A Possible Association with Rupture, *AJNR Am. J. Neuroradiol*, 30, 297-302.
- DeVault, K., Gremaud, P., Novak, V., Olufsen. M.S., Vernieres, G. and Zhao, P., 2008, Blood flow in the Circle of Willis: modeling and calibration, *Multiscale Modeling and Simulation: a SIAM Interdisciplinary Journal*, 27, 888--909.
- Fodale, V., Schifilliti, D., Conti, A., Lucanto, T., Pino, G. and Santamaria, L.B., 2007, Transcranial Doppler and anesthetics, *Acta Anaesthesiol Scand*, 51, 839-847.
- Grinberg, L., Anor, T., Cheever, E., Madsen, J.R. and Karniadakis, G.E., 2009. Simulation of the human intracranial arterial tree, *Phil. Trans. R. Soc.*, 367, 2371-2386.
- Kim, H.J., Figueroa, C.A., Hughes, T.J.R., Jansen K.E. and Taylor, C.A., 2009. Augmented Lagrangian method for constraining the shape of velocity profiles at outlet boundaries for three-dimensional finite element simulations of blood flow, *Comput. Methods Appl. Mech. Engrg.*, 198, 3551-3566.
- Kleinstreuer, C., Li, Z. and Farber, M. 2007. Fluid-structure interaction analyses of stented abdominal aortic aneurysms, *Annual Reviews in Biomedical Engineering*, 9, 169-204.
- Silvoinnen M.J., Kettunen M.I., Kauppinen R.A., 2003. Effects of hematocrit and oxygen saturation level on blood spin-lattice relaxation. *Magn. Res. Med.*, 49, 586-587.
- Thomas, D.L., Lythgoe, M.F., van der Weerd, L., Ordidge, R.J., Gadian, D.G., 2006. Regional variation of cerebral blood flow and arterial transit time in the normal and hypoperfused rat brain measured using continuous arterial spin labeling. *J.Cereb. Blood Flow Metab.* 26, 274-282.
- Watton, P.N., Raberger, N.B., Holzapfel, G.A. and Ventikos, Y. 2009. Coupling the Hemodynamic Environment to the Evolution of Cerebral Aneurysms: Computational Framework and Numerical Examples, *J. Biomech. Eng.*, 131 (2009), pp. 101003-1-101003-14.
- Zhao, P., Alsop, D.C., Abduljalil A., Selim M, Lipsitz, L., Novak P., Capln L., Hu K., Novak V., 2009. Vasoreactivity and periinfarct hyperintensities in stroke. *Neurology*, 72, 643-649.

EXPERIMENTAL AND NUMERICAL STUDIES OF DIGITAL ARTERIAL ELASTICITY BY VOLUME OSCILLOMETRIC ANALYSIS

Pichitra Uangpairoj^(a), Masahiro Shibata^(a)

^(a)Department of Bio-science and Engineering, College of Systems Engineering and Science, Shibaura Institute of Technology, Japan

^(a) m710506@shibaura-it.ac.jp, ^(b) shibatam@sic.shibaura-it.ac.jp

ABSTRACT

Arterial elasticity is one of the indicators for the existence of cardiovascular diseases. The present study introduces the elasticity evaluation of digital artery by using a volume oscillometric technique. We conducted the experiment by analyzing an alternating signal of photoelectric plethysmograph during the continuous change of transmural pressure and resulting in the form of new pressure-volume relationship. In addition, the theoretical model based on the continuum mechanics and hyperelasticity of arterial wall was also simulated under the passive tension to predict arterial elasticity and clarify the experimental analysis. As the results of the experimental approach, the new pressure-volume relationship provided the sensible response of arterial wall to the transmural pressure. The change of this relationship associated with the age of subjects. While, the theoretical results demonstrated the ability of current model for the prediction of arterial elasticity using the volume oscillometric technique under the influence of pulse pressure.

Keywords: arterial elasticity, volume oscillometry, photoelectric plethysmograph

1. INTRODUCTION

The stiffening of arterial wall is caused by several factors such as the endothelial cell dysfunction, atheroma, elastin fragmentation, calcium deposition, etc (Lee and Oh 2010). This condition relates to the existence and development of cardiovascular diseases (CVD) (Vlachopoulos, Aznaouridis and Stefanadis 2010). One of the arterial properties which is regarded to be a CVD marker is arterial elasticity. Arterial elasticity mainly associates with the mechanical properties of elastin and collagen fibers in arterial wall. It is basically investigated by monitoring the arterial motion in circumferential direction. The instantaneous change of vessel circumference correlates with arterial pressure pulse in the form of pressure-diameter relationship or pressure-volume relationship. Using isolated arteries in vitro test can obtain those relationships by varying intravascular pressure (Carew, Vaishnav and Petal 1968). Moreover, several noninvasive methods have been proposed to evaluate

arterial elasticity. For example, pulse pressure from sphygmomanometer (Franklin et al. 1997), pulse wave velocity from Doppler flow signals (Sutton-Tyrrell et al. 2005), indices derived from magnetic resonance imaging, MRI, and ultrasound imaging techniques and an augmentation index from applanation tonometry (Fantin et al. 2007). Another technique whose system is low-cost and easy to use is photoelectric plethysmography (PPG).

The system of PPG contains two essential components, an infrared light source and a photo sensor. The infrared light is absorbed by skin, bone, tissue and blood while the remaining light is detected at the photo sensor before transforming it into two signal components by the signal processing. The first signal is a steady signal, DC signal, which represents the absorption capability of skin, bone, tissue, venous blood and non-pulsatile arterial blood. Another is the alternating signal, AC signal, which represents the absorption due to pulsatile arterial blood. It also associates with the oscillation of vascular volume of arteries. The technique of using PPG for vascular elasticity analysis is conducted under the assumptions that surrounding tissues around vascular system are incompressible. The cuff pressure of the occluding cuff collapses the venous system. The continuous change of cuff pressure controls the vascular volume of arteries. Many researchers have utilized the DC signal of PPG coupled with the Lambert-Beer's Law to analyze arterial elasticity in the form of transmural pressure-relative volume change relationship (Ando et al. 1991, Kawarada et al. 1986). However, a conflict between the assumptions of their method and the actual DC signal has occurred. When cuff pressure is higher than systolic blood pressure, arterial blood flow disappears, then DC signal should be constant. However, the actual DC signal is not constant as in the assumptions. This phenomenon implies the influences of arterial surrounding tissues on the intensity of DC signal. In this study, therefore, we propose an alternative methodology to evaluate arterial elasticity using the AC signal instead of DC signal.

Furthermore, mathematical models derived from continuum mechanics and empirical data have been developed to describe the mechanical response of

arterial wall (Kalita and Schaefer 2008). In this study, theoretical method based on the continuum mechanics of arterial wall is also employed to model and predict arterial elasticity in various conditions, especially the different level of pulse pressure. The theoretical results are expected to validate and clarify the assessment of arterial elasticity using the AC signal of PPG.

2. MATERIAL AND METHOD

2.1. Experimental Approach

A newly developed PPG system in which contains the infrared light source and the photo sensor was used to obtain AC, DC and cuff pressure, P_c , signals from the target site. These signals were transformed to be digital outputs with a 24-bit A/D converter of PowerLab4/26 (ML846, ADInstruments, Bella Vista, NSW, Australia). Figure 1 shows a typical example of PPG signals obtained from human finger. At the first appearance of AC signal, P_c is equal to systolic blood pressure, P_{sys} . Meanwhile, the point where the amplitude of AC signal is maximum indicates the equality between P_c and mean blood pressure, P_m . Transmural pressure, P_{tr} , which means the subtraction of P_c from P_m is zero at this stage. Thus, we can obtain P_m and P_{sys} from the connection between the AC signal and P_c of PPG.

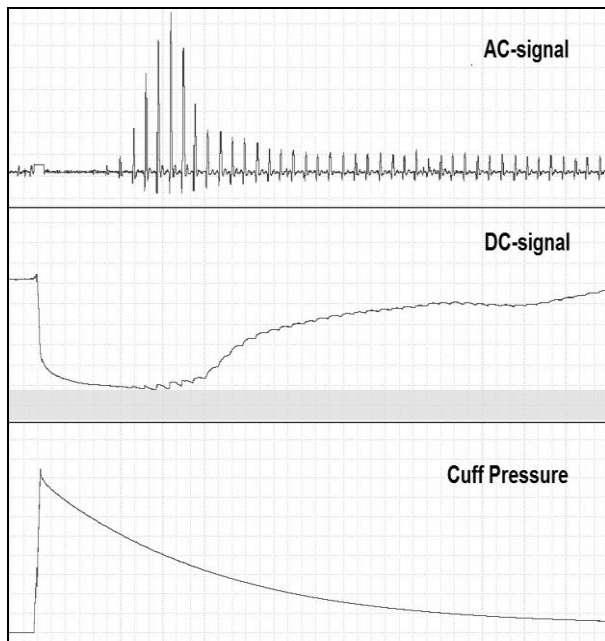


Figure1: The Signals Of Photoelectric Plethysmograph (AC, DC And P_c Signals)

Moreover, when P_c is equal or lower than P_m , the amplitude of AC signal reflects the volume difference between systole and diastole, ΔV . We therefore proposed the relative volume difference as a new index of vascular elasticity. The relative volume difference was defined as the volume difference at any P_{tr} normalized with that as zero P_{tr} , $\Delta V/\Delta V_0$. Then, we obtained the new pressure-volume relationship from the AC component of PPG signals in the form of P_{tr} – $\Delta V/\Delta V_0$ relationship.

We recorded the PPG signals from the right index finger of 40 subjects who were separated into six groups of subjects, eight females and eight males; 20-25 years old, five females and five males; 32-45 years old, six females and eight males; over 50 years old, at a temperature about 24 ± 1 °C. After that, the output AC and P_c signals were used to calculate P_{sys} , P_m , P_{tr} , diastolic blood pressure, P_{dia} , pulse pressure, ΔP , and $\Delta V/\Delta V_0$ in order to obtain the relationship between P_{tr} and $\Delta V/\Delta V_0$.

Two way analysis of variance, Two-way ANOVA, was utilized to estimate the differences of P_m and ΔP among six groups of subjects. The P_m and ΔP were exhibited as mean values with standard deviation, SD. While, the relationships between P_{tr} and $\Delta V/\Delta V_0$ of each group were shown as the mean curves.

2.2. Theoretical Approach

The theoretical approach of this study focused on the mechanical response of arterial wall due to pulsatile internal pressure. We combined the three layers of radial index arterial wall into one layer to simplify anatomical structure of arterial tube. The outer diameter of the tube was 1.54 mm (Bilge et al. 2006). The ratio of total wall thickness to outer diameter was 0.05. The initial length of the tube was 10 mm. This arterial tube confronted with pulsatile pressures which were given in the form of sinusoidal function to simplify the arterial pulsatile pressure as follows:

$$P(t) = P_m + P_{amp} \sin(2\pi ft) \quad (1)$$

where $P(t)$ is the instantaneous pressure (mmHg), P_m is the mean level of pulsatile pressure (mmHg), P_{amp} is the amplitude of pulsatile pressure (mmHg) which is the half of constant ΔP , 60 mmHg, f is frequency (Hz) which associates with normal heart rate, 75 beats per minute and t is time (s). The pulsatile pressure was applied at the inner wall of the arterial tube. The outside pressure was assumed to be zero.

Arterial wall stress due to pulsatile pressure was calculated from the Laplace's Law. It provides the relationship between internal pressure and circumferential Cauchy stress as follows:

$$\sigma_c = (P_i - P_o)r/\omega \quad (2)$$

where σ_c is circumferential Cauchy stress (KPa), P_i is the internal pressure (KPa), P_o is the outside pressure (KPa), r is the radius of arterial tube (mm) and ω is the wall thickness of arterial tube (mm). The Cauchy stress tensor, σ , related to the second Piola-Kirchhoff stress tensor, S , by the following relationship:

$$\mathbf{S} = \mathbf{J}\mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^T \quad (3)$$

where F is the deformation gradient and J is the determinant of deformation gradient. In this study, arterial wall was considered to be hyperelastic material

whose the relationship between stress and strain was derived in term of strain energy function, W , as follows:

$$S = \partial W / \partial E \quad (4)$$

where E is Green-Lagrange strain tensor. While the strain energy function of arterial wall was described by Fung model (Chuong and Fung 1983, Deng et al. 1994) as follows:

$$W = (C/2)\exp(Q) \quad (5)$$

$$Q = b_1 E_{\theta\theta}^2 + b_2 E_{zz}^2 + b_3 E_{rr}^2 + 2b_4 E_{\theta\theta} E_{zz} + 2b_5 E_{\theta\theta} E_{rr} + 2b_6 E_{zz} E_{rr} + b_7 (E_{r\theta}^2 + E_{\theta r}^2) + b_8 (E_{\theta z}^2 + E_{z\theta}^2) + b_9 (E_{rz}^2 + E_{zr}^2) \quad (6)$$

where C and $b_1 - b_9$ are material parameters. We firstly simplified the problem by neglecting shear stress and assuming arterial wall to be a very thin wall. Thus,

$$Q = b_1 E_{\theta\theta}^2 + b_2 E_{zz}^2 + 2b_4 E_{\theta\theta} E_{zz} \quad (7)$$

Material parameters, C , b_1 , b_2 and b_4 , of each group of subjects were obtained from the optimization by comparing with the $P_{tr} - \Delta V/\Delta V_0$ relationships of PPG measurement using the genetic algorithm. The theoretical results were also expressed as the $P_{tr} - \Delta V/\Delta V_0$ relationships.

Furthermore, the arterial model with the optimum material parameters was simulated under the different level of ΔP , 40, 50 and 60 mmHg, and resulted in the form of $P_{tr} - \Delta V/\Delta V_0$ relationships.

3. RESULTS AND DISCUSSION

3.1. Experimental Results

As the results of PPG measurement, the differences of P_m were found among six groups of subjects as in Figure 2. From the statistical analysis, over 50 years old subjects had the highest P_m in the group of women. The P_m of women tended to decrease when the age of subject decreased. These results corresponded to the study of Pearson et al. (1997). The increase in age induced the rises in systolic and diastolic pressure, especially in the range of 20 - 50 years old. However, the same level of P_m was seen in the different age of men. Moreover, the influence of gender on P_m was found in the 20-25 years old and over 50 years old subjects. Meanwhile, six groups of subjects had the similar level of ΔP as in Figure 2.

Figure 3. illustrates the relationships between P_{tr} and $\Delta V/\Delta V_0$ of six groups of subjects. The results showed that $\Delta V/\Delta V_0$ nonlinearly declined when P_{tr} increased. This informed that arteries were stiffer with the advancing P_{tr} . The groups of 20-25 years old and 32-45 years old subjects had the same characteristic of $P_{tr} - \Delta V/\Delta V_0$ curves, while the $P_{tr} - \Delta V/\Delta V_0$ curves of over 50 years old subjects explicitly differed from those curves of 20-25 years old and 32-45 years old subjects. The higher level of $\Delta V/\Delta V_0$ values at any P_{tr} of over 50

years old subjects, compared with those of 20-25 years old and 32-45 years old subjects, implied the lower elasticity of arterial blood vessel in human finger of the older subjects, especially subjects whose age beyond 50 years old. These results agreed with other methods in the study of Jayasree, Sandhya and Radhakrishnan (2008) and Kelly et al. (1989). Therefore, the $P_{tr} - \Delta V/\Delta V_0$ relationship would be one of the arterial elasticity markers as well. In addition, men and women had the same shape of $P_{tr} - \Delta V/\Delta V_0$ curves in all three groups of age. These results reflected that the different P_m between men and women, especially in 20-25 years old and over 50 years old subjects, did not influenced on the $P_{tr} - \Delta V/\Delta V_0$ relationship. Thus, using the $P_{tr} - \Delta V/\Delta V_0$ relationship would eliminate the effect of P_m on the arterial elasticity evaluation by using the AC signal of PPG.

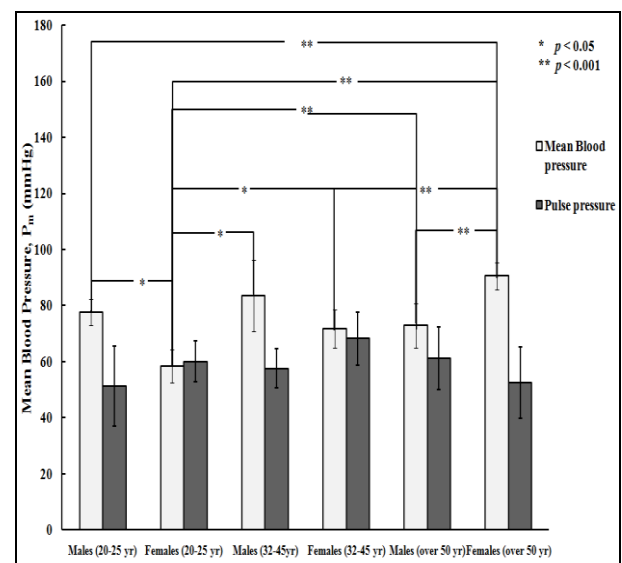


Figure 2: Mean Blood Pressure (White) And Pulse Pressure (Black) Of Six Groups Of Subjects From PPG Measurement

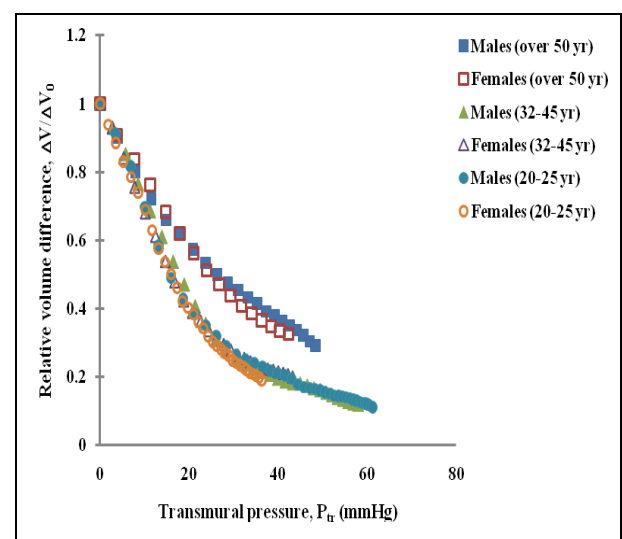


Figure 3: The Relationship Between Transmural Pressure And Relative Volume Difference From PPG

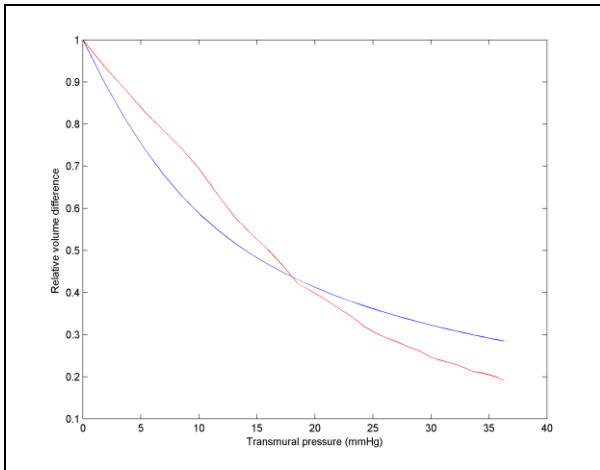


Figure 4: The Comparison Of $P_{tr} - \Delta V/\Delta V_0$ Relationship Between Experimental (Red Line) And Theoretical (Blue Line) Results Of 20-25 Years Old Women.

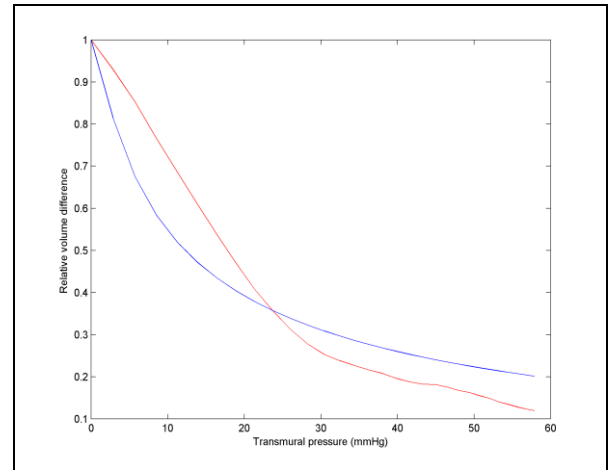


Figure 7: The Comparison Of $P_{tr} - \Delta V/\Delta V_0$ Relationship Between Experimental (Red Line) And Theoretical (Blue Line) Results Of 32-45 Years Old Men.

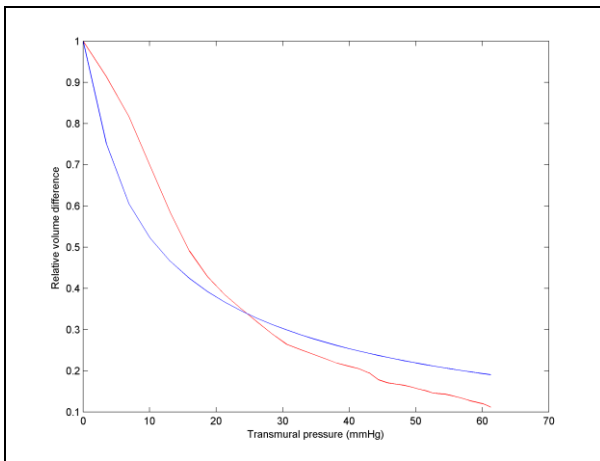


Figure 5: The Comparison Of $P_{tr} - \Delta V/\Delta V_0$ Relationship Between Experimental (Red Line) And Theoretical (Blue Line) Results Of 20-25 Years Old Men.

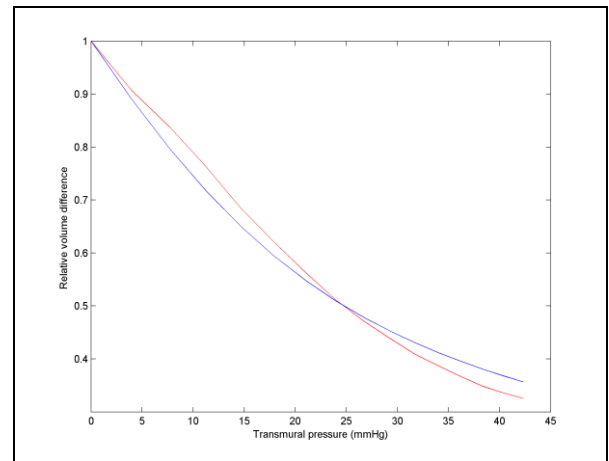


Figure 8: The Comparison Of $P_{tr} - \Delta V/\Delta V_0$ Relationship Between Experimental (Red Line) And Theoretical (Blue Line) Results Of Over 50 Years Old Women.

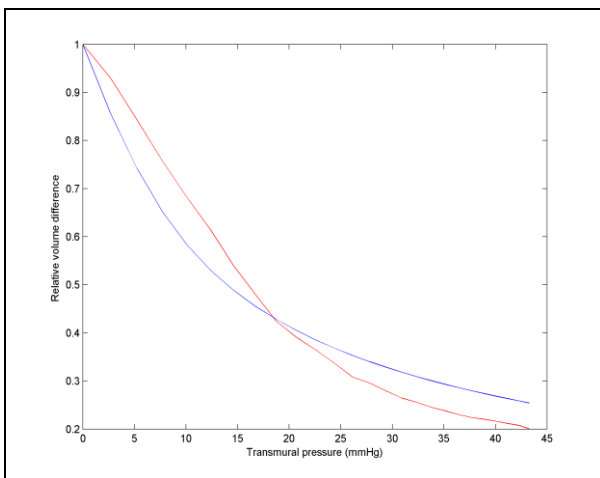


Figure 6: The Comparison Of $P_{tr} - \Delta V/\Delta V_0$ Relationship Between Experimental (Red Line) And Theoretical (Blue Line) Results Of 32-45 Years Old Women.

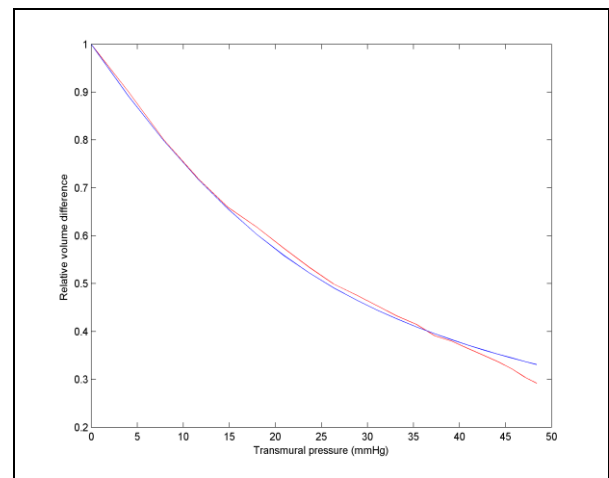


Figure 9: The Comparison Of $P_{tr} - \Delta V/\Delta V_0$ Relationship Between Experimental (Red Line) And Theoretical (Blue Line) Results Of Over 50 Years Old Men.

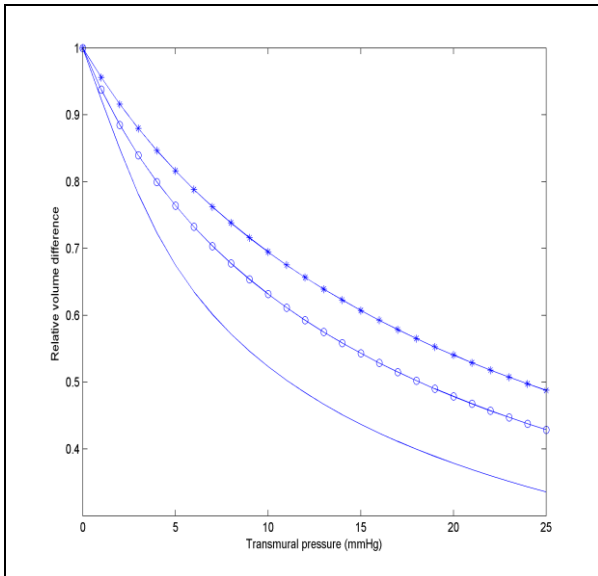


Figure 10: The $P_{tr} - \Delta V/\Delta V_0$ Relationships Of 20-25 Years Old Men With Different Level Of ΔP , 40 (*),50 (o) And 60 (-) mmHg.

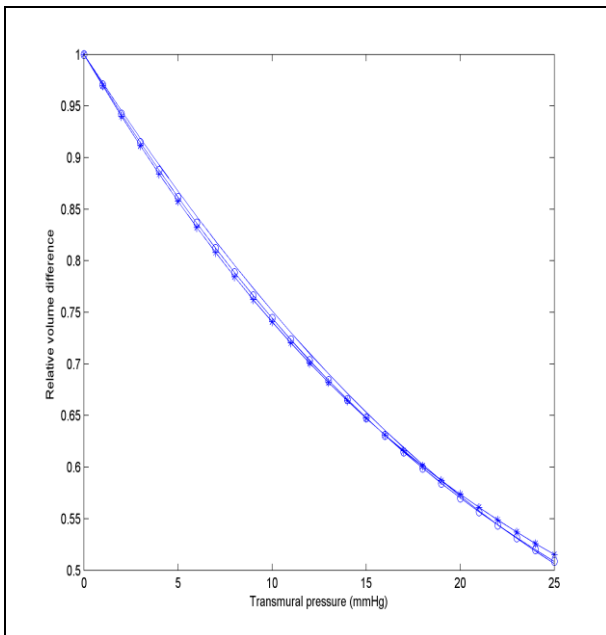


Figure 10: The $P_{tr} - \Delta V/\Delta V_0$ Relationships Of Over 50 Years Old Men With Different Level Of ΔP , 40 (*),50 (o) And 60 (-) mmHg.

3.2. Theoretical Results

According to theoretical approach, the relationships between P_{tr} and $\Delta V/\Delta V_0$ of six groups of subjects are shown in Figure 4-9. These results were obtained from the optimum material parameters of individual group. They showed that $\Delta V/\Delta V_0$ decreased when P_{tr} increased in all groups of subjects as same as the experimental results. These results indicated that Fung model which described the hyperelasticity of arterial wall was able to provide the reasonable response of arterial wall. It also generated the similar trend of $P_{tr} - \Delta V/\Delta V_0$ relationship as the experimental results from PPG measurement. Although there were some differences between

experimental and theoretical results in the groups of 20-25 years old and 32-45 years old subjects, theoretical results had the good fit curves with the experimental results in over 50 years old subjects (Figure 8 and 9). These results demonstrated that the current theoretical model was more efficient to predict the elasticity of stiff artery than that of flexible artery. Meanwhile, the difference between experimental and theoretical results might be caused by the limitation of 2D computation and the exponential term of the Fung model.

Figure 10 presents the example result of theoretical model under the influence of ΔP . The difference of $P_{tr} - \Delta V/\Delta V_0$ relationships in 20-25 years old men was found when the level of ΔP was changed. This relationship shifted downwards when the level of ΔP increased. However the effect of ΔP was difficult to see in older subjects as the example of $P_{tr} - \Delta V/\Delta V_0$ relationships in over 50 years old men in Figure 11. Because the blood vessel of older subjects had lower elasticity, the small magnitude of ΔP change could not induce the much different response of the stiff wall. We therefore suggest that the level of ΔP mainly affects the arterial elasticity evaluation in the flexible artery.

4. CONCLUSION

The novel method to quantify arterial elasticity using the AC signal of PPG and the first step of theoretical approach based on the continuum mechanics of arterial wall were proposed in this paper. According to using AC signal of PPG measurement, we obtained the $P_{tr} - \Delta V/\Delta V_0$ relationship which presented the reasonable response of arterial wall to the continuous change of P_{tr} . This $P_{tr} - \Delta V/\Delta V_0$ relationship could reflect the variation of arterial elasticity due to the increase in age of subjects. Moreover, the effect of P_m would be negligible when we analyzed arterial elasticity by this novel method. As the results of theoretical approach, the current theoretical model provided the response of arterial wall to the pressure similar to natural artery. This theoretical model also clarified the effect of pulse pressure on the $P_{tr} - \Delta V/\Delta V_0$ relationship that should be regarded in the analysis of arterial elasticity by using AC signal of PPG, especially in the flexible artery. However, the theoretical model will be improved in order to increase the efficiency of arterial elasticity prediction and evaluation by using the volume oscillometric technique

ACKNOWLEDGMENTS

The author is grateful to Prof. Masahiro Shibata at Shibaura Institute of Technology for his advice on performing this research. She would like to thank all subjects for their sacrifice in the experiment. She would also like to thank Mr. Bhanupong Petchlert for his advice on preparing theoretical computation. Finally, this study was supported by the grant of Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

- Ando, J., Kawarada, A., Shibata, M., Yamakoshi, K. and Kamiya, A., 1991. Pressure-volume relationships of finger arteries in healthy subjects and patients with coronary atherosclerosis measured non-invasively by photoelectric plethysmography. *Japanese Circulation Journal*, 55, 567-575.
- Bilge, O., Pinar, Y., O'zer, M. A. and Go'vsa, F., 2006. A morphometric study on the superficial palmer arch of the hand. *Surgical and Radiologic Anatomy*, 28, 343-350.
- Carew, T. E., Vaishnav, R. N. and Patel, D. J., 1968. Compressibility of the arterial wall. *Circulation Research*, 23, 61-68.
- Chuong, C. J. and Fung, Y. C., 1983. Three-dimensional stress distribution in arteries. *ASME Journal of Biomechanical Engineering*, 105, 268-274.
- Deng, S. X., Tomioka, J., Debes, J. C. and Fung, Y. C., 1994. New experiments on shear modulus of elasticity of arteries. *American Journal of Physiology*, 266, H1-H10.
- Fantin, F., Mattocks, A., Bulpitt, C. J., Banya, W. and Rajkumar, C., 2007. Is augmentation index a good measure of vascular stiffness in the elderly?. *Age and Aging*, 36, 43-48.
- Franklin, S. S., Gustin, W., Wong, N. D., Larson, M. G., Weber, M. A., Kannel, W. B., and Levy, D., 1997. Hemodynamic patterns of age-related changes in blood pressure. *Circulation*, 96, 308-315.
- Jayasree, V. K., Sandhya, T.V. and Radhakrishnan, P., 2008. Non-invasive studies on age related parameters using a blood volume pulse sensor. *Measurement Science Review*, 8, 82-86.
- Lee, H. and Oh, B., 2010. Aging and arterial stiffness. *Circulation Journal*, 74, 2257-2262.
- Kalita, P. and Schaefer, R., 2008. Mechanical model of artery walls. *Archives of Computational Methods in Engineering*, 15, 1-36.
- Kawarada, A., Shimazu, H., Yamakoshi, K. and Kamiya, A., 1986. Noninvasive automatic measurement of arterial elasticity in human fingers and rabbit forelegs using photoelectric plethysmography. *Medical & Biological Engineering & Computing*, 24, 591-596.
- Kelly, R., Hayward, C., Avolio, A. and O'Rourke, M., 1989. Noninvasive determination of age-related changes in the human arterial pulse. *Circulation*, 80, 1652-1659.
- Pearson, J. D., Morrell, C. H., Brant, L. J., Landis, P. K. and Fleg, J. L., 1997. Age-associated changes in blood pressure in a longitudinal study of healthy men and women, *Journal of Gerontology: MEDICAL SCIENCES*, 52A, M177-M183.
- Sutton-Tyrrell, K., Najjar, S. S., Boudreau, R. M., Venkitachalam, L., Kupelian, V., Simonsick, E. M., Havlik, R., Lakatta, E. G., Spurgeon, H., Kritchevsky, S., Pahor, M., Bauer, D. and Newman, A., 2005. Elevated aortic pulse wave velocity, a marker of arterial stiffness, predicts cardiovascular events in well-functioning older adults. *Circulation*, 111 (25), 3384-3390.
- Vlachopoulos, C., Aznaouridis, K. and Stefanadis, C., 2010. Prediction of cardiovascular events and all-cause mortality with arterial stiffness: a systematic review and meta-analysis. *Journal of the American College of Cardiology*, 55, 1318-1327.

AUTHORS BIOGRAPHY

Pichitra Uangpairoj received the B.Sc. (2007) in Food Technology and M.E. (2010) in Mechanical Engineering from Suranaree University of Technology, Thailand. She is the PhD student in the Department of Bioscience and Engineering, Shibaura Institute of Technology, Japan. Her current interests include the applications of numerical simulation in arteries.

Masahiro Shibata received the BS in Applied Physics and his PhD in Biomedical Engineering from Hokkaido University, Japan. Since 2008 he has been with the Department of Bio-Science and Engineering, Shibaura Institute of Technology, where he is a Professor of System Physiology. His research interests include oxygen dynamics in microcirculation.

BAROREFLEX SENSITIVITY DURING THE GRAVITATIONAL STIMULUS: PHYSIOLOGY AND PATHOPHYSIOLOGY

Raffaello Furlan^(a), Franca Dipaola^(b), Veronica Pacetti^(c), Carlo Selmi^(d), Francesca Meda^(e), Iaria Bianchi^(f), Franca Barbic^(g).

^(a,b,c,d,e,f,g) Internal Medicine, Humanitas Clinical and Research Center, Rozzano,
^(a,d) BIOMETRA Department, University of Milan, Milan, Italy

^(a) raffaello.furlan@unimi.it
^(b) franca.dipaola@humanitas.it
^(c) v.pacetti@libero.it
^(d) carlo.selmi@unimi.it
^(e) francesca.meda@humanitas.it
^(f) ilaria.bianchi@humanitas.it
^(g) franca.barbic@humanitas.it

ABSTRACT

Baroreceptor mechanisms play a crucial role in healthy humans during a gravitational stimulus to maintain blood pressure and cerebral perfusion by providing proper orthostatic tolerance. In addition, they concur to synchronize central neural discharge activity and hemodynamic spontaneous fluctuations leading to an optimal interaction between the cardiovascular neural regulatory activity and the heart and vessel response.

In pathophysiological conditions as encountered before tilt induced neurally-mediated syncope, the gain of baroreflex control of heart rate and muscle sympathetic nerve activity (MSNA) is remarkably impaired, revealing a diminished capability to adapt in response to similar changes of blood pressure compared to healthy subjects. The loss of an organized post-ganglionic sympathetic discharge activity to the vessel leads to orthostatic intolerance in subjects with baroreceptor failure.

Key words: baroreflex sensitivity, muscle sympathetic nerve activity, cardiovascular sympathetic activity, gravitational stimulus

1. PHYSIOLOGY

In healthy humans the change from rest to an orthostatic posture is followed by a blood pooling of about 800 ml into the high capacitance venous district below the heart leading to a drop of central venous pressure, a reduction of cardiac output and ultimately systemic hypotension (Mosqueda-Garcia 2000). The concomitant arterial baroreceptor and cardiopulmonary afferents unloading initiates a reflex neural sympathetic excitation and a concomitant vagal withdrawal that lead to arterial and venous vasoconstriction, an increase of heart rate and cardiac inotropism aimed at maintaining blood pressure and cerebral perfusion (Mosqueda-Garcia 2000). In addition, the overall increase of sympathetic activity is also sustained by positive feedback mechanisms induced by sympathetic afferents stimulation

generated by blood pooling in the low-pressure district attending the gravitational stress (Malliani 2000).

From a hemodynamic standpoint, during a 75° head-up tilt healthy humans manifest a mild increase of diastolic arterial pressure with a marked enhancement of heart rate, plasma norepinephrine levels and neural sympathetic discharge (muscle sympathetic nerve activity, MSNA), without major changes in systolic arterial pressure (Furlan 2000). Notably, beside changes in mean value under these conditions, every single variable undergoes spontaneous fluctuations. In particular, the discharge activity of the sympathetic fibres is grouped after an oscillatory pattern characterized by a period of 10 seconds and is coupled with analogous rhythmic fluctuations of heart rate and blood pressure (Furlan 2000).

The quantification of the observed oscillatory phenomena is obtained by power spectrum analysis as 0.1 Hz (low frequency, LF) oscillatory component (Task Force ESC and NASPE 1996, Malliani 2000, Pagani 1986). During the sympathetic activation induced by tilt, the LF component of muscle sympathetic nerve activity (MSNA) variability is prevalent in the spectrum, resembling the changes observed in the same oscillatory components of RR and systolic arterial pressure variability (Furlan 2000). In addition, the linear coupling between LF fluctuation of MSNA and heart rate, and of MSNA and systolic arterial pressure, assessed by the coherence function, was found to increase during the gravitational stimulus suggesting the onset of a common oscillatory pattern based on prevailing 0.1 Hz fluctuation (Furlan 2000).

It must be pointed out that sympathetic neurons located in areas within the central nervous system involved in cardiovascular regulation were found to be characterized by a 0.1 Hz rhythmical discharge activity linked to low frequency spontaneous fluctuations of RR interval and blood pressure (Kaminski 1970, Montano 1996, Preiss and Polosa 1974). Interestingly, these fluctuations persist following the removal of baroreflex afferents by a stabilizer device connected to the arterial system of the

animal or after bilateral vagal denervation in spinal dogs (Kaminski 1970, Preiss and Polosa 1974).

In patients with intractable heart failure, the suppression of spontaneous blood pressure fluctuations by a left ventricle assist device is associated with a persistent LF component of RR variability (Cooley 1998). These observations cumulatively suggest that baroreflex activity is not necessary to the *genesis* of LF fluctuations that seem to be the result of the activity of a central oscillator. However, the inhibitory role of human baroreflex activity on tonic MSNA is established and may concur to organize the sympathetic neural activity after a discharge pattern with a period of 10 seconds (i.e. low frequency fluctuation). Analogous phenomena may arise in heart rate spontaneous variability during both increase and decrease of arterial blood pressure as the result of baroreflex modulation of heart period (Pagani 1997). Indeed, a blood pressure decrease is associated with the increase of heart rate and sympathetic nerve discharge (Furlan 1998a). Conversely, arterial pressure enhancement is attended by neural silence and a drop of heart rate (Furlan 1998a).

Therefore, baroreflex mechanisms during tilt are more likely to play a role in *modulating* the amplitude and in *synchronizing* the LF fluctuations of heart rate, arterial pressure and MSNA (Furlan 2000). In keeping with this hypothesis is the observation of the absence of a LF component in both RR and systolic arterial pressure variability in subjects with pure autonomic failure (Furlan 1995) during a 45° head-up orthostatic challenge. In addition, we observed a remarkable orthostatic intolerance and hypotension in spite of high level of plasma catecholamine and MSNA in a patient with a baroreflex failure because of a rhinopharyngeal cancer who underwent neck radiotherapy (Furlan 2001). Notably, the post-ganglionic sympathetic discharge activity was completely disorganized suggesting that a proper neural frequency modulation (0.1 Hz) is required to induce a suitable vasoconstriction as already suggested in animal studies (Stauss 1998).

2. PATHOPHYSIOLOGY

In physiological conditions, the increase of the LF component of RR variability (LF_{RR} , in normalized units) during a gravitational stimulus tends to be constant over time. Indeed, in a previous study, based on time variant spectrum analysis of RR variability, which enabled us to follow the beat by beat change in LF_{RR} and HF_{RR} amplitude, the two spectral components and RR interval were relatively stable during a 15 minute lasting tilt procedure in healthy subjects (Furlan 1998b). Conversely, in the pathophysiological condition of a neurally mediated syncope, we observed that the progression of tilt was often characterized by wide fluctuations of LF_{RR} and HF_{RR} , suggestive of a marked instability of the cardiac autonomic control up to the onset of syncope (Furlan 1998b). Interestingly, that pattern occurred without concomitant changes in mean RR value, which remained stable. This underlies the discordance

between modifications in the oscillatory components of heart rate variability and the absence of concordant changes in heart rate (Furlan 1998b). In general terms, heart rate results from pacemaker intrinsic discharge, sympathetic and vagal neural modulation and circulatory neurohormones (Malliani 2000). Conversely, heart rate variability reflects the autonomic modulation of the sino-atrial node activity. Therefore, heart rate and its variability are separated entities and the former may not follow the changes of the latter (i.e. of the spectral components of RR variability). In this context the analysis of heart rate variability may furnish additional information on the state of the cardiac neural control which might have remained hidden if considering the simple changes of heart rate.

In a different group of subjects with analogous clinical problems, that is a single spell of syncope during a tilt test without a clinical history of loss of consciousness, the gain of baroreflex control of heart rate and MSNA computed during a 15° step-wise head-up tilt, revealed a diminished capability to increase heart rate and MSNA in response to similar changes of blood pressure compared with controls (Mosqueda-Garcia 1997). In these subjects MSNA was abolished before syncope and this suggests the existence of a remarkable impairment of baroreflex control of heart rate and MSNA in the period preceding syncope (Mosqueda-Garcia 1997).

Preliminary experiments within the context of a European Space Agency research project have recently provided somehow different results. In healthy volunteers who underwent -6° head-down bed rest for three weeks, the upright posture was associated, as expected, with an initial remarkable increase of the MSNA as assessed both in bursts per minute and in MSNA total activity, in heart rate whereas blood pressure was maintained. In addition, the discharge pattern of the post-ganglionic sympathetic fibres was grouped after a clear 0.1 Hz rhythm. However, in the subset of subjects who fainted, the minute preceding the loss of consciousness was characterized by a disruption of such rhythm. Thus, in spite of the fact that the sympathetic drive to the vessel was still present until the syncope, blood pressure progressively declined and heart rate was either maintained or dropped near zero. This observation confirms that the period preceding a syncope is characterized by an overall baroreceptor activity alteration. Importantly, it also points out that the presence of an organized sympathetic discharge at 0.1 Hz rhythm is mandatory in order to keep blood pressure and heart rate adequate during the gravitational stimulus.

REFERENCES

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: 1996, Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation*, 93, 1043-1065.

- Cooley, R.L., Montano, N., Cogliati, C., van de Borne, B., Richenbacher, W., Oren, R., Somers, V.K., 1998. Evidence for a central origin of the low-frequency oscillation in RR-interval variability. *Circulation*, 98, 556-561.
- Furlan, R., Jacob, G., Snell, M., Robertson, D., Porta, A., Harris, P., Mosqueda-Garcia, R., 1998a. Chronic orthostatic intolerance: a disorder with discordant cardiac and vascular sympathetic control. *Circulation*, 98, 2154-2159.
- Furlan, R., Piazza, S., Dell'Orto, S., Barbic, F., Bianchi, A., Mainardi, L., Cerutti, S., Pagani, M., Malliani, A., 1998b. Cardiac autonomic patterns preceding occasional vasovagal reactions in healthy humans. *Circulation*, 98, 1756-1761.
- Furlan, R., Porta, A., Costa, F., Tank, J., Baker, L., Schiavi, L., Robertson, D., Malliani, A., Mosqueda-Garcia, R., 2000. Oscillatory patterns in sympathetic neural discharge and cardiovascular variables during orthostatic stimulus. *Circulation*, 101, 886-892.
- Furlan, R., Piazza, S., Bevilacqua, M., Turiel, M., Norbiato, G., Lombardi, F., Malliani, A., 1995. Pure Autonomic Failure: complex abnormalities in the neural mechanisms regulating the cardiovascular system. *Journal of Autonomic Nervous System*, 51, 223-235.
- Furlan, R., Magatelli, R., Palazzol, L., Rimoldi, A., Colombo, S., Porta, A., 2001. Orthostatic intolerance: different abnormalities in the neural sympathetic response to a gravitational stimulus. *Autonomic Neuroscience: Basic and Clinical* 90, 83-88.
- Kaminski, R.J., Meyer, G.A., Winter, D.L., 1970. Sympathetic unit activity associated with Mayer waves in the spinal dog. *American Journal of Physiology*, 219, 1768-1771.
- Malliani, A., 2000. *Principles of cardiovascular neural regulation in health and disease*. Boston: Kluwer Academic Publishers.
- Montano, N., Gneccchi-Ruscione, T., Porta, A., Lombardi, F., Malliani, A., Barman, S.M., 1996. Presence of vasomotor and respiratory rhythms in the discharge of single medullary neurons involved in the regulation of cardiovascular system. *Journal of Autonomic Nervous System*, 57, 116-122.
- Mosqueda-Garcia, R., Furlan, R., Fernandez-Violante, R., Desai, T., Snell, M., Jarai, Z., Ananthram, V., Robertson, R.M., Robertson, D., 1997. Sympathetic and baroreceptor reflex function in neurally mediated syncope evoked by tilt. *Journal of Clinical Investigation*, 99, 2736-2744.
- Mosqueda-Garcia, R., Furlan, R., Tank, J., Fernandez-Violante, R., 2000. The elusive pathophysiology of neurally mediated syncope. *Circulation*, 102, 2898-2906.
- Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., Piccaluga, E., 1986. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, 59, 178-193.
- Pagani, M., Montano, N., Porta, A., Malliani, A., Abboud, F.M., Birkett, C., Somers, V.K., 1997. Relationship between spectral components of cardiovascular variabilities and direct measures of muscle sympathetic nerve activity in humans. *Circulation*, 95, 1441-1448.
- Preiss, G., Polosa, C., 1974. Patterns of sympathetic neuron activity associated with Mayer waves. *American Journal of Physiology*, 226, 724-730.
- Stauss, H., Anderson, E., Haines, W., Kregel, K., 1998. Frequency response characteristics of sympathetically mediated vasomotor waves in humans. *American Journal of Physiology*, 274, H1277-H1283

MODELING THE EFFECTS OF INTRA-ABDOMINAL HYPERTENSION

Jerry Batzel^(a), Stefan Fürtinger^(b), Daniel Schneditz^(c)

^(a)Institute of Mathematics and Scientific Computing, Univ. of Graz, Institute of Physiology Medical Univ of Graz

^(b)Institute of Mathematics and Scientific Computing, Univ. of Graz

^(c)Institute of Physiology, Medical Univ. of Graz

^(a)jerry.batzel@medunigraz.at,

^(b)stefan.fuertinger@uni-graz.at

^(c)daniel.schneditz@medunigraz.at

ABSTRACT

Intra-abdominal hypertension denotes elevated pressure in the abdomen. Such hypertension is common in patients in critical care hospital wards. The consequences of intra-abdominal hypertension involve constrained blood flow to vital organs such as the kidneys and abdominal organs (in connection with reduced cardiac output) potentially leading to multi-organ failure) Intra-abdominal hypertension represents an independent risk factor for mortality in acute and intensive care patients. Even given these observations, this condition is not always adequately monitored and much remains to be learned about its evolution and its impact on cardiovascular function. In addition, no significant modeling studies or models have been proposed to study this phenomenon. In this paper, we will discuss this problem, relevant mechanisms, and a model to simulate the effects of intra-abdominal hypertension on the cardiovascular system.

Keywords: venous return, systemic resistance, blood flow, Abdominal Compartment Syndrome, renal failure

1. INTRODUCTION

Intra-abdominal hypertension (IAH) and abdominal compartment syndrome (ACS) (Ivaturi et al [2006], Li [2012]) have only recently received significant attention. In fact, the World Society of the Abdominal Compartment Syndrome which defines the four grades of IAH was only established in 2004. The establishment of this society was in recognition of the importance, prevalence, and negative consequences of IAH in critically ill hospital patients. IAH is recognized as an independent risk factor for death in intensive care patients.

Definitions

Intra-abdominal pressure (IAP) is the pressure within the abdominal cavity. Typical steady state values for IAP in normal healthy individuals are 0 to 5 mm Hg. IAH is diagnosed if the IAH is greater than 12 mmHg either over a continuous period or with repeated

episodes. IAH is divided into four grades with grade IV IAH reflecting pressure greater than 25 mmHg.

Consequences

The abdomen includes many important organs including the intestines, kidney, and liver. IAH can influence the function of such systems (De Waele and De Lit [2007] and in particular cause a local reduction in blood perfusion to such organs leading to organ damage (see, e.g., Shibagaki [2006]). In addition, IAH can reduce venous return thereby affecting organ systems outside of the abdominal region such as interfering with diastolic filling of the heart and thereby reducing cardiac output. These effects can in turn influence blood perfusion in the brain.

2. MOTIVATION AND METHOD

Limited modeling studies exist related to intracellular and extracellular fluid volume change which may influence IAP (Tartara and Tashiro [2007]) and little modeling on the influence of IAP on hemodynamics. We describe a mathematical model that can in a straight forward way be adapted to test the impact of increased abdominal hypertension. The ultimate goal is to develop a test model of IAH which will provide the basis for a more comprehensive model that can provide qualitative analysis of the consequences of IAH and potentially lead to important insights for detection of IAH in the clinical critical care setting. Given the number of cardiovascular consequences of IAH, such a model for diagnosis or diagnostic training would be very useful.

3. MODEL

The mathematical model is adapted from a model for examining orthostatic stress using experiments from lower body negative pressure (LBNP) and head up tilt or (HUT) and blood loss due to hemorrhage (Batzel et al [2006, 2009], Kappel et al [2007], Fink et al [2004]). Based on the above applications, the model is very well suited for the proposed extension to study IAH. In particular:

1. The initial model includes 10 compartments (Figure 1) representing various body tissues compartments as well as control mechanisms, and plasma-interstitial fluid exchange. The model thus distinguishes between the abdominal compartments and other compartments.
2. The implementation of changes in transmural pressure in intra-abdominal vessels can be directly applied via the application of positive intra-abdominal pressure on the vasculature.
3. The model focuses on control responses which will also be a focus here. Unstressed venous volume also is implemented in the model which is relevant for the current proposed work.

Mass balance relations

Model compartment equations are mass balance relations. The generic form for these mass balance relations is illustrated in the equation below where for a generic compartment comp in Figure 1, P represents the pressure, c compliance, V the volume of the compartment, F the flow (in and out) of a compartment, and V_u unstressed volume:

$$(1) \quad c \frac{dP_{comp}}{dt} = F_{in} - F_{out} - \frac{dc}{dt} (P_{comp} + P_{bias}) - c \frac{dP_{bias}}{dt} - \frac{dV_u}{dt}$$

Transmural pressure can be implemented via positive or negative changes in the term P_{bias} .

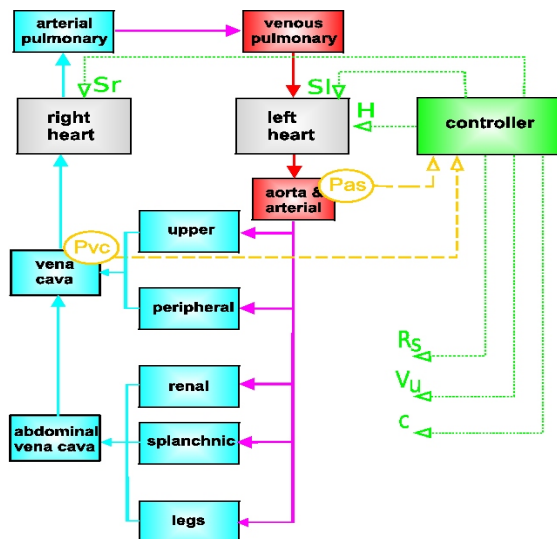


Figure 1: Model block diagram

The model incorporates lumped volume compartments where each compartment is considered to be a vascular compartment of blood or fluid characterized by the total fluid volume, by fluid pressure, and by vascular compliance (that reflects the vascular stretch due to pressure). Compartmental unstressed volume V_u represents the volume of a compartment before the vascular walls are subjected to a transmural pressure that stretches the vascular wall to accommodate the incorporation of additional fluid volume. The control of unstressed volume is an important response to loss of dynamic blood volume (the volume under pressure).

Flows between compartments

The equation for blood flow in or out of a compartment takes the form of Ohm's law so that, for example, the left equation below describes the flow into the upper body compartment dependent on arterial and venous pressures depicted in Figure 1 and the right equation give the flow into the splanchnic region:

$$(2) \quad F_{up,in} = \frac{P_{as} - P_{vc}}{R_{up}} \quad \text{and} \quad F_{spl,in} = \frac{P_{as} - P_{avc}}{R_{spl}}$$

An important mechanism to consider is the impact of IAH on vascular resistances due to constriction of vessels under positive transmural pressure (see below). Additional features included in the vascular flow structure includes venous valves to block back flow and also minimal flow constraints that act as autoregulatory action to override the baroreflex control of vascular resistance which induces contraction in certain arterioles to support blood pressure.

Cardiac output

Cardiac output is determined by heart rate and stroke volume. The latter is determined by ventricular preload and afterload and a relation for heart muscle contractility determined by state equations for contractility influenced by heart rate (see Batzel et al [2009]). The model is nonpulsatile as currently configured but pulsatile effects can be implemented as in Olufsen et al [2005].

Control mechanisms

The model incorporates control response to changes in arterial pressure P_{as} and venous pressure P_{vc} through the baroreflex which alters heart rate H , (and hence indirectly heart muscle contractility S), unstressed volume V_u , compliance c , and vascular resistance R (See Figure 1 controller) to stabilize arterial pressure after perturbation from a steady state operating point. These controls are based on the model of Olufsen et al [2005].

1. Each control equation is described in terms of a set point function and a differential equation that determines the dynamics of the control response based on deviations in the arterial and venous blood pressures

- Arterial pressure is sensed by the arterial baroreceptors in the aorta and carotid bodies and venous pressure by the cardiopulmonary sensors in and around the right atrium and in the pulmonary circuit
- Mobilizable unstressed volume which can be shifted to dynamic circulation to support blood pressure is incorporated into several compartments including the lower body (legs) compartment and in the splanchnic region (abdomen, intestines, and including liver) and kidneys. In addition the global control of systemic resistance is apportioned proportionally to the arterial vascular elements subject to a maximal increase that ensures a minimal blood flow (autoregulatory effect).

Fluid volume control and exchange with interstitium

The exchange of fluid between the interstitial space and the vascular space is an important mechanism for regulating volume in normal healthy individuals and also in patients depending on dialysis [Schneditz [1992, 2005]]. The current model includes a submodel of this exchange which can aid in supporting circulating blood volume. This aspect of fluid volume control can have important effects on interstitial volume and edema which can have negative effects on structures such in the microvasculature.

Additional model details for the model given in Figure 1 can be found in Appendix A.1 of Batzel et al [2009].

Model adaptations for IAH and validation

Model adaptations to reflect IAH need to include several effects produced by a relative increase in exterior pressure relative to the internal pressure of key organs (compartments) and the vascular structures (arteries and veins) connecting them. This pressure differential will be referred to as a positive transmural pressure (exterior to interior). In contrast, during studies using lower body negative pressure (LBNP), exterior air pressure is reduced creating a negative transmural pressure differential that stretches the vascular elements causing pooling of blood in compartments where the negative pressure is applied. This acts as an effective sequestration of blood reducing the blood volume under dynamic pressure which determines blood flow.

The key effects to consider include the following.

- IAH influences the transmural pressure and this effect will be introduced via a positive bias pressure P_{bias} in the mass balance equation of form (1) for any compartment subject to IAH. This pressure acts to reduce the compartment volume. A positive bias pressure will be applied to the abdominal compartments (abdominal vena cava, renal, and splanchnic compartments).
- This positive pressure will also reduce compartment volumes of inflow arteries and

arterioles (incorporating capillaries) and outflow venules and veins, changing the vascular flow resistances. Such variations in inflow and outflow resistances reflect constriction due to the positive external pressure applied to arteries and veins. This effect can be implemented as variations in inflow and outflow resistance parameters but the effects are complex and a comprehensive model will require additional modeling studies to implement.

The changing hemodynamic environment due to pressure changes and resistances can influence blood flow to organs where IAH exists and may also cause changes in cardiac output and cerebral blood flow. These interactive effects are under investigation in current work.

The model can be used to assess in particular the interaction of several key variables including

- Changes in intra-abdominal pressure
- Changes in vascular resistances to abdominal blood inflow and outflow
- Degree of lower body blood volume sequestration
- Venous return
- Arterial and femoral, blood pressure and intracranial pressure.

4. FIRST SIMULATIONS

The following figures are generated with a +10 P_{bias} induced at $t=0$ over a period of 1 minute and applied to the abdominal vena cava, splanchnic, and renal compartments. No change in arterial resistance is assumed for these simulations. Pressures are provided in Figure 2 and control responses in Figure 3.

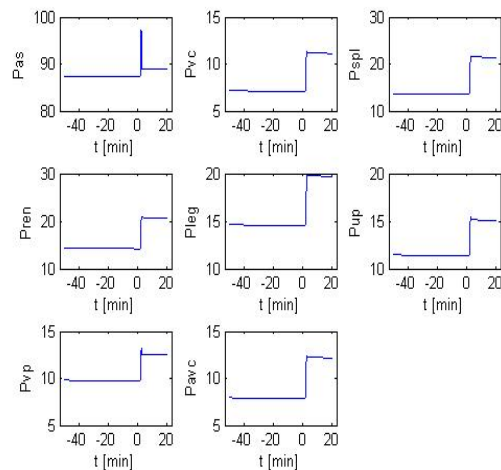


Figure 2: Model outputs for pressures. P represents pressure, for the various compartments in Figure 1.

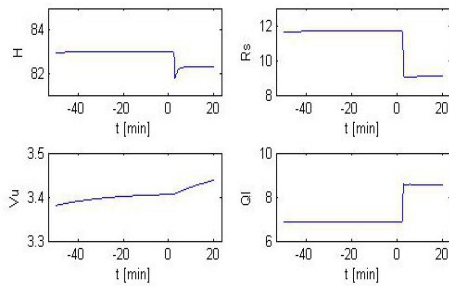


Figure 3: Model outputs for controls and cardiac output, V_u represents unstressed volume control. H represents heart rate, while R_s represents systemic resistance, and Q_l left ventricular cardiac output,

Figure 2 and Figure 3 illustrate the impact of a change in P_{bias} pressure alone. The following points are made:

1. In general the compartment pressures rise as a consequence of the external pressure which physiologically act to compress compartment volumes. In terms of modeling the bias pressures reduces the volume via the difference between pressure P and P_{bias} in (1).
2. The increase in pressures generate and perhaps are partly produced by an increase in left cardiac output Q_l as depicted in Figure 3.
3. As a consequence of the spike in P_{as} shown in Figure 2, the control response acts to reduce systemic resistance R_s in Figure 3 which also acts to increase Q_l .

The above effects are due only to the change in P_{bias} . Additional factors must be included to reflect various effects of IAH. As mentioned above, it certainly is the case that IAH constricts arteries and veins. Currently in the model, no direct impact of IAH on resistances is included.

Arterial and venous blood flow resistances R are currently incorporated in the model as resistance to flow between compartments as given in (2). An increase in local resistance due to constriction induced by IAH can be implemented as a local increase in resistance in those vascular elements subject to IAH. This will impact, in conjunction with other effects such as pressure differentials between compartments, the blood flow to those compartments. In turn, reduced blood flow could damage tissues such as is sometimes seen in the kidneys of patients with IAH. In fact high compartment pressures may also have an impact as discussed below.

In the above simulations, cardiac output rises due to the global control response of reducing resistance (Figure 3) in response to the spike in arterial pressure

(Figure 2). Including an additional effect of increasing local resistance due to constriction caused by IAH will counteract this effect and will certainly increase flow resistance to compartments where IAH is experienced. The net result could reduce overall cardiac output.

5. DISCUSSION

An important area where IAH has serious consequences is in kidney function (Mohmand and Goldfarb [2011]), Renal dysfunction is a common early consequence of IAH with renal hypoperfusion beginning approximately at abdominal pressures of 8-12 mm Hg and peritoneal dialysis is a risk factor for IAH (Shibagaki et al [2006]). The mechanisms by which IAH can influence renal health and function include many factors including reduced cardiac output, increased renal vascular resistance, and decreased glomerular filtration etc.

In addition the increase in compartmental renal pressure can have a negative impact on renal tissue health and function, which could reflect intrarenal venous congestion and also act to reduce the renal filtration rate seen in renal function under IAH (Mohmand and Goldfarb [2011]).

The overall impact of IAH can extend beyond those compartments where IAH is induced. Figure 2 indicates that just the inclusion of positive bias pressure in some compartments can cause pressure changes even in compartments where IAH is not induced. In addition other effects can be introduced. For example, intracranial pressure has been observed to rise (Mohmand and Goldfarb [2011]),

Recently the problem of IAH has been given greater attention in critical care situations where IAH can arise as a consequence of several factors including techniques during surgery. Increased efforts have been made to monitor and probe for IAH in critical care patients and greater efforts have also been made to train critical care workers in this area. A well validated model of IAH could be used to aid in this training and in testing the hemodynamic implications of IAH.

SUMMARY

As can be seen from the above discussion, the interaction of local effects of IAH and global control response to perturbations in steady state pressures and flows is very complex. Quantifying these interactions can lead to better detection and treatment in patients as well as better training for those dealing with IAH, especially in the critical care area. Further work will be directed to generate a more comprehensive model of the effects and interactions of mechanisms related to IAH.

ACKNOWLEDGMENTS

We wish to acknowledge and to thank the Steiermarkischen Landesregierung für Wissenschaft und Forschung, Austria, for support of this work.

REFERENCES

Batzel, J. J., Fürtinger, S., Fink, M., Bachar, M., and Kappel, F., 2009. Sensitivity analysis of a model for the cardiovascular control system. Technical report IMA0309 2009, Institute of Mathematics and Scientific Computation, University of Graz.

Batzel, J. J., Kappel, F., Tran, H.T., Schneditz, D., 2006. *Cardiovascular and Respiratory Systems: Modeling, Analysis, and Control*. Siam, Philadelphia.

De Waele, J. J. and De Laet I., 2007. Intra-abdominal hypertension and the effect on renal function. *Acta Clin Belg Suppl*, 2007(2), 371-374.

Fink, M., Batzel, J.J., Kappel, F., 2004. An optimal control approach to modeling the cardiovascular-respiratory system: An application to orthostatic stress. *Cardiovasc Eng, an International Journal*, 4, 27-- 38.

Kappel, F., Fink, M., and Batzel J. J., 2007. Aspects of control of the cardiovascular-respiratory system during orthostatic stress induced by lower body negative pressure. *Math Biosciences*, 206, 273-308.

Ivatury,R.R.,Cheatham, M. L., Malbrain, M., and Sugrue, M., eds. 2006. *Abdominal Compartment Syndrome*. George-town, TX: Landis Bioscience; 2006:1-7.

Lee, R. K., 2012. Intra-abdominal hypertension and abdominal compartment syndrome: a comprehensive overview. *Crit Care Nurse*, 32(1), 19-31.

Mohmand H, and Goldfarb S., 2011. Renal dysfunction associated with intra-abdominal hypertension and the abdominal compartment syndrome. *J Am Soc Nephrol*. 22(4), 615-621

Olufsen, M.S., Ottesen, J.T., Tran, H.T., Ellwein, L.M., Lipsitz, L.A., Novak, V., 2005. Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation. *J Appl Physiol*, 99, 1523-1537.

Schneditz, D., Roob, J., Oswald, M., Poggitsch, H., Moser, M., Kenner, T., Binswange, U., 1992. Nature and rate of vascular refilling during hemodialysis and ultrafiltration. *Kidney Int*, 42,1425-1433.

Schneditz, D., 2005. Technological aspects of hemodialysis and peritoneal dialysis. In: A.R. Nissenson, R. Fine, eds, *Clinical Dialysis*, McGraw-Hill, New York, 4th edn., 47 - 83.

Shibagaki, Y., Tai, C., Nayak, A., Wahba, I., 2006. Intra-abdominal hypertension is an under-appreciated cause of acute renal failure. *Nephrol Dial Transplant*, 21(12), 3567-70.

Tatara T, Tashiro C., 2007. Quantitative analysis of fluid balance during abdominal surgery. *Anesth Analg*, 104(2):347-54.

AUTHORS BIOGRAPHY

Jerry Batzel is a mathematician research associate of the Institute of Mathematics and Scientific Computing University of Graz and the Physiology Medical University of Graz with extensive experience in mathematical modeling of physiological systems in particular the cardiovascular and respiratory control systems. His area of interest includes short-term and long-term control of blood volume and blood pressure.

Stefan Fürtinger has a Masters Degree in applied mathematics from the University of Graz. He currently holds a PhD-position in the field of mathematical image processing at the University of Graz.

Daniel Schneditz is Assoc. Professor of Physiology at the Institute of Physiology Medical University of Graz and an expert on renal function and hemodialysis including kinetic modeling of solute and fluid changes in normal subjects and patients undergoing dialysis.

MORBISIMMOD - MORBIDITY BASED NEEDS ASSESSMENT USING MICROSIMULATION

M. Gyimesi ^(a), I. Czasny ^(b), G. Fülöp ^(c), S. Mathis-Edenhofer ^(d)

^(a) ^(b) ^(c) ^(d) Gesundheit Österreich GmbH, Österreichisches Bundesinstitut für Gesundheitswesen, Vienna, Austria

^(a)michael.gyimesi@goeg.at, ^(b)ines.czasny@goeg.at, ^(c)gerhard.fueloep@goeg.at, ^(d)stefan.mathis-edenhofer@goeg.at

ABSTRACT

Needs-based health care planning is a major objective of modern health care systems and has become a major approach in the 1990s in many countries. Despite need can be seen in different ways and is not always objectively measurable, it is not questioned that health care services should be designed to meet the needs of the community based on the community's morbidity. In order to improve Austrian health care planning we developed a morbidity-based simulation model, the MorbiSimmod, where we combined the protocol of Health Care Needs Assessment with a population- and needs-based policy approach and choose microsimulation as simulation technology. The basic morbidity information is derived from inpatient routine health care data, data from the AT-HIS 2006/2007 plus sociodemographic and socioeconomic data. Together with spatial data we are able to calculate stratified figures of morbidity for in-depth analysis of structural requirements for specific health care services in Austria.

Keywords: microsimulation, needs-based health care planning, health services research, Austrian health care system

1. INTRODUCTION

Health care planning in Austria has a long tradition and is done at the moment primarily by extrapolating routine health care data. In order to improve future projections of the Austrian health care system a so called macro analytical approach of health services research was developed, where the health care need can be related to structural requirements of the Austrian health care system and the need for health becomes a more important part in the planning process (Gyimesi 2012).

Needs based health care planning is a major objective of modern health care systems and has become a major approach in the 1990s, particularly in the United Kingdom, Netherlands, New Zealand, Australia and the United States (Stevens 2004). But besides the beguiling simplicity about the idea that health care services should be designed to meet the needs of the community need can be seen in different ways and the concept of needs as objective states that

can be measured is not always feasible or applicable. (Mooney 2006). Therefore various methods for health care planning are in use and each country follows its own way to solve the methodical difficulties as well as the political constraints when implementing health care planning systems.

Despite the common problem of lacking valid data there is also a big discussion in defining the meaning of "need". In general three different aspects should be distinguished:

- Desire
- Need
- Utilization / Demand

Concerning data, surveys on the one hand give an impression of the personal judgment, how much and which kind of health care services are desired within a community. Routine health care data on the other hand, which are produced by the health care system itself, give an overview to the actual utilization of the existing health care services. The real "need" in fact is somewhere situated in between. The definition, when a desire becomes a need, is hard to set and will differ within countries depending on societal values as well as budgetary constraints. When only focusing on the actual utilization, the real need can be underestimated for example in case of missing or too expensive health care providers. On the other hand we also have to bear in mind the so called supply-induced demand, which exceeds the actual need.

Since finding a common definition for the need of health care services is one of the basic requirements to quantify the actual need, this discussion can be a very controversial one and in the end a political decision is needed.

Methods of Health Care Planning

Although the conceptual difficulties concerning health care needs are not negligible, the question how to plan specific services for specific populations groups has still to be answered and different concepts have been developed to tackle the problem. Depending on the

available data different methods of estimating the requirement of health care services have been developed. Some of them are listed below (Spycher 2004):

The *Utilization-based Method* focuses on the utilization of the existing health care services. The main advantage of this method is the usually easy availability of data. The big disadvantage is that using this method the existing structure and utilization of the health care system will be continued into the future. Neither medical improvements nor possibly existing under- or over-supply are taken into account.

Another often used method is the *Manpower-Population-Ratio*. Using this method benchmarks concerning the number of services per population are set. These numbers are easy to calculate but do not take into account regional or epidemiological diversities. The main challenge using this method is the selection of reasonable ratios. Mostly measures from other countries or previous periods are taken over, which can lead into controversial discussions concerning the suitability of the chosen ratios.

Accessibility-oriented-Methods fix a distance or time within the next specific health care service has to be reached by the population living in a certain area.

As none of these concepts is alone and straight forward applicable for Austrian health care planning mostly a combination of these three methods is used. Since all of these methods are limited by the fact that epidemiological diversities between regions and time are not taken into account, our aim is to complement these methods with the results of a *morbidity based needs assessment* and we developed the so called MorbiSimmod.

The MorbiSimmod is a morbidity driven model that derives the necessary epidemiologic information by a microsimulation model. It starts with the Needs-related policy model, presented by C. Sanderson and R. Gruen in 2006 (Sanderson 2006). This very simple model starts with the number of people in a population or subgroup of the population and analyses a specific health services situation in a stepwise process by reducing these numbers in a way that in the end the number of people who are planned to receive an health service intervention is calculated.

To close the gap between morbidity information of a population group and the structural requirements of health care services we adopt some ideas from the Health Care Needs Assessment Project (HCNA), funded by the Department of Health and latterly the National Institute of Health and Clinical Excellence (NICE) and managed at the University of Birmingham. The project started in the 1990s and produced 38 needs

assessments covering many diseases and populations. (Stevens 2004).

HCNA starts with the overlapping but not identical concepts of need, demand and supply, thereby addressing areas of potential health care improvement (Figure 1) and defines a general protocol which is applicable for a wide range of needs assessment scenarios.

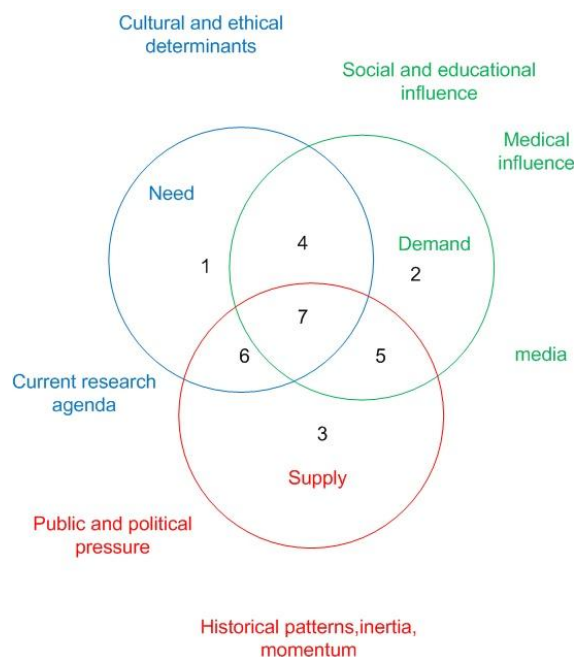


Figure 1: Relation of health care need, demand and supply with influencing factors (Stevens 2004)

Both models rely heavily on the calculation of epidemiologic figures. In the MorbiSimmod model we choose to integrate an (agent-based) microsimulation model, based on information of the Austrian health care system like the documentation of diagnoses and procedures in Austrian hospitals, survey data of the last Austrian Health Interview Survey (AT-HIS), spatial data and further sociodemographic data. The epidemiologic figures generated with the microsimulation model are then transformed into structural requirement data, which can be used for different purposes according to the initial question.

The paper is organized as follows: Section two and three will introduce the Needs-related policy model and HCNA in more detail, section four will explain concepts of micro-modeling and why it is useful for calculation of morbidity scenarios. Section five presents the input data for the MorbiSimmod and section six explains the MorbiSimmod in more detail. Finally, we will discuss the strength of the MorbiSimmod model as well as the limitations and further development.

2. NEEDS-RELATED POLICY MODEL

The Needs-related policy model is an epidemiology based needs assessment approach and was defined by Sanderson and Gruen in 2006 as a special case of population-based needs assessment (Sanderson 2006).

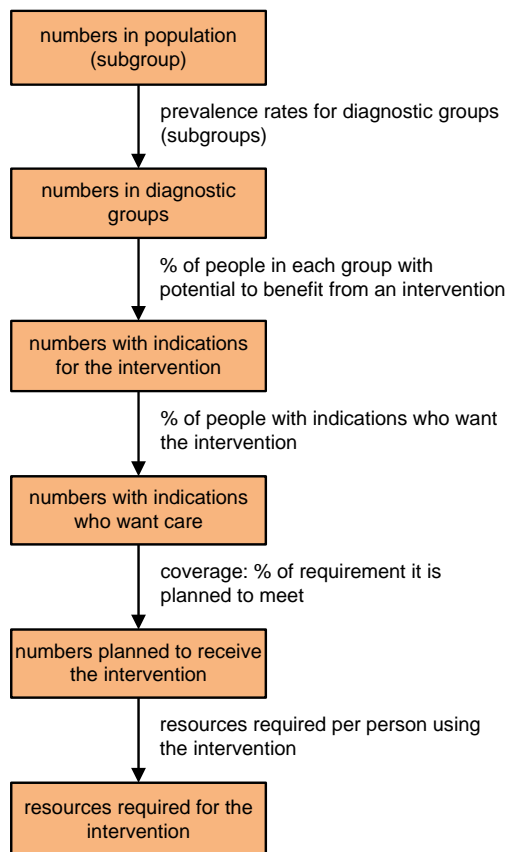


Figure 2: Needs-related policy approach (Sanderson 2006)

This model calculates numbers of population groups with decreasing numbers step by step (Figure 2).

1. The Model starts with the number of the population or an arbitrary subgroup of the population (e.g. people in a specific federal state, people in a specific age group).
2. For this group of people epidemiologic information is gathered to identify the group of people who have a specific disease or group of diseases.
3. As not all of the remaining people are eligible for an appropriate intervention, only these with indication for an intervention are taken.
4. In the next step, people who are not willing to get the intervention are eliminated. The remaining group includes people who have a relevant disease, have the indication for an intervention and want the intervention.

5. Therefore a decision has to be made, how many of these people's requirements are planned to met.
6. Based on known structural requirements, numbers for needed structures can be calculated.

The modeling for this population-based approach seems technically straight-forward as it connects easily to microsimulation models (cf. section 4) but it depends heavily on the available data to perform each of the recommended steps. Which data we use and how to use it will be explained in section 5.

Although the Needs-related policy model serves well to formulate the calculation process of morbidity of a given population with a given medical condition, it does not emphasize the distinction of different care models like different possible procedures for the same medical problem or different locations of service provision.

Therefore, we use another epidemiology based needs assessment model with more focus on models of care.

3. HCNA

HCNA became established in the years from 1980 and 1990 in the United Kingdom (Stevens 2004). The aim of HCNA is to provide information to plan, negotiate and change services for the better and to improve health in other ways. Because HCNA bases much on conclusions from epidemiological data (but also on 'effectiveness and cost effectiveness' and on 'existing services') the method is described by its authors as an 'epidemiological approach to needs assessment'. Unlike other approaches that capitalise on existing services and make incremental changes to them, HCNA has a strong focus on non-local, but population based epidemiological data.

For applying the method on a certain health problem the authors of the methodical handbook on HCNA have defined a 'universal protocol' (Stevens 2004) for the 'epidemiological approach' of needs assessment. This protocol is summarized in the following text.

3.1. Statement of the problem

At the beginning the authors of an individual HCNA make a precise statement of the problem also including an appropriate description of the context and perspectives of related controversies.

3.2. Sub-categories

The health problem is sub-divided into categories that best support the requirements for service planning.

3.3. Prevalence and incidence

Epidemiological data may derive from the epidemiological literature, from official statistics and from national health surveys. Data must be adjusted to estimate regional frequencies. Beside these epidemiological information two other types of information is needed in order to define the need for health care by means of 'the population's ability to benefit from health care': existing services and (cost-) effectiveness.

3.4. Services available and their costs

The situation of currently provided services can serve as starting point for considering change of services. For this section it is relevant how many specialists are available per unit of population (structure) and how many people are treated per unit of population (process).

3.5. Effectiveness and cost-effectiveness of services

Data on effectiveness and cost effectiveness may be derived from the Cochrane Collaboration initiative and other initiatives of the Evidence Based Medicine movement. The evidence for an intervention can be graded regarding the size of effect and the quality of evidence.

3.6. Quantified models of care and recommendations

The authors try to find a formula that calculates the need from incidence and prevalence figures. From disparities of this calculated need and the current service provision recommendations emerge. Due to several lacks of information (inadequate data, lack of relevant thresholds, definitions and because of complex pathways) various flexible approaches must be developed in dependence of the actual problem.

3.7. Outcome measures, audit methods and targets

To support long term agreements on services and other health care activities the authors of HCNA must identify measures and targets that might be used. Agreements may include specific services and health and quality objectives.

3.8. Information and research requirements

The assessment may bring out new research questions respectively a lack of information or data. Such challenges and questions can be forwarded to national research programs or a national health priority setting program.

4. MICROSIMULATION

As stated in recent publications the use of dynamic models is a small but emerging field in health services research. It can provide an important complement to static and statistical analysis of health care problems and as shown by Einzinger the addition of routine health care data can be particularly valuable (Einzinger et al., 2012). Furthermore Austria has a history in using dynamic modeling and simulation as decision support

instrument in health care (Miksch 2010, Breitenecker 2010a, Zauner 2010, Miksch 2011).

While simulation models with a top-down approach are usually easier to generate, they lack the possibility of modeling complex dependencies of the system (Zuchelli 2010). Therefore we decided to use a bottom-up microsimulation model, a very flexible modeling technique being able to account for population heterogeneity, multiple outcomes, the capacity to capture long-run effects and fitting to our health care needs assessment approaches. It gained raising popularity in recent years as a tool for ex-ante evaluation of health policies, and particularly in the field of public health interventions, where evaluations are often challenging and costly. (Zuchelli 2010, Brown 2011, Lymer 2011a, Lymer 2011b, Spielauer 2007, IQWIG 2009, Brouwers 2011, Décarie 2012).

The Austrian health care data the model is build on are mainly collected for reimbursement purposes and despite the Austrian health care payment system is highly fragmented the data from inpatient sector and outpatient sector can be linked statistically to suit microsimulation modeling (Katschnig 2012). More details on the data aspect can be found in section 5.

To analyse the structural requirements of specific Austrian health care situations we developed a fictitious analogue to the Austrian population with statistical representatives or agents. These agents do have diagnosis information from intramural care as well as extramural care. By adding spatial data the agents can be differentiated by municipal districts.

At the moment we do not think of our model as an agent-based one, because agents are not aware of other agents and environmental parameters, e.g. the dynamic evolution of the model population is generated just corresponding to the demographic forecast of Statistics Austria up to 2030 and no feedback mechanism concerning raising or falling morbidity is implemented. Although there is no clear separation between agent-based modeling and micro-modeling, we refer to MorbiSimmod as a micro-simulation model because of its data-driven approach.

5. DATA

5.1. Routine health care data

The health care system in Austria mainly consists of two sectors: the inpatient care in hospitals and the outpatient care, which is provided by physicians (general practitioners and specialists) and medical institutes as well as by ambulances in hospitals.

Since these sectors are financed in different ways the quality, the reliability and the completeness of the existing billing data are quite different.

Concerning the **inpatient care** comprehensive routine data are available for all hospitals (DLD). Since the payment of hospital care in Austria is case-based, taking into account the diagnosis (ICD coded) and the main undertaken procedures, detailed hospital admission statistics also including basic personal information for each patient (e.g. sex, age and place of residence) are provided by each hospital. Since these data are already collected for several years and comprehensive plausibility checks are obligatory, data quality and completeness of the inpatient data are quite satisfying.

Nevertheless, the usefulness of these routine data for estimating morbidity is seen controversial. For instance, an important methodological limitation is that these data are originally collected for the purpose of payment. Since a corresponding effect on the selection of diagnoses cannot be ruled out, the reliability of these routine data for epidemiologic purposes is limited. Another possible limitation is the missing unique personal identifiers in the data available for our studies. Instead the data records are only linked to single hospital admissions. Therefore the actual number of patients with a specific diagnosis can only be estimated by identifying readmissions using the given personal indicators.

Concerning the **outpatient care** the situation is quite different: until now no systematic recording of the diagnosis is obligatory. Therefore at the moment we do not have reliable routine data providing evidence on the “outpatient morbidity”.

For using the inpatient data in the MorbiSimmod the DLD-data concerning the years 2006 and 2007 were analysed and aggregated. The resulting dataset DLD-A contains anonymous information for each patient, who was discharged from an Austrian hospital (including sanatoria and rehabilitation centers) in the year 2006 or 2007. The data set includes information about sex, age, nationality, place of residence and the information whether a diagnosis within a specific ICD-10 chapter was reported.

It has to be taken into account that these routine hospital data can only be used for calculating the “inpatient morbidity”. Whether or not this can be seen as a good estimation for the general morbidity depends on the specific diagnosis. Whereas these hospital data will be very suitable for prevalence estimation of diseases such as heart attack or cataract the prevalence of diseases, which do not necessarily require hospital care, will be underestimated.

5.2. AT-HIS

To complement the prevailing results, the Austrian Health Interview Survey (AT-HIS) 2006/2007 was chosen as a second data source. The results of this

survey are representative for the Austrian population in the age of 15 years and older (Klimont 2007).

The main advantage of this data source is the direct connection to the population’s morbidity without referring to routine data, and therefore to the existing health care services and their utilization. Nevertheless it has to be taken into account that self reporting health problems can also induce underestimation (e.g. psychological problems) as well as overestimation (e.g. migraine headache) of morbidity. That the information of AT-HIS is available just for some pre-defined major chronic diseases has to be seen as another severe limitation.

5.3. Sociodemographic and socioeconomic stratification

Besides age and sex other determinants of health like sociodemographic and socioeconomic factors are known to have an influence on the individual health status. E.g. recent analyses showed that the risk of chronic diseases depends on the educational level as well as on the available income. Also unemployment is a relevant risk factor, which influences morbidity. (e.g. Klimont 2008)

Since the DLD-dataset only includes basic patient-specific information such as sex and age other datasets are used to impute relevant sociodemographic and socioeconomic indicators.

In the end the characteristics of the statistical representatives of the Austrian population include

- Place of residence,
- Sex,
- Age,
- Educational level,
- Employment status and
- Income level.

5.4. Addition of further data

A main advantage of the micro-modeling approach is the extensibility of the basic data used for the simulation model. As there are other research groups in Austria, who use partially different data (e.g. the GAP-DRG database) for similar purposes (Weisser 2010, Endel 2010, Breitenacker 2010b), the MorbiSimmod is build in a way that data from additional databases can be added easily in the future.

6. THE MORBISIMMOD MODEL

The MorbiSimmod consists of two main models. On the one hand it propagates a predefined structured process to analyse the structural requirements for health care based on health care needs. On the other hand a microsimulation model was developed that enables us to investigate exactly these questions with a data-driven bottom-up approach for the Austrian health care system.

6.1. The MorbiSimmod Process model

In order to produce policy-relevant answers to health services research and health care planning questions not only microsimulation but some pre- and post-processing steps are part of the MorbiSimmod. In the preparation phase we integrate these tasks to specify the problem or question and to update the MorbiSimmod model according to the problem. In the post-processing phase we transfer the results of the microsimulation model calculation into corresponding models for health care system planning. Process steps are defined in figure 3.

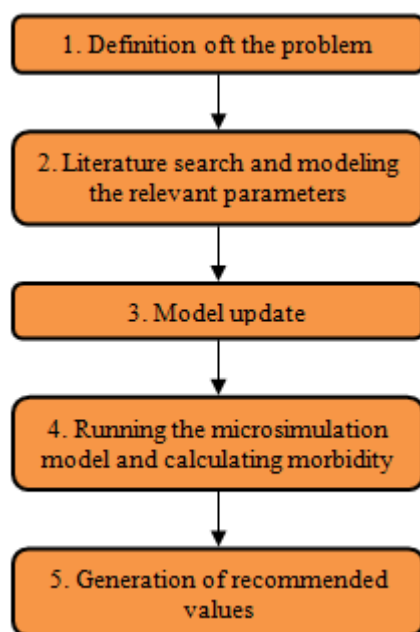


Figure 3: Process steps of MorbiSimmod

6.1.1. Definition of the problem

Based on the original health policy or public health question we operationalise the question by

- defining the the morbidity spectrum and the focused granularity (single diagnosis, diagnosis group, risk factor, etc...),
- defining the geographical target population,
- defining the target population regarding to specific characteristics (age stratum, social class,...),
- defining the focused time period respectively the time horizon for prognostic calculations,
- the identification of relevant planning benchmarks, guidance values, indicators levels or strategic controlling parameters that relate to the morbidity figures of the specific health problem.

6.1.2. Literature search and modeling the relevant strategic planning parameters

We search for basic information regarding the above defined health service problem in the literature. The search includes reports on epidemiological figures, new medical developments and additional information on relevant benchmarks, guiding values, strategic parameters or indicators. After the search we analyse the relations of relevant benchmarks, parameters and indicators for other influencing factors to be able to define certain analytical scenarios. Finally, we prepare specific models for benchmarks, parameters and indicators based on available morbidity input data of the microsimulation model. For each model underlying basic assumptions and definition of scenarios have to be set.

6.1.3. Model update

The basic microsimulation model is updated with the latest available data on morbidity. Additionally missing data in the context of the problem defined in the first step might be imputed by assumptions based on data from populations that are similar to the target population or by a demographic extrapolation of the target population by extrapolating medical trends.

6.1.4. Running the microsimulation model and calculating morbidity figures according to the initially formulated problem

This step delivers figures on the morbidity in the defined context (time horizon, target population and spatial granularity).

6.1.5. Generation of recommended values

Step five connects the results of the generated morbidity scenarios to the secondary models for the selected planning benchmarks, guidance values, indicator levels or strategic controlling parameters. These results have to be calculated on the base of the assumptions and scenarios defined in step 2.

6.2. The MorbiSimmod simulation model

The MorbiSimmod simulation model overlaps with the process model in steps 3 to 5 (figure 3), when the appropriate indicators for the defined health care question are chosen, the simulation is run and specified output figures are calculated.

7. CONCLUSIONS

Needs based health care planning is a major objective of modern health care systems despite the fact that the concept of needs as objective measurable states that can be measured is not always feasible or applicable. Up to now in Austria health care planning is based on a mixture of the utilization-based, manpower-population-ratio and accessibility-oriented planning method.

Our aim with the development of the MorbiSimmod is to complement these methods with a morbidity based needs assessment method. The

MorbiSimmod derives the necessary epidemiologic information by a microsimulation model. It incorporates ideas from the Needs-related policy model developed by Sanderson and Gruen as well as the Health Care Needs Assessment project developed at the University of Birmingham.

The MorbiSimmod model consists of two main models. On the one hand it propagates a predefined structured process to analyse structural requirements for health care based on health care needs. On the other hand a microsimulation model was developed, that enables us to investigate exactly these questions with a data-driven bottom-up approach for the Austrian health care system. Comprehensive data from inpatient care, survey data from the Austrian Health Interview Survey (AT-HIS) 2006/2007, spatial data on a municipal level as well as sociodemographic and socioeconomic data are used to compute result figures for defined scenarios. At the moment the dynamics of the model are triggered just by demographic extrapolation.

Future developments include the implementation of more complex relations of agents with the environment and inter-agent relations. Furthermore, it is planned to integrate additional data sources.

REFERENCES

- Breitenecker, F., Einzinger, P., Endel, G., Gyimesi, M., Meier, L., Pfeffer, N., Popper, N., Weisser, A., 2010a. Development of a Dynamic Model Structure for Comparing Ambulatory Reimbursement Systems. *Proceedings of the SMDM Society for Medical Decision Making Europe - Public Health Decision Making*, May 30 - June 2, Hall in Tyrol, Austria.
- Breitenecker, F., Miksch, F., Popper, N., Urach, C., Wilbacher, I., Zauner, G., 2010b. Distance Dependence on the Willingness of Patients to Participate in Ambulatory Heart Rehabilitation Programs. *Proceedings of the SMDM Society for Medical Decision Making Europe - Public Health Decision Making*, May 30 - June 2, Hall in Tyrol, Austria.
- Brouwers, L., Ekholm, A., Janlöv, N., Johansson, P., Mossler, K., 2011. Simulating the need for health- and elderly care in Sweden — a model description of Sesim-LEV. *Proceedings of the The third General Conference of the International Microsimulation*, 8. - 10. June, Stockholm, Sweden.
- Brown, L. 2011. Editorial Special Issue on 'Health and Microsimulation'. *International Journal of Microsimulation* 4:1-2.
- Décarie, Y., Boissonneault, M., Légaré, J., 2012. *An Inventory of Canadian Microsimulation Models*. Available from: <http://socserv.mcmaster.ca/sedap/p/sedap298.pdf> [accessed 19. July 2012].
- Einzinger, P., Jung, R., Pfeffer, N., 2012. Modeling Health Care Systems – An Approach Using Routine Health Care Data. *Proceedings of the MATHMOD*, February 15 - 17, Vienna.
- Endel, F., Weisser, A., Filzmoser, P., Eisl, A., Endel, G., 2010. Obtain valid diagnoses from prescription data. *Proceedings of the 26th PCSI Conference*, 15. -18. September, Munich.
- Gyimesi, M., Mathis-Edenhofer, S., Grabner, I., 2012. MorbiSimmod- ein Mikrosimulationsmodell zur morbiditätsbasierten Bedarfsschätzung in Österreich. *Accepted Talk Deutscher Kongress für Versorgungsforschung*, 27. - 29. September, Dresden.
- Institute for Quality and Efficiency in Health Care (IQWiG), 2009. *Working Paper Modelling*. Available from: https://www.iqwig.de/download/Working_Paper_Modelling.pdf [accessed 19. July 2012].
- Katschnig, H., Endel, F., Endel, G., Weibold, B., Filzmoser, P., 2012. Dementia and pathways of health services utilization in Austria: A record linkage study in a country with a fragmented provider payment system and only partially available unique patient identifiers. *Proceedings of the International Data Linkage Conference*, 2-4 May 2012, Perth, Australia.
- Klimont, J., Ihle, P., Baldaszi, E., Kytir, J., 2008. *Sozio-demographische und sozio-ökonomische Determinanten von Gesundheit - Auswertungen der Daten aus der Österreichischen Gesundheitsbefragung 2006/2007*. Wien: Statistik Austria.
- Klimont, J., Kytir, J.L., Barbara, 2007. *Österreichische Gesundheitsbefragung 2006/2007- Hauptergebnisse und methodische Dokumentation*. Wien: Statistik Austria.
- Lymer, S., 2011a. *More than just population ageing: an evaluation, using microsimulation, of the escalation of Australian Government expenditure on health over the next 40 years*. Available from: http://www.canberra.edu.au/events/home/view_by_event_id/206 [accessed 2012-07-12].
- Lymer, S., Brown, L., Duncan, A., 2011b. *Modelling the health system in an ageing Australia, using a dynamic microsimulation model*. Available from: https://guard.canberra.edu.au/natsem/index.php?mode=download&file_id=1174 [accessed 19. July 2012].
- Miksch, F., Popper, N., Zauner, G., Endel, G., Schiller-Frühwirth, I., Breiteneker, F., 2010. Evaluation of Dynamic Modelling Approaches for Vaccination Strategies. *Proceedings of the 4th Vaccine and ISV Annual Global Congress*, 03.10. - 05.10., Vienna, Austria.
- Miksch, F., Urach, C., Popper, N., Zauner, G., Endel, G., Schiller-Frühwirth, I., Breiteneker, F., 2011. New insights on the spread of influenza through agent based modeling. *Proceedings of the ISPOR*

- 14th Annual European Congress*, 5. - 8. November, Madrid, Spain.
- Mooney, G., Jan, S., Wiseman, V., 2006. Measuring health needs. In: R. Detels, J. McEwen, R. Beaglehole and H. Tanaka, eds. *Oxford textbook of public health*. Oxford: Oxford Univ. Press, 1765-1772.
- Sanderson, C.J., Gruen, R., 2006. *Analytical models for decision making*. Maidenhead: Open University Press.
- Spielauer, M. 2007. Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: survey and review. *International Journal of Microsimulation* 1:35-53.
- Spycher, S., 2004. *Prognose und Planung in der ambulanten Gesundheitsversorgung. Literaturanalyse und Expertengespräche zur Prognose und Planung des ambulanten medizinischen Personals in der Schweiz*. Neuchâtel: Schweizerisches Gesundheitsobservatorium.
- Stevens, A., Raftery, J., Mant, J 2004. *An Introduction to HCNA - The epidemiological approach to health care needs assessment*. Available from: <http://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/HCNA/intro/index.aspx> [accessed 13. July 2012].
- Weisser, A., Endel, F., Endel, G., Filzmoser, P., 2010. Results of the project ATC-ICD. *Proceedings of the 26th PCSI Conference*, Munich.
- Zauner, G., Popper, N., Zauner, G., Endel, G., Schiller-Frühwirth, I., Breitenecker, F., 2010. Herd immunity as a result of model complexity. *Proceedings of the 4th Vaccine and ISV Annual Global Congress*, 03.10. - 05.10., Vienna, Austria.
- Zucchelli, E., Jones, A., Rice, N., 2010. The evaluation of health policies through microsimulation methods. *HEDG working paper 10/03*. York.

MARIA: An Agent Driven Simulation for a Web Based Serious Game devoted to Renew Education Processes in Health Care

Agostino G. Bruzzone¹, Marco Frascio², Francesco Longo³, Marina Massei¹, Anna Siri², Alberto Tremori¹

¹ Simulation Team MITIM DIME University of Genoa, Italy
Email {agostino, massei, tremori}@itim.unige.it – Url www.itim.unige.it

² Simulation Center DISC, University of Genoa, Italy
Email {mfrascio, annasiri}@unige.it – Url www.itim.unige.it

³ MSC-LES, Mechanical Dept., University of Calabria, Italy
f.longo@unical.it

ABSTRACT

The paper proposes an innovative approach for applying Serious Games concepts in developing a new generation of agent driven models devoted to support and renew the educational processes in Health Care. This research is devoted to develop new models of virtual humans to be used as patients in a distributed simulation game together with students to teach them concepts related to the continuous care of patients; the research goal is to create a strong relationship between medical doctor students and their set of personal virtual patients, driven by intelligent agents. By following this approach, it is possible to reproduce the patient life cycles and teach students on subjects such as pathology evolution, preventive actions, symptoms identification and human relationships.

MARIA (Model for Advanced and Realistic patient simulation driven by Intelligent Agents) is the solution developed by the authors and it has to be regarded as an innovation in the educational programs for health students enabled by new technologies.

KEYWORDS: Simulation, Serious Games, Health Education, Mobile Training, Intelligent Agents

1. INTRODUCTION

Modeling & Simulation (M&S) is a very broad research area with an impressive range of applications; therefore in some sectors, M&S was historically most popular for training and education, for instance defense training is one of the major application areas of M&S. In such a context, it is evident the necessity to use live simulation, as well as constructive and virtual, to provide learning experience and training sessions over simulated mission environments.

Another sector where education and training are traditionally very important is the health care: medical doctor education requires strong and long efforts even just by using traditional approaches; in fact it is over one decade that simulation becomes popular even in this sector and a while range of solutions have been

developed for training in different domains from emergency medicine to intensive care and surgery. Innovative solutions and new technologies were developed from different center of Excellence of Simulation (From the Center for Medical Simulation at Harvard Medical School to Simulation Team in Genoa University). In particular virtual patients libraries are progressively evolving and are actually used in different educational centers and universities.

Anyway multiple studies and researches were carried out to demonstrate the effectiveness of simulation in medicine education above all in teaching clinical knowledge, medical protocols and procedures, team working and communication capabilities (Okuda et al. 2009).

The synergy between M&S and Medical Doctors provided many benefits in medical simulation evolutions, therefore since now most of these important steps forward was driven by combining different backgrounds: needs and tools were paired, so for instance new training equipment and models are developed by simulationists in order to create solutions able to support health students based on doctor training procedures.

Today the technology advances and the joint experience allows moving forward from multi culture (where each one brings its one background) to trans-discipline (Trindade et al. 2002): by this approach the different cultures (i.e. simulation and medicine) work together to create new concepts and processes with benefits of a common development.

In this paper, the authors outline a new concept (for new medical doctors) that represents a change in the traditional educational programs: teaching doctoral students to take care of their patients along the entire patient life cycle, dealing with their needs and fears. To this end the authors propose MARIA. MARIA (the acronym stands for *Model for Advanced and Realistic patient simulation driven by Intelligent Agents*) is an advanced serious game for multiple players based on the combination of web-based simulation and intelligent agents.

In the sequel a brief description of the paper organization is reported. Section 2 presents an overview on the last advances of the Education solutions for medical doctors. Section 3 proposes a detailed survey of the current state of the art. Section 4 presents the MARIA architecture with a focus on the virtual patients driven by Intelligent Agents (IAs), while finally section 5 summarizes the conclusions and future research activities.

2. EDUCATION OF NEW MEDICAL DOCTORS

In last decades medical training programs and courses are changing in order to face critical issues such as, necessity to improve and simplify students curriculum due to the huge amount of information, major attention to patients safety and quality and to their concern for being subject of students practice. As matter of fact while traditional medical education was historically based on didactic lectures and directed practice on real patient, current program and professional courses are introducing new technologies and methodologies, even due to the fact that young clinicians have more experience with computer-based technologies or video games (Roy et al. 2006). For this reason new solutions and new technologies are required and computer-based clinical simulations are introduced to provide medical students with realistic medical experiences in medical problem solving, diagnoses definition and clinical procedures (Felciano & Parvati 1994). In particular Modeling and Simulation has proved to be a valid approach to improve students learning. In effect since 1980 simulation based training became an opportunity for educators and medical trainees to present different patients cases, to reproduce medicine knowledge and procedures, to test students skills.

Actually different kinds of simulators are used in different educational centers:

- Low-tech/part task simulators, that are models often representing parts of the body and are useful to learn simple procedures (i.e. the use of a prosthetic arm for venepuncture, cannulation or arterial blood gas sampling)
- Simulated/standardised patients, that are well selected and trained actors playing the role of patients; this training is useful to allow students to practice both medical and communication skills
- Computer screen-based simulators, that usually include an interactive patient vital sign screen which responds to user interventions; the major purpose is to allow students to make treatment decisions and then evaluate its effectiveness
- Complex task trainers that are used to train the students in the use of technology- specific equipment and procedures such as intubation or endoscopy
- Realistic patient simulators are also called high fidelity simulators; they are sophisticated computer controlled mannequins that replicate real patients by reproducing symptoms and human behaviors. They can be Instructor driven or Model driven simulators.

Different researches compare the above mentioned techniques: in Australia the Flinders University and Medical Center have carried out a experimentation on three simulation-based training methods for management of medical emergencies (Owen et al. 2006) and in this case full-mission simulation including the use of physical patient simulators appeared to be the most useful training method; other studies highlight the benefits of computer screen-based simulation for procedures learning (Bonnetain et al. 2010), while other ones have not remarked high differences between a computer screen-based simulator and a high-fidelity patient simulator (Nyssen et al.).

Any case training based simulation improve students learning and provide a set of benefits such as:

- Possibility to recreate different medical scenarios and environment
- Students can practice medical procedures without procuring harm to real patients
- Speed up learning in decision making by playing in virtual environment and by operating on virtual or simulated humans reacting as real people
- Development of not only professional knowledge and skills but also attitudes (communication with the patient and relationship management)
- Opportunity for teamwork training in simulated environment
- Opportunity to assess, give feedback and quantitatively measure students performance
- Opportunity to repeat the experience

Many researches highlight the effectiveness of simulation use for medical education: for instance a experimentation was carried out to compare a group of students using the Harvey Cardiology Patient Simulator to practice cardiac bedside skills and a group trained by traditional methods. The simulation group performed at twice as well as the ward group, with only half the training time (Woolliscroft et al. 1987). From January to June 2004 a similar experiment was carried out to train cardiac arrest team response at the Northwestern University Feinberg School of Medicine: two groups of students were considered, one trained by a human patient simulator and the other one traditionally trained. Their performances were assessed based on the adherence to ACLS (advanced cardiac life support) response quality indicators deriving from American Heart Association (AHA) guidelines. Simulator-trained group remarked significantly higher adherence to the quality standards (mean correct responses, 68%) versus traditionally trained residents (mean correct responses, 44%) (Wayne et al. 2008). An additional example of cohort study or control study was developed for training the procedure of central venous catheter (CVC) insertion: ninety-two internal medicine and emergency medicine residents were required to complete a simulation-based mastery learning program in CVC insertion and, by comparing the rates of CRBSI before and after the simulation-based educational intervention over a 32-month period, after the simulator-trained program CRBSI rate significantly decreases from 3.20 infections per 1000 catheter-day to

0.50 infections per 1000 catheter-days (Barsuk et al. 2009). Therefore different cases and examples demonstrate the effectiveness of Simulation Use in Medical Education. A new meta-analysis carried by a medical-education specialist at Mayo Clinic in Rochester has confirmed the effectiveness of technology-enhanced medical simulation in clinical training (Cook et al. 2010). In addition computer connected databases and computer assisted learning (CAL) provide the clinician with huge amount of information for decision-making and patients treating; these solutions, by combining audio and visual data in an interactive framework, recreate clinical situations in an interactive way. In the future these solutions will be more sophisticated and integrated with virtual simulated patients (Schitek et al. 2001).

The authors propose an innovative solution, based on serious game concepts and distributed simulation, to provide medical students with a set of virtual patients to be cared. In this way the student has the opportunity to follow the whole patient cycle life and to evaluate the effects of his clinical decisions. The simulated patient are driven by Intelligent Agents and so are characterized by realistic human factors (i.e. pain, fear, psychological issues etc.) (Bruzzone et al. 2011). In addition this simulator is web- based, so it is possible to access it from every place and from any advice (personal computer, ipad, smartphone, other tablets).

3. STATE OF ART MEDICAL SIMULATION TOOLS AND TRAINING PROGRAM

During the 1930s simulation acquired great popularity due to flight and military applications (i.e. Pilot Trainer by Link Aviation Devices Inc.), while in health care field it spread only later due mainly to the lack of standardization of medical knowledge. The first use of medical simulation is related to a full body mannequin simulator, named Sim1, developed by anesthesia physicians at the University of Southern California to reduce accidents in late 1960s (Lateef 2010). In the same university, in 1963, a neurologist used standardized patients for the first time during the third year neurology clerkship (Wallace 2007) and in 1960 the first task trainer was developed by Laerdal Medical Company: it was named Annie and it was used to practice of mouth-to-mouth breathing (Yasuharu et al. 2009). Few years later it was developed a cardiology patient simulator, Harvey, to practice resuscitation, physical examination, and procedural skills. In 1970s Massachusetts General Hospital produces first computerized clinical encounter simulations. In the early 1990s, more comprehensive anesthesia simulation environments were developed by Gaba's team at Stanford University in partnership with CAE-Link. These environments were the precursor of Medical Education Technologies Inc. (METI) Advanced Human Patient Simulator. In the last years medical simulation spread worldwide and new applications and software were developed, even complex integrated systems (such as TheraSim 360, a multi-tier architecture

encompassing an artificial intelligence engine, an analytics harness, content creation tools and the simulation user interface) in order to reproduce more and more realistic environment and to collect many information and many case studies. For instance Wahlgren et al. proposed a new computerized interactive case simulation system for dermatology teaching; students are able to select a patient and to make questions and propose diagnosis and to obtain feedbacks on their own suggestions that are compared of a specialist (Wahlgren et al. 2006). At University of Leicester Professor Reiko Heckel in collaboration with Dr. John Barry Omara is developing a Virtual Teaching Hospital System to assist medical students in rehearsing the problem solving process and help them to decide what patient information is needed and to determine a possible diagnosis and management. This expert system will provide guidance and direction with the evolving notion of what might support or refute the diagnosis (Heckel et al. 2009).

In particular virtual patients are becoming a good alternative to standard patients for training medical personnel even in mass disaster and crisis management context (Andreatta et al. 2010) due to their efficiency, easy accessibility, interactivity, capability to be personalized for specific purposes and capability to provide feedback (Saleh 2010).

Virtual reality allows to reproduce real environments in a realistic way and to provide a better experience to the students. An example of the use of virtual reality combined with intelligent tutoring systems is represented by URO Mentor (Simbionix, Tel Aviv, Israel) computer-assisted simulator, that overcomes the disadvantages of traditional training methods in ureterorenoscopy and percutaneous nephrolitholapaxy fields. It provides 3D environment and includes several procedures and different types of cases and virtual patients (Knoll et al. 2002). The University of Florida is working to develop highly-immersive virtual human (VH) to train health students (medical, nursing, and physician assistant) on communication and interpersonal skills (Rodriguez et al. 2008).

Wide knowledge about the patient, interactivity with the virtual patient and a realistic environment are critical issues for an effective training in Health Care. For this reason the authors propose an innovative solution based on serious game concept combined with Intelligent Agents driving the virtual patient in order to develop a more immersive and engaging framework where it is possible to experiment medical decisions and procedures. In effect well developed serious games present great opportunities for creating immersive experiential learning environments for decision making training, enabling students to become active learners in a safe benign environment, but one which encourages them to take risks and explore the solution space, with the benefit of immediate feedback, and subsequent review of performance.

Computer-based (serious) gaming is a new field in medical education and has a big potential to become an

important tool for healthcare professionals in order to learn a range of clinical skills. A research conducted in June 2011 highlighted that in term of video games, education, training, gaming and healthcare there are about thirty relevant papers. The studies showed that serious gaming is a stimulating learning method and that students are enthusiastic about its use. In fact in addition to surgical skills, serious gaming is potentially a good method for learning clinical decision-making and patient interaction in this sector (Wit-Zuurendonk & Oei 2011). Innovative studies are in addition focusing on investigating the clinical effectiveness of serious gaming on skills used in patient care.

Obviously it is evident the importance of coupling serious games with human factors in order to guarantee proper fidelity by using Intelligent Agents (IA) reproducing humans interaction and features.

Intelligent agents are very critical in order to develop new games where different players and entities are managed by the computer; from this point of view it is interesting to consider the IA-CGF (Intelligent Agents Computer Generated Forces) developed by Simulation Team (Bocca et al. 2007, Bruzzone et al. 2004-2011). By using Agent Driven Simulation based on Human Behavioral Models in the last 15 years the Research Team of the University of Genoa led by Prof. Bruzzone has created models for addressing different kind of Crisis to be managed, for instance Decision Support and Training for Country Reconstruction and CIMIC operations (Bruzzone A.G., Massei M. Tremori A., et al.,2009, 2010, 2011).

4. THE MARIA ARCHITECTURE AND TECHNOLOGIES

As already mentioned before, in this paper the authors propose an advanced serious game (MARIA) for medical students training and education. The MARIA serious game is based on the joint use of web-based simulation and intelligent agents located on a centralized server. The main idea behind the serious game is to educate medical students to establish a relationship with their patients in order to follow the evolution of the patients healthcare conditions along the entire patients' life cycle.

The architecture of the MARIA serious game is based on a Client-Server architecture where multiple players may simultaneously access and play the game. Figure 1 shows a schematization of the MARIA architecture. The main idea is to design and develop, based on multiple scenarios evolutions, an advanced web-based game obtained by the combination of several parts. In particular, the web based simulation engine implements a layer of intelligent agents playing the role of a virtual patient. The virtual patients are located on a central server and they are able to interact dynamically with students through traditional computers as well as through mobile solutions such as smart-phones and tablets. Multiple MARIA virtual patients can be assigned to each students and they request healthcare assistance 24/7 through several channels (i.e. email, sms).

Therefore for the first time a medical student will be required to interact continuously with the same patients with their own anamnesis, behaviors and pathologies. One of the most important features of MARIA is that the virtual patients' life cycle can be accelerated therefore providing the medical students with the possibility to experience the entire life of a patient within a two-year course, observing his growth, the appearance of diseases, trying to correct attitudes and behaviors in order to support prevention.

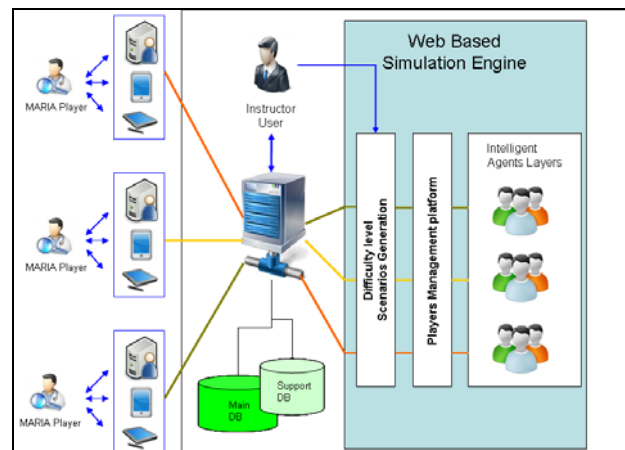


Figure 1: Schematization of the MARIA architecture

Within the game each intelligent agent (patient) can recreate different types of disease and, for each disease different types of evolutionary scenarios. In order to provide medical students with a game that will improve their skills step by step, different difficulty levels can be provided (i.e. low, medium, high). The difficulty levels can be selected by the medical instructor or directly by the players. Therefore based on the disease, on the difficulty level and on the actions taken by the player, the simulation engine rules the evolution of the game (in other words the evolutions of the virtual patients life).

A dedicated Players Management Platform provides the players with the possibility of communications with the other players. This aspect is particularly useful in order to improve and test the team working skills of medical students. The players' management platform also provides to each player feedbacks on the performed actions; every action is traced and in every moment the interaction between the students and the virtual patients can be examined by the teaching committee (medical instructor).

The MARIA architecture will allow to different players to play simultaneously. Each medical student may have multiple virtual patients and he/she will operate as normally done by real medical doctors in real-life situations. At the beginning of the game each player is provided with a description of his/her duties within the game. Each player will start to interact with one or more patients having the possibility to act based on the information received by the patient (anamnesis), results of medical analysis and experience therefore positively (or negatively) affecting the patient health situation.

Each player will be therefore provided with a set of possible actions and by executing these actions the player interacts with the virtual patients and, if required, with other players.

By interacting with the virtual patients, the medical students will learn how crucial is the doctor-patient relationship and the possibility of delivering treatment disease over the time (and as function of patient's needs)

During the game and at the end of the game each player can be informed about his learning performances. Such evaluation will be based on key performance indicators that will consider diagnostics, decision making and treatment protocols. After using the game, the medical students may review their actions, examine the negative and positive choices and collaboration activities and assess the dynamics of the relationship with his/her virtual patients.

A particular modeling effort is required to define and implement (within the web based simulation engine) the intelligent agents recreating the role of the virtual patients. The following section provides the reader with a more accurate description of the MARIA virtual patient.

4.1 The MARIA Intelligent Agents

The intelligent agents should be able to behave like real patients therefore providing consistent responses to the treatment protocols proposed by the medical students. In addition the intelligent agents should provide ability to communicate continuously with the medical students both when the situation is getting worse and when healthcare conditions are improving therefore asking advices to remain in good shape as the time goes by. Therefore the intelligent agents implemented within the MARIA architecture have the following characteristics:

- **Autonomy:** they are able to act based on their own capabilities
- **Social Ability:** they are able to interact with other agents and with the context
- **Reactivity:** they are able to react to the events
- **Pro-activeness:** based on their adaptive learning they act to achieve objectives and accomplish their tasks

In addition an intelligent agent is conceptualized and implemented using concepts that are usually applied to humans and therefore it should be able to reproduce human behaviors such as fear, aggressiveness, stress etc. In particular in the case of a virtual patient, the agent is able to reproduce a real patient with his/her needs, anamnesis, diseases, pains and feelings.

The concept to develop a virtual life form that establishes a relationship with the human player devoted to taking care and cure it, was already experimented (even if for different purposes) in the past (Trindade et al. 2002); this approach demonstrates to be a very effective engagement mechanism able to generate strong empathy between the player and the virtual life (Huang and Tettegah 2010). A very good example of these concepts was extensively applied in '90s with the

Tamagotchi application: a handheld digital pet, developed in Japan by Akihiro Yokoi of WiZ and Aki Maita of Bandai. This application was integrated into in a small simple egg-shaped computer and it allows following the whole pet life cycle from the birth to the death. Therefore, the player has to feed and follow pet needs in order to make him happy during its growth. In particular, the tamagotchi grows based on the care the player provides. The pet can also "die" due to poor care, old age or sickness.

The solution developed by authors is, in some way, inspired to Tamagotchi game in order to provide medical students with an advanced framework in which they can improve their skills in patients care.

The virtual patients interact with the students exactly as a real patient does: he can make a phone call, send an e-mail or a message to the student who will feel exactly as a medical doctor feels when is managing his patients. So students can live a real clinical experience having the responsibility to decide what do and when; for instance if it is correct to prescribe examinations, specialist consultations and to advise the patient about serious as well as trivial problems.

The target of the pedagogical aspect of the game is completely open: the intelligent agents can be tuned for supporting the education and training of graduate students or general practitioner (postgraduate); the virtual patient can be a patient that suffers of every kind of disease or even a person in good shape that aims at maintaining is good health conditions by means of prevention care and life style advises. The MARIA program could be aimed to check medical ability, relationship with patients, economic aspects of medical practice, adequacy of diagnostic or therapeutic course.

5. CONCLUSIONS

The paper proposes an innovative approach based on serious game concept and Intelligent Agents driven simulation to enhance education in Medical Programs and Courses. The solution proposed by authors (called MARIA) allows students following the whole life cycle of a virtual patient and check the effects of their medical decisions and suggestions. The MARIA game guarantees a virtual engaging environment and an easy access by web both by using traditional computer and mobile solutions (smart-phones, tablets, etc.).

In order to test the effectiveness of the developed solutions, the authors are now planning an exercise by using the MARIA serious game by involving a group of medical students (students at their sixth year of medical school) for two days. The students will receive in care a virtual patient; based on the anamnesis, the student has to take in care of the virtual patient by taking the correct decisions in terms of medical protocols and procedures also trying to obtain from the virtual patient all the information that could be critical for the patient healthcare conditions. The purpose of the exercise will be to measure the approach effectiveness in term of

learning improvement with particular focus to communication and interpersonal skills.

REFERENCES

- Andreatta PB, Maslowski E, Petty S, Shim W, Marsh M, Hall T, Stern S, Frankel J., (2010) Virtual reality triage training provides a viable solution for disaster-preparedness *Acad Emerg Med.* 2010 Aug;17(8):870-6, Michigan Medical School, Ann Arbor, MI, USA.
- Barsuk J. H., Cohen E. R., Feinglass J., McGaghie W. C., Wayne D. B. (2009) Use of Simulation-Based Education to Reduce Catheter-Related Bloodstream Infections FREE, *Arch Intern Med.* 2009;169(15):1420-1423.
- Bocca E., Pierfederici, B.E. (2007) "Intelligent agents for moving and operating Computer Generated Forces" Proc. of SCSC, San Diego July
- Bonnetain E., Boucheix J. (2010), Benefits of computer screen-based simulation in learning cardiac arrest procedures, *Medical Education* 2010: 44: 716–722
- Bruzzone A.G. (2008) "Intelligent Agents for Computer Generated Forces", Invited Speech at Gesi User Workshop, Wien, Italy, October 16-17
- Bruzzone A.G. (2009) Intelligence and Security as a Framework for Applying Serious Games, Proceedings of Serixgame, Civitavecchia, November
- Bruzzone A.G. Tremori A., Massei M. (2011) Adding Smart to the Mix, Modeling Simulation & Training: The International Defense Training Journal, 3, 25-27, 2011
- Bruzzone A.G., Cunha G., Elfrey P., Tremori A. (2009) "Simulation for Education in Resource Management in Homeland Security, Proceedings of SumemrSim2009, Istanbul, Turkey, July 13-16
- Bruzzone A.G., Massei M. (2007) "Polyfunctional Intelligent Operational Virtual Reality Agent: PIOVRA Final Report", EDA Technical Report, Genoa
- Bruzzone A.G., Massei M. (2010) "Intelligent Agents for Modelling Country Reconstruction Operation", Proceedings of AfricaMS 2010, Gaborone, Botswana, September 6-8
- Bruzzone A.G., Massei M. Tremori A., Bocca E., Madeo F., Tarone F, (2011) "CAPRICORN: Using Intelligent Agents and Interoperable Simulation for Supporting Country Reconstruction", Proceedings of DHSS2011, Rome, Italy, September 12 -14
- Bruzzone A.G., Massei M., Tremori A. (2009) "Serious Games for Training and Education on Defense against Terrorism" - NATO MSG-069 Symposium Use of M&S in: Support to Operations, Irregular Warfare, Defence Against Terrorism and Coalition Tactical Force Integration", Brussels, Belgium October 15, 16
- Bruzzone A.G., Petrova P., B.M., B.C. (2004) "Poly-functional Intelligent Agents for Computer Generated Forces", Proceedings of Wintersim2004, Washington DC, December
- Chin, H.L.and Cooper, G.F., (1989), Case-based tutoring from a medical knowledge base. *Comput Methods Programs Biomed.* 1989. 30(2-3): p. 185-98.
- Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Acad Med.* 2010;85:1589–1602.
- Cundick, R., et al. (1989), ILIAD as a patient case simulator to teach medical problem solving, 13th Annual Symposium on Computer Applications in Medical Care, Washington, DC: American Medical Informatics Association
- Dichter, M.S., Greenes, R.A., and Bergeron, B.P. (1991), Authoring multimedia clinical problemsolving exercises with CaseBase, Proc. of Annu Symp Comput Appl Med Care.
- Edelbring S, Fors U, Hindbeck H, Ståhle M., (2006), Evaluation of an interactive case simulation system in dermatology and venereology for medical students, *BMC Med Educ.* 2006 Aug 14;6:40
- Felciano RM and Dev P. (1994) Multimedia clinical simulation based on patient records: authoring, user interface, pedagogy, Proc Annu Symp Comput Appl Med Care. 1994:59-63.
- Felciano, R.M., Daane, S.P., and Dev, P.(1994), Clinical Pearls and Short Rounds: Two Shells for Multimedia Case Presentations, AMIA Spring Congress, San Francisco, CA: AMIA.
- Friedman, C.P., France, C.L., and Drossman, D.D., (1991) A randomized comparison of alternative formats for clinical simulations. *Med Decis Making.* p. 265-72.
- Greco M.; Murgia G., (2007), "Improving negotiation skills through an online business game", Proceedings Of The European Conference On Games-Based Learning Pages: 97-104 Published: 2007
- Heckel R., Donyina A., Omara J. B. (2009) Virtual Teaching Hospital System (VTHS) Project aims to transform medical training, University of Leicester Press
- Huang W. D. , Tettegah S., (2010) Cognitive Load and Empathy in Serious Games: A Conceptual Framework , *Gaming and Cognition: Theories and Practice from the Learning Sciences*, IGI Global, ch006
- Lateef F. (2010) Simulation-based learning: Just like the real thing, *Emerg J., Trauma Shock.* 2010 Oct-Dec; PMID: PMC2966567
- Lyon, H.J., et al., PlanAlyzer, (1992) An Interactive Computer-assisted Program to Teach Clinical Problem Solving in Diagnosing Anemia and Coronary Artery Disease. *Academic Medicine.* 67(12): p. 821-828.
- Mayne A. (2002) Comparison of the training value of two types of anaesthesia simulators: computer

- screen-based and mannequin-based simulators. *Anesth Analg* 2002; 94:1560–5.
- Michel MS, Knoll T, Köhrmann KU, Alken P., (2002), The URO Mentor: development and evaluation of a new computer-based interactive training system for virtual life-like simulation of diagnostic and therapeutic endourological procedures, *BJU Int.* 2002 Feb;89(3):174-7.
- Miller, R.A., et al., (1986) The Internist-1/Quick Medical Reference Project-Status Report. *The Western Journal of Medicine*, 1986. 145 (6): p. 816-22.
- Nyssen AS, Larbuisson R, Janssens M, Pendeville P, Mayné A., A comparison of the training value of two types of anesthesia simulators: computer screen-based and mannequin-based simulators, *Anesth Analg.* 2002 Jun;94(6):1560-5, table of contents. PubMed PMID: 12032027
- Okuda Y., Bryson E. O., DeMaria S. Jr., Jacobson L., Quinones J., Shen B., (2009) Levine A. I. The Utility of Simulation in Medical Education: What Is the Evidence?, *Mount Sinai Journal of Medicine* 76:330–343
org/wallace.htm (Originally published in *Caduceus*. 1997;13:5-28.).
- Owena H., Mugford B., Follows V., Plummer J. L., (2006) Comparison of three simulation-based training methods for management of medical emergencies, PubMed PMID: 16987587 April 2006
- Perper, EJ., Felciano, R.M., and Dev, P., (1993) Real Problems: A Layered Approach to Constructing a Patient Simulation. in 1 7th Annual Symposium on Computer Applications in Medical Care. 1993. Baltimore, MD: American Medical Informatics Association.
- Rodriguez, H., B. Lok, D. Lind, D. Beck, "Audio Analysis of Human/Virtual-Human Interaction" *Intelligent Virtual Agents 2008*, Sept. 1-3, Tokyo, JP.
- Roy MJ, Sticha DL, Kraus PL, Olsen DE (2006), Simulation and virtual reality in medical education and therapy: a protocol. *Cyberpsychol Behav.* 9(2):245-7
- Saleh N., (2010), The Value of Virtual Patients in Medical Education *Annals of Behavioral Science and Medical Education* 2010, Vol. 16, No. 2, 29-31 2010 by the Association for the Behavioral Sciences and Medical Education
- Schitteck M, Mattheos N, Lyon HC, Attström R. (2001). Computer assisted learning. A review. *Eur J Dent Educ.*, Vol 5(3). pp. 93-100.
- Trindade, J., Fiolhais, C., & Almedia, L. (2002). Science learning in virtual environments: A descriptive study. *British Journal of Education Technology*, 33(4), 471-488.
- Wallace P. (2007) Following the threads of an innovation: the history of standardized patients in medical education. *Association of Standardized Patient Educators Web site*. Available at: www.aspeducators.
- Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie (2008) WC, Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest.* 2008 Jan;133(1):56-61. Epub 2007 Jun 15. PubMed PMID: 17573509
- Wit-Zuurendonk L. D.; Oei S. G (2011), Serious gaming in women's health care, *Bjog-An International Journal of Obstetrics and Gynaecology*, Volume: 118 Special Issue: SI
- Woolliscroft JO, Calhoun JG, Tenhaken JD, Judge RD. (1987), Harvey: the impact of a cardiovascular teaching simulator on student skill acquisition. *Med Teach.* 1987;9:53–7, PubMed.

AGENT BASED SIMULATION MODEL FOR OBESITY EPIDEMIC ANALYSIS

Agostino Bruzzone^(a), Vera Novak^(b), Francesca Madeo^(c)

^(a) DIME, University of Genoa - URL www.itim.unige.it

^(b) BIDMC, Harvard Medical School- URL <http://www.bidmc.org/SAFE>

^(c) M M&S Net - URL www.m-s-net.org

^(a) agostino@itim.unige.it, ^(b) vnovak@bidmc.harvard.edu, ^(c) francesca.madeo@m-s-net.org

ABSTRACT

This research proposes a simulation model based on Intelligent Agents to reproduce human behavior and its influence on the evolution and impact of obesity epidemics.

Based on our previous experiences on Human Behavior Models, we have developed a Library including Intelligent Agents for Computer Generated Forces (IACGF Libraries), designated to reproduce complex scenarios, with particular attention to non-conventional frameworks mediated by human behaviors on progression of obesity epidemics in the world. In this paper, we compared two scenarios (in Italy (obesity prevalence ~10%) and in the U.S. (obesity prevalence ~35%) based on different social and cultural condition in order to test, validate and calibrate the simulation model. The authors are interested in analyzing how simulation model results change by considering different social and cultural conditions in different countries.

Actually Obesity is a real big problem for both USA and European countries, so it is necessary to take under control this phenomenon and above all to provide people and government with simulation models in order to promote specific actions and to guarantee population healthy and even less social costs.

Keywords: Simulation, Intelligent Agents, Human Behavior Modeling, Health Care, Obesity

1. INTRODUCTION

Obesity is becoming an increasingly common and growing public health problem in America and in European Countries. Currently, two of three Americans are considered overweight, while one out of three people in America is obese.. According to the Centre for Disease Control report, U.S. Obesity Trends 1985-2007, in 1985,

there were only 8 states in the U.S. with prevalence of obesity ~ 10%, while in 2010, prevalence of obesity increased dramatically, and at least 35.7% of the adult population was obese, affecting all 50 states and men and women (Figure 1a,b). Even more importantly, obesity affects 16.9% of U.S. children and adolescents (Ogden et al. 2010). There was no change in prevalence of obesity from 2009-to 2010, and adults over age 60 were more likely to be obese than younger adults. The question how the growing prevalence of obesity in younger population will affect the overall trend in epidemic.

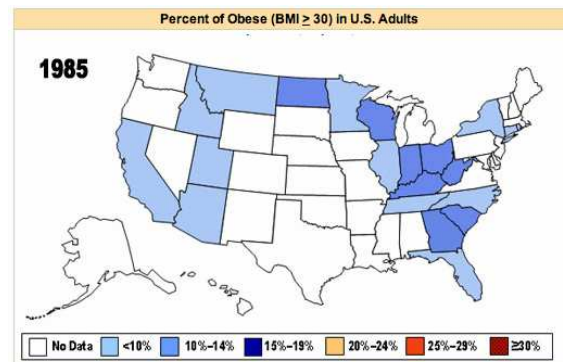


Figure 1a. Obesity Rates in United States in 1985

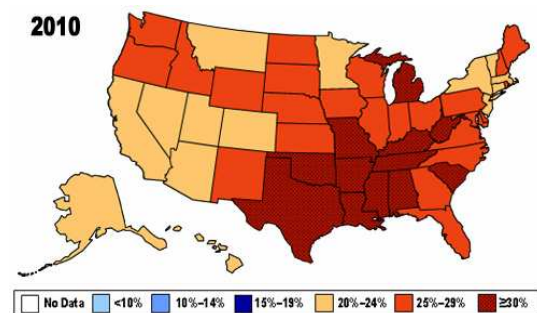


Figure 1b. Obesity Rates in United States in 2010

Obesity is a global problem. Obesity epidemic (Wolf et al. 2007) has been increasingly spreading around the world in the past three decades, involving countries that never in the past reported obesity in their population. Recently, special scientific series published in leading medical journals such as Lancet examine the global obesity pandemic (Swinburn et al. 2011; Gortmaker et al. 2011) and emphasize that obesity is a global issue and these reports call upon for government interference to turn around the epidemic with cost-effective programs and policies, supported with continuous evaluation and monitoring (Wang et al. 2010; Bruzzone 2004). In European Region, since 1980s, obesity prevalence has tripled in many countries and the numbers of those affected continue to rise at an alarming rate, particularly among children (Figure 2 and Figure 3).

THE GLOBAL OBESITY PROBLEM

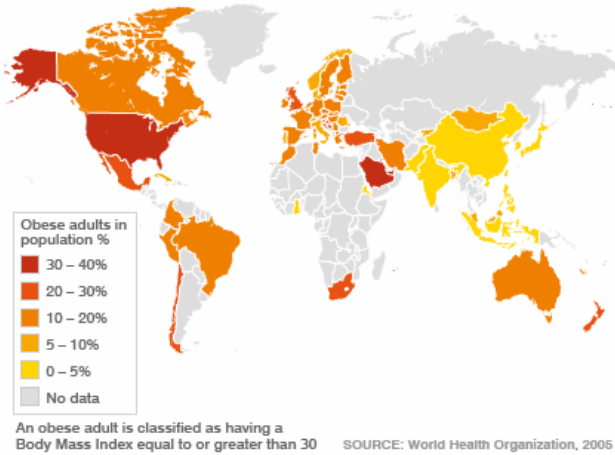
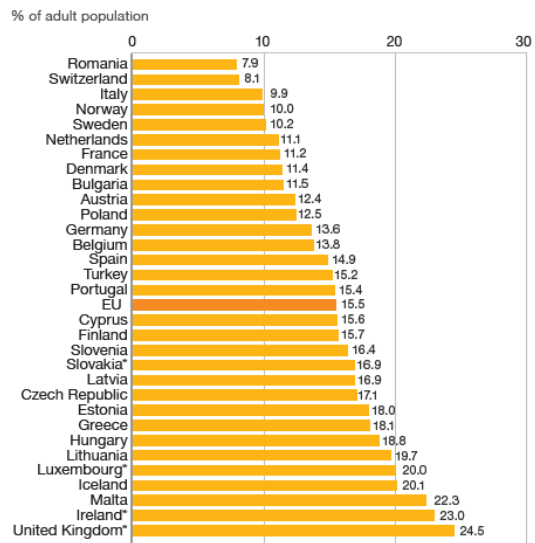


Figure 2. Obesity as a Global Problem

The European Commission and the Organisation for Economic Co-operation and Development (OECD) published the Health at a Glance Europe 2010 report to encourage better eating habits in children. This report emphasized that the rate of obesity has more than doubled over the past 20 years in most European countries and that just over 50 percent of Europeans are now either overweight or obese (OECD 2010).

Obesity rates among adults, 2008



*Ireland, Luxembourg, Slovak Republic and United Kingdom figures are based on health examination surveys, rather than health interview surveys. Source: OECD Health Data 2010, Eurostat Statistics Database, WHO Global Infobase.

Figure 3. Obesity rates in Europe

The current lifestyle based on a diet with excessive calories intake (fast food and carbonated beverages) and lack of an adequate physical exercise has widely contributed to the obesity increase.

The consequences affect the whole society from a social and economic point of view i.e. health care burden, increased consumption demands and special needs to adapt to normal life in many different sectors i.e. furniture, transportation, and consumer spending, increased waste and lifestyle cost. Furthermore, obesity problem is related to many other factors such as urban environments, transport systems, socio-economic conditions, and cultural factors, population aging, etc. Obese people are susceptible to various health problems i.e. cardiovascular diseases, certain types of cancer, and type 2 diabetes, musculoskeletal, neurological and psychiatric disorders, increased overall morbidity, disability and mortality.

The impact of obesity is evident in health care more than in other sectors, because of the strong influence of stochastic factors and very complex correlations that make analysis of large scale data difficult without using modeling and simulation. Therefore obesity poses an important and very complex problem, that is suitable for modeling and simulation that may take into account into account complex nature of this condition including social a cultural conditions and economical aspects.

The authors have developed several simulation models reproducing human behavior in different frameworks (i.e. for country reconstruction operations, as well as, port security and safety) (Bruzzone et al. 2007). We have developed a set of Libraries, named IA-CGF (Intelligent Agents for Computer Generated Forces), in order to

simulate units (i.e. police, gangs, terrorist, etc.), as well as, non-conventional frameworks (i.e. food distribution and humanitarian support, disaster relief). Our goal is to use these libraries to address obesity problem world wide, and thus to provide theoretical support for decision makers most cost-effective approaches to curb obesity epidemic. This paper is aimed to test and validate BACCUS (Behavioral Advanced Characters and Complex System Unified Simulator) Model by reproducing two different scenarios, one related to Massachusetts State in the U.S. and second related to an Italian Region, by taking into account the mutual influence of different factors related to physiological and psychological issues, behavioral aspects and regional/ethnic/social/geographical and economical factors.

This kind of model may be also helpful to simulate and predict the impact of behavioral, social and economic interventions aimed to prevent further development of obesity epidemic and to curtail its costs. Therefore, BACCUS could become a strategic tool for designing preventive strategies and decision as well as evaluating the impact of actions and countermeasures of public and private institutions and organizations on such a critical sector of the health care.

2. THE OBESITY EPIDEMIC AND CRITICAL FACTORS

The obesity is defined as a medical condition in which an excess proportion of body fat accumulates in human body; this causes various health problems. Obesity is most commonly defined in term of Body Mass Index or BMI.

A person is considered obese when he or she has 20% higher than their normal body weight. According to the World Health Organization (WHO) an overweight person has a BMI between 25 and 29.9, and obesity is marked at a BMI of 30 and above.

Where:

BMI Body Mass Index

W Weight [kg]

H Height [m]

Extreme form of obesity is termed as “morbid obesity” which means either a person is more than 100 pounds over normal weight, 50%-100% over normal weight, has

$$BMI = \frac{W}{H^2}$$

a BMI of 40 or above, or is enough overweight to suffer from various health concerns. The main causes of obesity are related to:

- Diet or consumption of excessive calories

- Age: With age the human body’s power to metabolize food decreases, leading to gaining of weight.
- Gender: gender difference in obesity rate is disappearing. In general women have slightly a lower resting metabolic rate than men, more body fat and less muscle mass.
- Physical activity: Lack of physical activity increases energy storage and accumulation of fat and weight gain. .
- Genetics: genes play an important role in the prevalence of obesity. There is a good chance of 75% to be fat or slender if a person has biological parent who is obese or normal weight. This predisposition, however, is likely to be enhanced by eating habits and social factors during childhood.
- Psychological factors: eating habits and obesity run parallel and, similarly, potentiate factors and habits that lead to overeating. Many people tend to eat more when they are lonely, bored, sad, depressed or angry.
- Illness: some disorders like hypothyroidism, depression, and some rare diseases of the brain, which are mostly hormonal diseases, may lead to obesity.
- Medication: certain medications may be associated with weight gain.

Therefore, obesity is a multifactorial process that results from interactions among the individual health status, functional and social habits, social networks, education and other factors that cannot be predicted from a single variable (i.e. body mass) but requires nonlinear modeling of multiple variables and their interactions.

Mathematics and Statistics provide different modeling tools that can be used to measure and control obesity.

For example, a model predicting how the body composition changes in response to what it ate, developed by K. Hall at NIDDK (National Institutes of Diabetes and Kidney Diseases) (Hall et al. 2011).

This is a mathematic model of a human being that contains several variables (i.e. height, weight, food intake and composition, exercise). The model can predict what a person’s weight, dependent on their body size, food intake and exercise. However, the model is complicated: hundreds of equations; and new attempts are being made to simplify it to a single equation.

In addition, Dragone and Savorelli (2010) underlined the importance of psychological factors and of the ideal body weight concept. They analyzed obesity and anorexia together and demonstrated that increasing the ideal body weight is socially desirable. Therefore, if a majority of population is overweight, but an ideal body weight is low and therefore difficult to achieve, then increasing the ideal body weight may be socially desirable and would reduce social pressure. Their model is based on forward-looking agent depending on food consumption, health condition, and the conformity of body weight to an ideal

weight. So, the agent is aware of how food consumption affects body weight, and it explicitly takes this information into account when choosing how much to eat (Dragone & Savorelli 2010).

In this study we propose a model based on Intelligent Agents, that include the following characteristics of the study population:

- Social Status and employment status: in our sample, including 307 adults from Massachusetts, obese people are both employed or unemployed, while retirees have higher percentage of severe obesity (Figure 4).

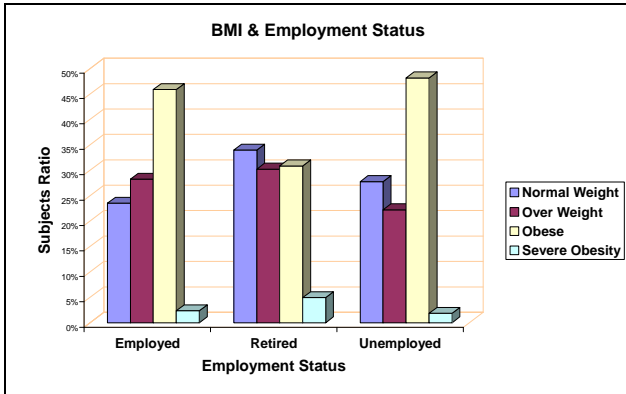


Figure 4. BMI & Employment Status

- Gender: women had higher average BMI in this sample (Figure 5).

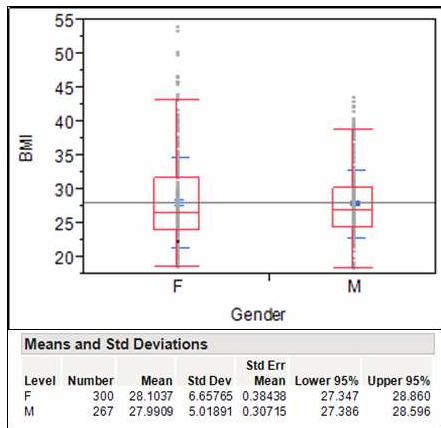


Figure 5. BMI & Gender

- Age: Our analysis shows that obesity is becoming more prevalent in younger people e; in fact on average, people younger than 39 years are more overweight than elderly. The age range the most affected by obesity is between 40 and 70 years old (Figure 6).

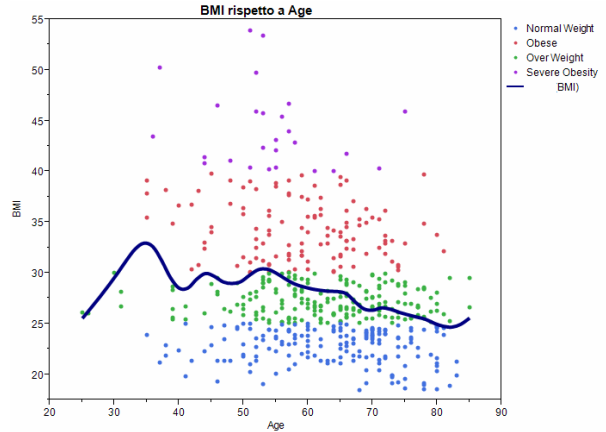


Figure 6. BMI & Age

- Ethnicity: majority (97%) subjects in our samples were not Hispanic or Latinos. However, there was a trend for suggesting that American Native and African American are more prone to obesity.
- Religion: no information about religion was collected in our sample. The model however includes this variable by considering statistics related to Ethnicity and Race on Massachusetts area.
- Education: the average duration of schooling was from 11 to 20 years (Figure7), but there was no clear trend in relation between obesity and education.

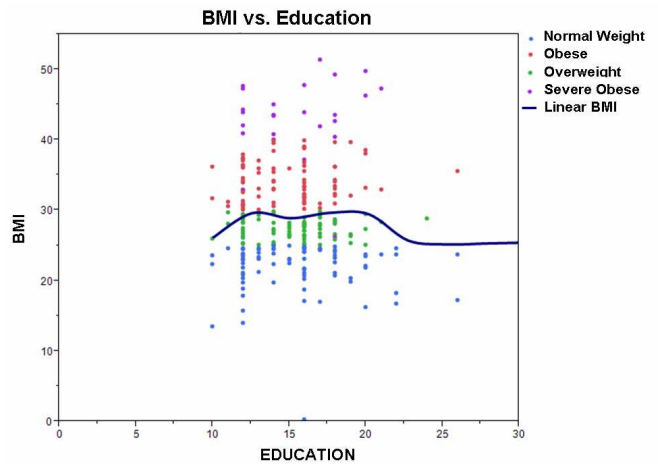


Figure 7. BMI and Education

Furthermore, the model considers other variables in order to simulate the social network, i.e. marital status, Figure 8 shows that majority of obese people are not married, which may indicate underlying social or/and psychological issues.

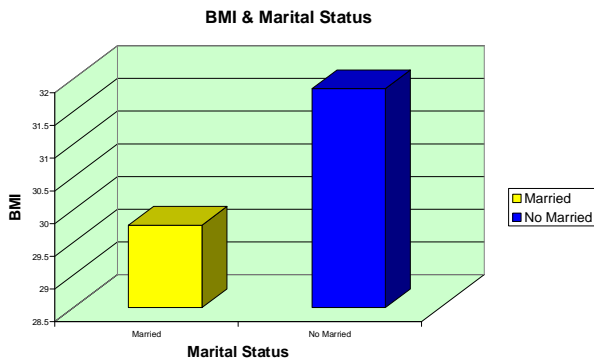


Figure 8. BMI and Marital Status

The data collection and analysis has involved many other parameters in order to verify correlations with BMI trend. In particular among the others:

- Previous Tobacco Use and Current Tobacco Use and the related quantitative variable (i.e. Tobacco Pack Years)
- Previous Alcohol Use and current Alcohol Use and the related quantitative variable (Alcohol Dose/Week)
- Family History: related to cancer, heart disease, hypertension, diabetes, Stroke; for instance by considering Bother Decease number it's interesting the result: people with more than one brother decease, have on average a higher BMI (see figure below):

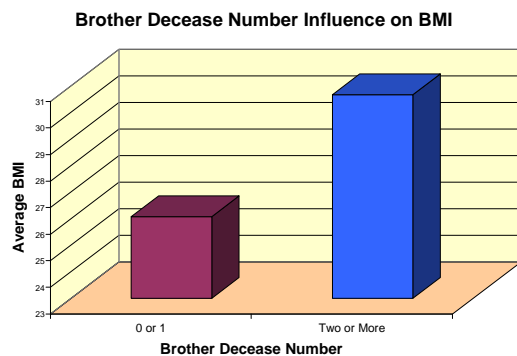


Figure 9. BMI and Brother Decease Number

- Patient Medical History in term of cancer, heart disease, hypertension, diabetes, Stroke and Falls
- Medications (Antiplatelets, Anticoagulant, Antihyperlipidemic, Antiparkinsonian, Statins, Estrogen, ACE Inhibitors, ARBS, Beta Blockers, Diuretics, CA ++, glucose mg/dL, Insulin, Cholesterol mg/dL, triglycerides mg/dL, hemoglobin A1C%, heart rate, systolic and diastolic blood pressure at baseline)
- Other parameters: walking speed and brain volumes measure by on magnetic resonance imaging (MRI) global intracranial volume(ICV), gray matter volume (GM), white matter volume (WM),l cerebrospinal

fluid volume (CSF) and white matter hyperintensities volume (WMHs).

These analyses are based on the studies conducted in the Syncope and Falls in the Elderly (SAFE) laboratory at the Beth Israel Deaconess Medical Center. These data were collected during different five studies including clinical tests and questionnaires from patients. All the data were reviewed and validated for consistency, missing and outlying values. The analytical approach included the following steps:

1. Distribution, Variance, Standard Deviation and Average Value for each variable to identify possible outliers
2. Estimating BMI respect each variable of interest
3. Multivariate regression analysis

Even we have found correlations between BMI and each involved variable (i.e. education, age, gender etc.), it is difficult to define a linear correlation combining all these variables. This is due to the fact that BMI is influenced by too many factors interacting with each other, that are related to human behavior and stochastic elements. Therefore, it is important to introduce a simulation approach to investigate obesity phenomenon the effects of different cultural and social factors on the trend of this epidemic. .

3. THE SIMULATION MODEL FOR OBESITY EPIDEMIC

We propose a model based on Intelligent Agents driving people acting and moving in the area in order to simulate the Obesity epidemic evolution in a specific region and scenario. This model is based on a stochastic discrete event simulation. The simulator, named BACCUS, allows to simulate the population behaviour and to reproduce different scenarios and geographical areas. The goal is to reproduce the scenario of Massachusetts, with particular focus on Boston Area in order to test the simulator and compare its results with the real data, and then compare these results with a region in Italy.

The input data are provided by statistical analysis developed at the SAFE Lab that was combined with statistics by the World Health Organization.

The simulator allows to generate over 600.000 agents based on statistical distributions of social and cultural characteristics, and then it allows to generate a social network including families relationships, friendships and working relationships. These links are generated based on compatibility algorithms and social algorithms allowing to compare the agents based on their social and cultural characteristics in order to measure a level of compatibility; if this level overcomes a predefined threshold the connection link is generated (Giribone & Bruzzone 1999).

The BACCUS model, is defined as a IA-CGF framework that takes into account population behaviour and obesity factors. IA-CGF is an innovative solution

provided by DIPTeM Genoa University and MAST srl, and it includes behavioural libraries for Non-Conventional Frameworks (i.e. disaster relief, humanitarian support, natural disasters etc.) to reproduce human behaviour in town or regions (Bruzzone et al. 2008). These models were designed to study evolution of obesity epidemic (Avalle et al. 1999), analyzing urban disorders (Bruzzone et al. 2006) and country reconstruction (Bruzzone & Massei 2010).

The IA-CGF Libraries include the following Human Factors (Figure 10):

- Physiological Characteristics
- Social Factors
- Political Factors
- Ethical and Tribal Factors
- Religious Factors
- Cultural Factors

The BACCUS model provides the following functionalities:

- To generate the population
- To generate the families
- To generate the social network
- To visualize zones characteristics and statistics
- To visualize the operative status of agents
- To set up population characteristics
- To provide report on specific parameters

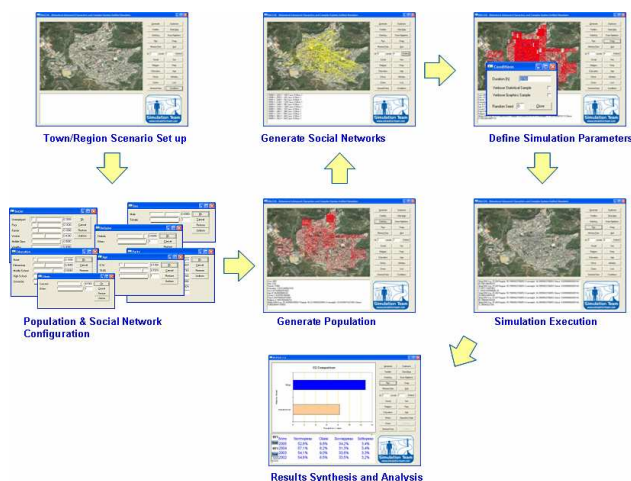


Figure 10. Model Functionalities and Key Elements

The population is grouped in zones and is characterized by different parameters (social, religious, ethnic).

Based on these parameters the population is generated randomly respecting original statistics on the geographic region, but even considering the different groups characteristics with their specific structure (i.e. some ethnic group have different social status statistics respect other one living in the same region); this process generates the people agents; in addition inside the region it is possible to define Zones as objects that have affinities with the main factors and so the people agents are distributed geographically in term of living and

working locations in consistency with their affinities; it is possible to overlap zones with different affinities in order to represent inconsistent mixes of different groups of the population.

Daily regimen of an individual is defined as well. Each individual operates through the whole day by getting up, going work, working, having lunch, having relax, doing exercise, having diner, sleeping.

The model includes relationships among obesity and people's behavior based on data collected and analyzed at the SAFE Lab. In particular the following aspects are included in the model:

- Children, having an obese parent, increases the risk of by 30%.
- Not-married, being single increases the risk of obesity.
- BMI decreases at advanced age (>70 years).
- Obese people tend to have obese friends.
- Exercise: regular exercise three times a week for 40 minutes negatively correlates with BMI.
- Alcohol: moderate to severe alcohol consumption increase the risk of obesity.
- Morbidity and mortality: obesity people have higher morbidity and mortality.

The authors propose to implement two different scenarios in BACCUS model in order to test and validate it. In fact dealing with human modeling the verification and validation of the simulator, as well as data validation, collection and analysis is critical in order to test hypotheses about the conceptual model of behavior leading to obesity epidemic.

4. COMPARATIVE RESULTS ANALYSIS

The model results are based on two different scenarios: an Italian Town including about 40.000 people (Figure 11) and Boston Area scenario including over 600.000 agents (Figure 12), with obesity distribution collected by the SAFE lab studies. Target functions include the different obesity classes and the average BMI:

- Normal Weight
- Overweight
- Obese
- Average BMI

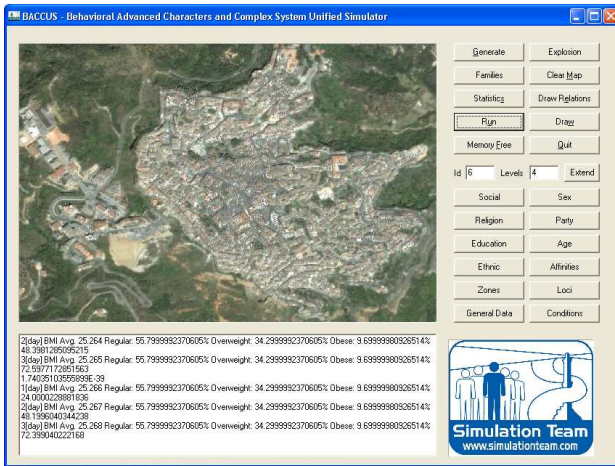


Figura 11. BACCUS Model Interface: Italian Town

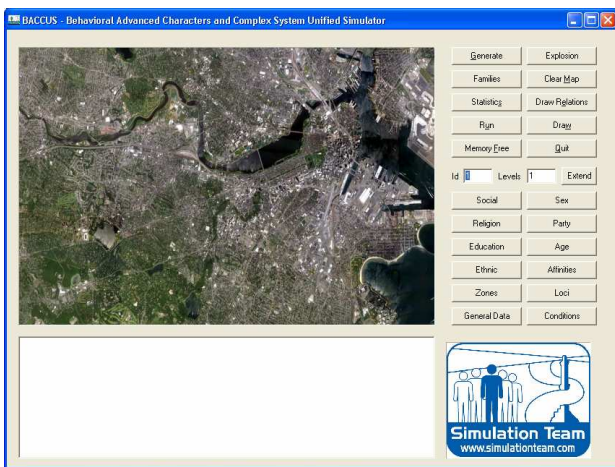


Figure 12. BACCUS Model Interface: Boston Area Scenario

The VV&A (Verification, Validation & Accreditation) of BACCUS is based analysis of MSPE (Mean Square pure Error analysis) as measure of the variance of the target functions among replicated runs over the same boundary conditions; by this approach it becomes possible to identify the number of replications and the simulation duration able to guarantee a desired level of precision; MSPE values in correspondence of these experimental parameters determines the amplitude of the related confidence band:

$$MSPE^m(t, n_0) = \frac{\sum_{i=1}^{n_0} \left[Sr_i^m(t) - \frac{1}{n_0} \sum_{j=1}^{n_0} Sr_j^m(t) \right]^2}{n_0}$$

$$CBA^m(t, n_0, \alpha) = \pm t_{\alpha, n_0} \sqrt{MSPE^m(t, n_0)}$$

- t simulation time
- m m-th target function estimated by the simulator
- no number of replications with same boundary conditions and different

- random seeds
- $Sr_k^m(t)$ m-th target function value at t time of the k-th replicated simulation
- $MspE^m(t, n_0, \alpha)$ Mean Square pure Error at t time and with no replications for the m-th target function
- α percentile
- $CBA^m(t, n_0, \alpha)$ Confidence Band Amplitude at t time, with no Replications for the m-th target function

In fact the MSPE allows to quantify the experimental error due to influence of the stochastic components. In the figure below it is reported the result for the first scenario: the variance of all the target function reach steady state situation over a reasonable number of replications and over a time duration for about a one year period . This confirms that simulator provides consistent results, when a situation is a stable with capability to define the confidence band for estimating the obesity target functions.

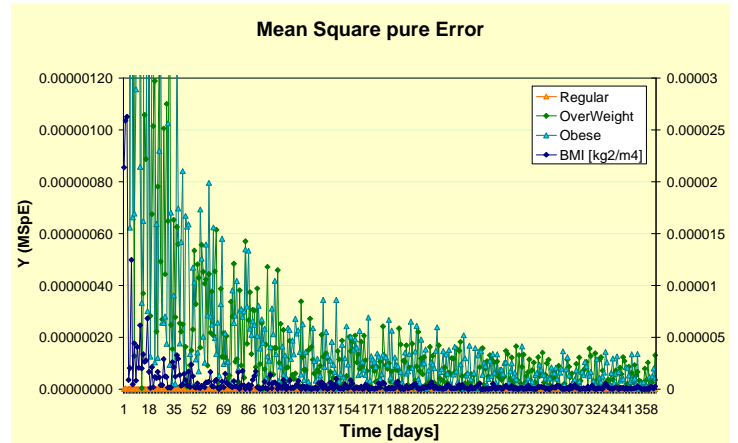


Figure 13. BACCUS VV&A based on Dynamic Analysis of Mean Square pure Error on Population Obesity Classes in Italian Scenario

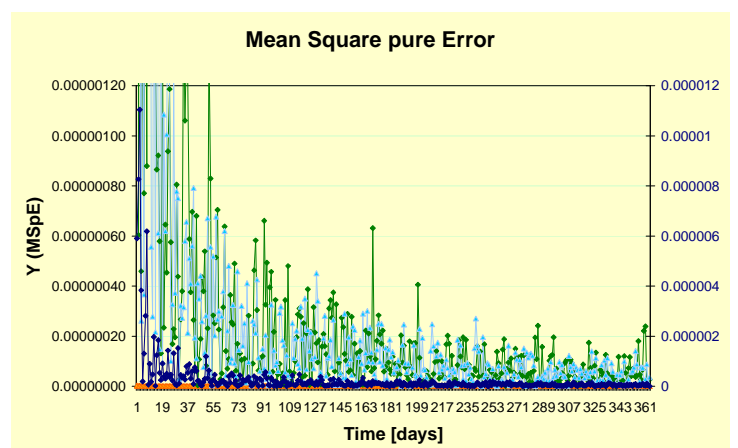


Figura 14. BACCUS VV&A based on Dynamic Analysis of Mean Square pure Error on Population Obesity Classes in Boston Scenario

Figure 13 and Figure 14 show the Mean Square pure Error trend for both the scenarios in Italy and in Boston. In the first one the experimental error is correctly evaluable after about 120 replications, while in the second one the optimal number of replication is about 199.

At the end of the Simulation a report is available to analyze how the obesity evolves day by day:

6[day]	BMI Avg. 22.793	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
7[day]	BMI Avg. 22.794	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
8[day]	BMI Avg. 22.795	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
9[day]	BMI Avg. 22.796	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
10[day]	BMI Avg. 22.797	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
11[day]	BMI Avg. 22.798	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
12[day]	BMI Avg. 22.799	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
13[day]	BMI Avg. 22.8	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
14[day]	BMI Avg. 22.801	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
15[day]	BMI Avg. 22.802	Regular: 14.7000007629395%	Overweight: 55.5999984741211%	Obese: 29.7000007629395%
16[day]	BMI Avg. 22.803	Regular: 14.2999992370605%	Overweight: 56%	Obese: 29.7000007629395%

Figure 15. Extract from Simulation Final Report

Considering the two scenarios and replicating the simulation for more years the results indicate that the trends of obesity growing in Italy follow the USA Obesity Trend, but more slowly. These preliminary results indicate that it is possible to simulate and validate trends in population behavior related to obesity, and to prospectively predict its impact on the society around the globe.

5. CONCLUSIONS

This study proposed an agent based model to analyze and to evaluate the obesity trend in different scenarios around the globe. In particular we compared and evaluated two scenarios in order to test and calibrate the simulation model. Our model is designed to reproduce population behaviour in order to evaluate how social and cultural conditions impact on obesity epidemics. It was developed an extension of the IA-CGF intelligent agents developed by Simulation Team to represent the behaviour of the population and the critical factors related to obesity epidemics. The use of these agents and simulation allows to investigate large scale health care problems, and represent an important opportunity to for prediction and early interventions. The obesity epidemic represents a very important and interesting application framework that could be very useful to consolidate research in this area of modeling and simulation related to Medicine and Health Care.

In addition the research highlighted the critical aspects related to collecting, mining and filtering the data to define the conceptual models related to such complex problems as well as to support parameter fine tuning and simulator VV&A (Verification, Validation and Accreditation).

REFERENCE

- Australian Bureau of Statistics, National Health Survey 2004-05: Summary of results. ABS cat.no. 4364.0. Canberra, 2005
- Avalle L, A.G. Bruzzone, F. Copello, A. Guerci, P.Bartoletti Epidemic Diffusion Simulation Relative to Movements of a Population that Acts on the Territory: Bio-Dynamic Comments and Evaluations, Proc. of WMC99, San Francisco, January, 1999
- Bruzzone A.G., Fadda P, Fancello G., Massei M., Bocca E., Tremori A., Tarone F., D'Errico G. (2011) "Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes", SIMULATION October 2011, vol. 87 no. 10, p. 857-874, ISSN: 857-874, DOI: 10.1177/0037549711418688.
- Bruzzone A.G. Tremori A., Massei M., Adding Smart to the Mix, Modeling Simulation & Training: The International Defense Training Journal, 3, 25-27, 2011
- Bruzzone A.G., Tremori A., Madeo F., Tarone F, (2011) "Intelligent agents driving computer generated forces for simulating human behaviour in urban riots", International Journal of Simulation and Process Modelling, 2011 Vol. 6 No. 4 p. 308-316, DOI: 10.1504/IJSPM.2011.048011
- Bruzzone A.G., Massei M., Intelligent Agents for Modelling Country Reconstruction Operation", Proceedings of AfricaMS 2010, Gaborone, Botswana, September 6-8, 2010
- Bruzzone A.G., Scavotti A., Massei M., Tremori A., Metamodelling for Analyzing Scenarios of Urban Crisis and Area Stabilization by Applying Intelligent Agents, Proceedings of EMSS2008, September 17-19, 2008, Campora San Giovanni (CS),Italy, 2008
- Bruzzone A.G., Briano A., Bocca E., Massei M. (2007). Evaluation of the impact of different human factor models on industrial and business processes". SIMULATION MODELING PRACTICE AND THEORY, vol. 15, p. 199-218, ISSN: 1569-190X
- Bruzzone A.G., Bocca E., Rocca A., Algorithms devoted to Reproduce Human Modifiers in Polyfunctional Agents, Proc. of SCSC2006, Calgary, Canada, July 30-August, 2006
- Bruzzone (2004). Preface to modeling and simulation methodologies for logistics and manufacturing optimization . SIMULATION, vol. 80, p. 119-120, ISSN: 0037-5497, doi: 10.1177/0037549704045812
- Cereda C., Models and Analysis of Complex Systems for the Evaluation of Future Scenarios and of

- Related Infrastructural and Operational Needs, Genoa University Thesis, DIPTTEM Press, 2011
- Christakis N.A., Fowler J.H., The Spread of Obesity in a Large Social Network Over 32 Years, *The New England Journal of Medicine*, 357-4, 370-379, 2007
- Cournot M, Marquie JC, Ansiau D, Martinaud C, Fonds H, Ferrieres J, Ruidavets JB: Relation between body mass index and cognitive function in healthy middle-aged men and women. *Neurology* 67:1208-1214, 2006
- Dragone D., Savorelli L. (2010) "Thinness and Obesity: A Model of Food Consumption, Health Concerns, and Social Pressure", LSE STICERD Research Paper No. EOPP017, June 2010. Available at SSRN: <http://ssrn.com/abstract=1717449>
- Falkstedt D, Hemmingsson T, Rasmussen F, Lundberg I: Body Mass Index in late adolescence and its association with coronary heart disease and stroke in middle age among Swedish men. *International Journal of Obesity* 31:777-7783, 2006
- Foster G.D., What is a reasonable weight loss? Patients' Expectations and Evaluation of Obesity Treatment Outcomes, *Journal of Consulting and Clinical Psychology*, 65 (1): 79-85, 1997
- Giribone P, Bruzzone A. (1999). Artificial Neural Networks as Adaptive Support for the Thermal Control of Industrial Buildings. *International Journal of Power & Energy Systems*, vol. 19, No.1, p. 75-78, ISSN: 1078-3466
- Gortmaker S. L., Swinburn B. A., Levy D., Carter R., Mabry P. L., Finegood D. T., Huang T., Marsh T., Moodie M. L. (2011) "Changing the future of obesity: science, policy, and action", *Lancet*, Volume 378, Issue 9793, Pages 838 - 847, 27 August 2011
- Hall KD, Sacks G, Chandramohan D, Chow CC, Wang YC, Gortmaker SL, Swinburn BA. Quantification of the effect of energy imbalance on bodyweight. *Lancet*. 2011 Aug 27;378(9793):826-37. (PMID:21872751)
- Kivipelto M, Ngandu T, Fratiglioni L, Viitanen M, Kareholt I, Winblad B, Helkala EL, Tuomilehto J, Soininen H, Nissinen A: Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease. *Arch Neurol* 62:1556-1560, 2005
- Magarey AM, Daniels LA & Boulton JC, Prevalence of overweight and obesity in Australian children and adolescents: reassessment of 1985 and 1995 data against new standard international definitions. *Medical Journal of Australia* 174: 561-564, 2001
- National Institutes of Health, Clinical Guidelines on the identification, Evaluation and Treatment of overweight and Obesity in Adults: The Evidence Report, 1998
- OECD (2010), *Health at a Glance: Europe 2010*, OECD Publishing. http://dx.doi.org/10.1787/health_glance-2010-en
- Ogden C., Carroll M., Kit B., Flegal K. (2012), "Prevalence of Obesity in the United States, 2009–2010", *NCHS Data Brief*, No. 82, January 2012
- Pliskin, J., Shepard, D, Weinstein, M, Utility Functions for Life Years and Health Status". *Operations Research (Operations Research Society of America)* 28 (1): 206–224, 1980
- Popkin B.M., Kim S., Rusev E.R., Du S., Zizza C., Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases, *Obesity Review*, 7, 271-293, 2006
- Spearman C.E. "The Abilities of Man: Their Nature and Measurement", Macmillian, London, 1927
- Swinburn B. A., Sacks G., Hall K. D., McPherson K., Finegood D. T., Moodie M. L., Gortmaker S. L. (2011) "The global obesity pandemic: shaped by global drivers and local environments", *Lancet*, Volume 378, Issue 9793, Pages 804 - 814, 27 August 2011
- Wang L.Y., Yang Q., Lowry R. and Wechsler H., Economic Analysis of a School-Based Obesity Prevention Program, *Obesity Research* 11, 1313–1324, 2003
- Wang Y, Beydoun MA: The Obesity Epidemic in the United State- Gender, Age Socioeconomic, Racial/Ethnic and Geographic Characteristics : A systematic Review and Meta-Regression Analysis. *Epidemiologic Reviews* 29:6-28, 2007
- Whitmer RA, Gunderson EP, Barrett-Connor E, Quesenberry CP, Jr., Yaffe K: Obesity in middle age and future risk of dementia: a 27 year longitudinal population based study. *BMJ* 330:1360, 2005
- WHO, Preventing and Managing the Global Epidemic of Obesity: Report of the World Health Organization Consultation of Obesity, Geneva, June 1997
- WHO, World Health Statistics: Reports, 2011
- Wolf PA, Beiser A, Elias MF, Au R, Vasan RS, Seshadri S: Relation of obesity to cognitive function: importance of central obesity and synergistic influence of concomitant hypertension. *The Framingham Heart Study. Curr Alzheimer Res* 4:111-116, 2007
- Wang C., McPherson K., Marsh T., Gortmaker S. L., Brown M. (2011) "Health and economic burden of the projected obesity trends in the USA and the UK", *Lancet*, Volume 378, Issue 9793, Pages 815 - 825, 27 August 2011

Authors' Index

Abo Hamad, 149
Adam, 1
Amann, 125
Arisha, 149
Backfrieder, 59
Barbic, 172
Bardon, 10
Batzel, 175
Bauer, 101
Biagini, 10
Bianchi, 172
Breitenecker, 101, 113, 119
Bruzzone, 188, 195
Cairns, 1
Caruso, 156
Cès, 68
Chiêm, 68
Chung, 10
Crowe, 149
Czasny, 180
De Micheli, 21
Dipaola, 172
Einzinger, 87, 93, 119
Endel, 119
Fragomeni, 15
Frascio, 188
Fülöp, 180
Furlan, 172
Fürtinger, 175
Gordon-Wright, 162
Gray, 10
Gremaud, 162
Guida, 42
Gyimesi, 180
Harrison, 10
Hedrich, 26
Iannone, 42
Joy, 10
Kantianis, 10
King, 125
Krishnan, 131
Laforet, 26
Larue, 32
Leskovar, 93
Liu, 131
Longo, 188
Lorenz, 143
Luangmul, 54
Macq, 68
Madeo, 195
Mahdi, 75
Marque, 26
Martens, 162
Massei, 188
Massote, 156
Mathis-Edenhofer, 180
Matsuka, 107
Mazzitelli, 15
Meda, 172
Mehlsen, 81, 107
Miksch, 101, 119
Miranda, 42
Nkambou, 32
Novak, 162, 195
Olsen, 81
Olufsen, 75, 81, 107, 136
Ottesen, 75, 81, 136
Pacetti, 172
Parragh, 87
Percy, 1
Pferzinger, 143
Pichitlamken, 49, 54
Poirier, 32
Pöll, 101
Popper, 113
Renzulli, 15
Riemma, 42
Sarno, 42
Schmitz, 68
Schneditz, 175
Selmi, 172
Serraino, 15
Shibata, 166
Simalatsar, 21
Siri, 188
Smeathers, 1
Sousa, 156
Speybroeck, 68
Teschl G., 125
Teschl S., 125
Tran, 81, 107
Tremori, 188
Uangpairoj, 166
Unterkofler, 125
Van Durme, 68
Vandendorpe, 68
Weerawat, 49, 54
Wilbacher, 113
Williams, 107
Winterer, 101
Wongsammacheep, 49
Wurzer, 143
Wytrzens, 93
Xu, 131
You, 21
Zauner, 119
Zechmeiste, 119r
Zwettler, 59