HADA: TOWARDS A GENERIC TOOL FOR DATA ANALYSIS FOR HOSPITAL SIMULATIONS

I. Castilla[a], M. M. Gunal[b], M. Pidd[c], R. M. Aguilar[d]

[a][d]Department of Systems Engineering and Automation, and Computer Architecture, La Laguna University, Spain
[b][c]Lancaster University Management School, UK.

[a]ivan@isaatc.ull.es, [b]m.gunal@lancaster.ac.uk, [c]m.pidd@lancaster.ac.uk, [d]r.aguilar@ull.es

ABSTRACT

Discrete Event Simulation (DES) in healthcare modeling has been an active research area for many years. However one of the drawbacks of this method is the need for meaningful data for building valid models. This paper discusses Hospital Activity Data Analysis (HADA) which is software specifically designed to be used with a generic hospital simulation model (DGHPSim). The DGHPSim model is built for UK healthcare system and is presented conceptually in this paper. HADA integrates with raw data from different sources to evaluate a hospital's past performance. Its results can be used by hospital managers for statistical inference and general understanding, and by DGHPSim users for estimating appropriate parameters of the simulation models. As well as its use in DGHPSim HADA is well-suited to be generically used for any patient-flow type hospital simulation models.

Keywords: data analysis, discrete event simulation, healthcare

1. INTRODUCTION

It is a sad fact that many people must wait a long time before receiving the healthcare they need. The source of this problem in countries such as UK and Spain, where health care is financed through taxation, is the use of waiting lists in order to try to ration hospital care. The UK National Health System (NHS) has a long history of waiting lists as the service struggles to cope with a huge number of patients using limited resources.

Among other analytical techniques, Discrete Event Simulation (DES) has been widely used in health care analysis and improvement for many years. However, most DES applications tend to be highly focused with a microscopic scope on single services such as emergency departments (Jurishica 2005), outpatient clinics (Harper and Gamlin 2003), and operation theatre capacities (Sciomachen, Tanfani, and Testi 2005).

Although DES is known to be a flexible tool, and hence is used frequently in modelling in healthcare, one of its burdens in applications is the requirement for extensive data and its manipulation (Banks and Carson 1984). Data analysis is an important phase in the development of most simulation models. When dealing with a complex social system such as a hospital, some data may be easily obtainable but others may be very difficult to acquire, making it hard to obtain a clear representation of what the modeller wishes for (Jurishica 2005; Katsaliaki, Brailsford, Browning, and Knight 2005).

Modelling a hospital requires information (and data) from various sources such as a hospital information system, interviews with hospital staff, and personal observations at the hospital. All of these sources help the modeller gain understanding of the important aspects that need to be simulated.

Interviews and visits offer a qualitative view of the real system and interviews are challenging tasks in which the skills of the interviewer and the predisposition of the interviewee are crucial. Visits are useful if a general view of the hospital is required and can also fill some information gaps which can not be explained only with numerical data, such as the disposition of rooms and wards.

For large hospitals only source of verifiable, quantitative data from a hospital is its information systems from which the modeller may generate inputs and other characteristics of the model, though it is important to be wary of data generated by information systems when modelling (Pidd 2002). Moreover, being generally stored in databases, automated data extraction is possible. However, knowledge extraction is not as straightforward as it seems as the data structure used in the hospital information system may differ from those used in the simulation model. These differences can lead to heavy pre-processing and reorganization of hospital data. This problem is aggravated when trying to use the same simulation model for several hospitals: each hospital handles its own data structure and thus getting a generic transformation mechanism from data structure to simulation model is an extremely complex task. Other important sources of complexity include;

- incompatible data types (numeric data stored as text),

- different codifications for the same topic,
- missing data, and
- data errors

Although there are many generic simulation software packages and libraries which can be used to build a hospital simulation model, custom solutions for data analysis seem more appropriate for dealing with data problems. Certainly, a generic software which solves all the above problems is not achievable. However, a reasonable option is to make a specific software in a controlled environment that could be used to homogenize the data sources in order to reduce human interaction as much as possible.

This paper presents the development of a data analysis software which is specifically designed for estimating input parameters of a generic hospital simulation model (District General Hospital Performance Simulation-DGHPSim) [www.hospitalsimulation.info] built for evaluating hospitals' waiting time related performance in the UK. The model is not presented in detail but a discussion is given at conceptual level to provide enough information for a discussion of its data requirements.

2. A GENERIC WHOLE HOSPITAL SIMULATION MODEL

At conceptual level, a typical general hospital can be divided into three main parts: Accident and Emergency Department (A&E) for emergency patients, Outpatient Clinics for elective patients, and Inpatient wards/units for both elective and emergency patients. Patients arrive from the outside world and are, therefore endogenous. As well as entering via A&E, emergency General Practitioner (GP) referrals, are also significant and must be taken into account.

The DGHPSim suite comprises of four discrete event simulation models; A&E, outpatient, waiting list, and inpatient. These models are designed for simulating patient flows to a general hospital from a holistic view to investigate possible ways of reducing waiting times at various stages in patient journeys. The UK government's waiting time targets have put a great pressure on general hospitals in the UK, and hospital managements are forced to use their limited resources more efficiently than previously. The DGHPSim suite is generic and data-driven, that is, it can be fitted to particular hospitals by specifying parameters and other data.. A more comprehensive description of each sub-model can be found in Gunal and Pidd (2007b).

Not surprisingly Gunal and Pidd's main finding in building a generic hospital model is that although huge amounts of data are available with today's information systems in healthcare (hence it may seem like a heaven for simulation modellers), it is difficult to use these data to estimate system parameters which characterises a hospital. Hospital Activity Data Analyser (HADA) shown in Figure 1, is software designed to feed simulation models with the required inputs from the real world system. Note that the real data is not being used directly by the models but instead inferences from the data are used.

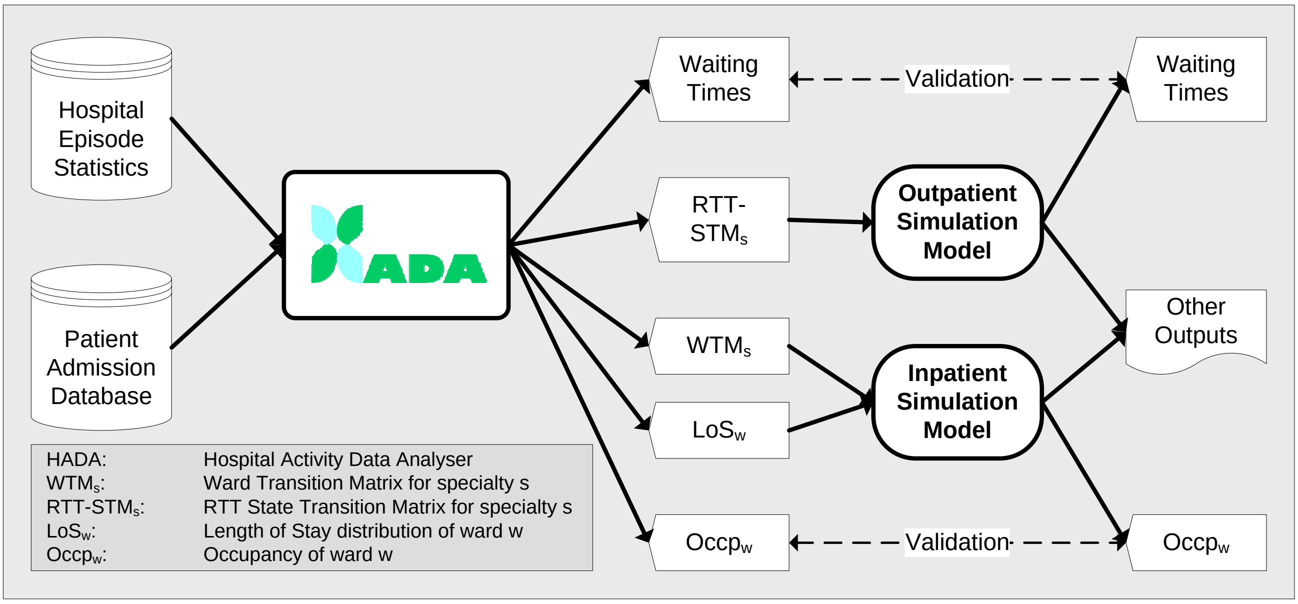


Figure 1: Methodology for Estimating Input Parameters of Simulation Models

3. DATA SOURCES

There are two main sources of hospital data that are integrated in HADA: a hospital's patient admission system (PAS) and hospital episode statistics (HES).

3.1. Hospital Episode Statistics

HES is a UK-wide, routinely collected dataset capturing details of all inpatient and outpatient hospital episodes in the NHS. HES includes two huge datasets: one for outpatients (hereafter HESOP), which includes data regarding all outpatient appointments; and one for inpatients (hereafter HESIP), which gathers decision to admit, admission, operations. HESIP and HESOP structures are defined in HES-Online Data Dictionary (The NHS Information Centre 2008).

This data could be used to draw the general picture of Referral-To-Treatment (RTT) patient journeys of a hospital. However, there is no unique identifier which matches HESIP and HESOP to enable them to be linked to form patients' full journeys. Consequently, there is no way to link with absolute certainty episodes from outpatients and inpatients belonging to the same pathway.

3.2. Patient Admission System

Since HES is routinely collected for all hospitals, it can be seen as a source for generic simulation models. However, more detailed and customized models of a specific hospital would require a direct access to the hospital's information systems.

This data source, as opposed to HES data, is not homogeneous. Therefore, each hospital handles its own data and assumptions. Indeed, two different hospitals can even use the same term referring to different topics. For example, one hospital could treat spell and episode as synonyms.

Errors and incoherencies are also a frequent source of problems. In contrast to HES, which filters, or at least marks, most errors, hospital data can require a more comprehensive and careful review, and a more powerful error handling mechanism.

4. HADA

Hospital Activity Data Analyser (HADA) is standalone PC software which is designed for analyzing PAS and HES data for understanding a hospital's past performance as well as for estimating parameters of a hospital simulation model (such as DGHPSim). There is one software module per data source.

4.1. PAS Analysis

HADA can be used to evaluate hospitals' past performance based on the data provided to the software. The PAS data is pre-processed by HADA, with the help of a data conversion wizard, and processed to display information of two kinds: Bed occupancy, and LoS in each ward (or ward group).

This HADA's module relies on the identification of some basic fields in the original data source. The required fields are shown in Table 1.

Table 1: Fields Required from PAS Data

Field name	Meaning
Patient identifier	Anonymous patient Identification number
Spell number	Spell Identification number
Admission date	Admission date/time
Discharge date	Discharge date/time
Episode start date	Consultant (or bed) episode start date
Episode end date	Consultant episode end date
Elective date	For elective patients, decision to admit date
Ward code	Ward code number (or name)
Patient classification	Ordinary/Day case/Regular category
Admission source	Admission source
Admission method	Type of admission code (e.g. Emergency (21/22/23), Elective (11/12/13))
Primary diagnosis code	Diagnostic code
Consultant specialty	Consultant main specialty
Consultant	Anonymous consultant code

A user needs to identify these fields in the original data source. If the fields are not directly available, he or she should provide the software with a table where this data is accessible. The data conversion wizard allows the user to select a raw data source, to match the original fields with the corresponding expected fields, and, additionally, to add SQL statements which perform extra processing over the raw data. Once the wizard finishes, a preprocessed table is available. HADA makes use of this table to show a set of results.

By default, the resulting information about hospital wards and units is displayed by each ward. For example, if there are 20 wards in the PAS, the output can be displayed for each one of these 20 wards. Alternatively, wards (or units) can be logically grouped. For example if there are 4 general medicine wards in the hospital, they can be grouped as one, to be able to observe the general medicine wards' activity as whole. The grouping is especially useful and necessary for the transitions, which will be explained below.

HADA generates three kinds of outputs: bed occupancy, length of stay and transitions.

- HADA shows the number of occupied beds in two categories (Figure 2); Overnight stays (blue lines) and same-day discharges (red lines). This categorization is necessary mainly because HADA uses only "Admission Date" and "Discharge Date" to calculate occupancy. For example if a patient's admission and discharge dates are the same, this means that the patient occupied one bed during the day however did not stay overnight. This is especially possible for observation wards (or units) such as Medical Assessment Wards, or Clinical Decision Units.
- The second type of output are related to the Length of Stay (LoS). Figure 3 shows the LoS histogram of the General Medicine wards from a UK hospital;. LoS histograms are especially useful to observe the LoS distributions in wards. One should generally expect some sort of decreasing curve (negative exponential), like the one in Figure 3.
- The PAS analysis section also generates the transition counts and probabilities of patients moving between wards whilst in the hospital. This is calculated for each specialty in the hospital, and for each type of patient (Emergency, Elective) separately. For example Table 2 shows the Ward Transition Matrix (WTM) for emergency patients admitted to general medicine. The first column shows "From" wards and the first row "To" wards. The "Gate" symbolizes the entrance and "Disc" the discharge. The table shows how 4711 patients are first admitted to the Assessment ward (Gate-GASM) and, of these, 2452 patients are transferred to a Medical ward (GASM-GMED); finally, 1712 patients are discharged from the Assessment ward (GASM-Disc). Remember that these wards could be actual wards in the hospital or the group of wards as the user defines; in this case, they are ward groups.

Use of WTMs in generic hospital bed management simulation models has been first introduced by Gunal and Pidd (2007a). This method depicts complex relationships between hospital units, based on historical data, and gives simulation model users full flexibility to experiment with different alternatives of bed configurations

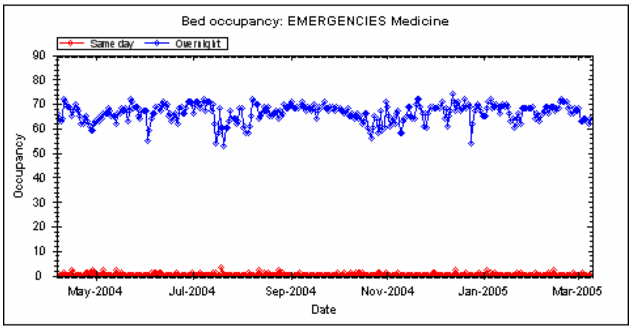


Figure 2: Bed Occupancy Related Outputs.

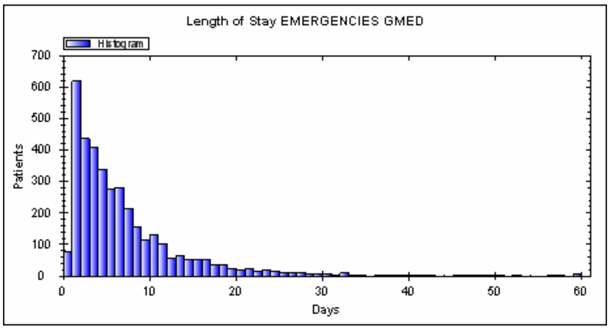


Figure 3: Length of Stay Distribution Related Outputs.

Table 2: General Medicine – Emergency Patients Ward Transition Matrix.

	GASM	GCAN	GCRI	GELD	GMED	GSPE	GSUR	GWOC	Disc
Gate	4711	10	455	9	397	3	15	3	0
GASM	24	125	80	242	2452	12	115	26	1712
GCAN	2	0	2	11	6	0	2	0	157
GCRI	25	3	8	5	203	0	14	0	330
GELD	6	0	4	50	19	0	9	0	472
GMED	18	41	32	224	15	23	159	64	2533
GSPE	0	0	0	2	5	0	2	0	30
GSUR	2	1	7	16	11	1	11	5	276
GWOC	0	0	0	1	1	0	3	0	93

4.2. HES Analysis

The second main function of HADA is related to analysis of the HES data. The final aim is to get a general picture of Referral-To-Treatment (RTT) patient journeys of a hospital (or a trust), that is, the succession of events which describe the different stages a patient goes through. This succession should include the events shown in Table 3.

Table 3: RTT Journey Events.

Event Type	Meaning	Source table
GP	First referral event, generally by a GP	HESOP
OP1	First outpatient appointment event	HESOP
OP2	Pre-operation follow-up outpatient appointment event	HESOP
IPDC	Inpatient as day-case admission event	HESIP
IPOR	Inpatient as ordinary admission event	HESIP
DADC	Decision to admit event (day-case)	HESIP
DAOR	Decision to admit event (ordinary)	HESIP
POP	Post-operation follow-up outpatient appointment event	HESOP
END	Discharged from consultant's care (in OP stage)	HESOP

As stated before, there is no implicit link between HESOP and HESIP tables apart from a patient identifier. It is not the objective of this analysis to detect relations between pathologies in a patient. Thus, it is assumed that different pathologies of the same patient are considered as different journeys. Consequently, we must rely on intuitive or heuristic knowledge in order to identify these journeys.

HES events consists of the following basic fields:

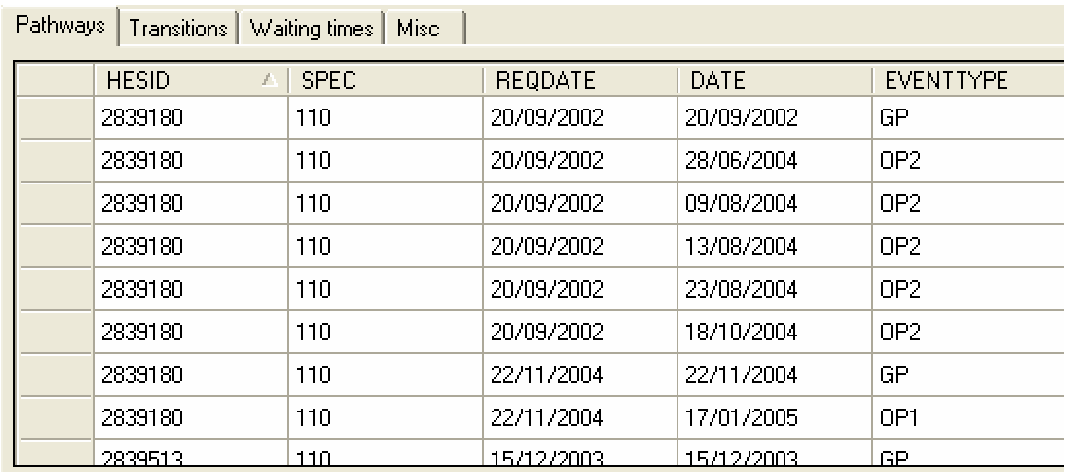
- HESID: Anonymized patient identifier in HES tables.
- Specialty: Specialty of the consultant who is responsible for the patient's treatment.
- Referral Date: Referral date of the referrer (generally General Practitioners).
- Date: The date when this event happened.

HESID, Specialty and Referral Date constitute the patient journey key. Referral date is the trickiest part of this key, because it does not appear in HESIP.

HADA firstly identifies individual events in both tables. HESOP events are ordered by <HESID, Specialty, Referral Date>; and HESIP events are ordered by <HESID, Specialty, Admission Date, Episode End>. On the one hand, HESOP events with the same key constitutes a HESOP pathway, which is part of the whole patient journey; on the other hand, only pairs of (Decision to Admit, Admission) events can be identified as part of the whole patient journey in HESIP.

The next step is to link these pairs in HESIP with the corresponding HESOP pathway. A heuristic algorithm is used on this purpose. The algorithm looks for the first HESOP pathway with HESID and Specialty equals to one HESIP pair, and referral date previous to Decision to admit date. The Referral Date of this pathway is taken as the Referral Date of the whole patient journey, including HESIP events.

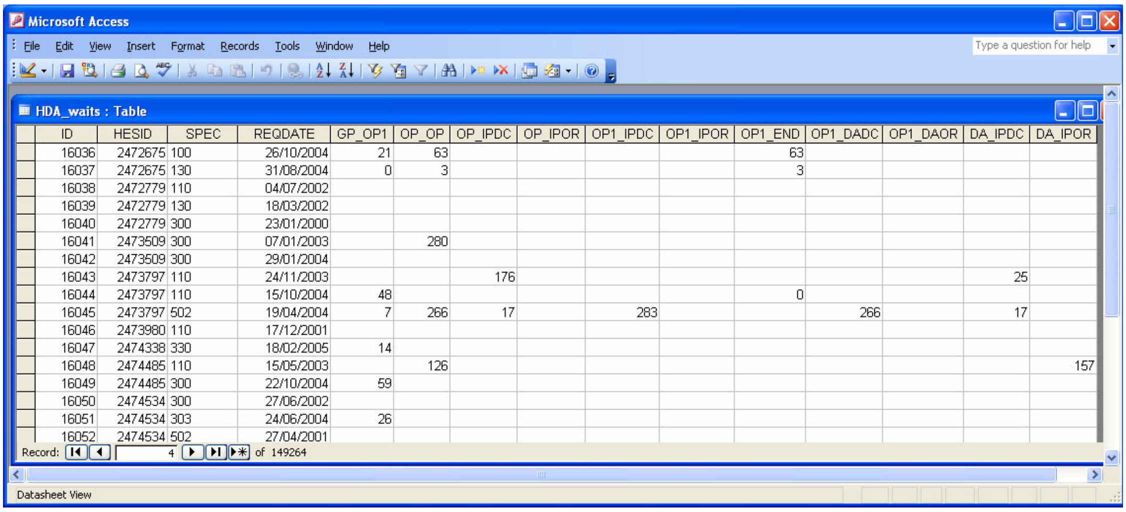
Once used, this HESOP pathway is disregarded and not used again. Thus, some HESIP pairs will not find their source referral date. These loose pairs are neglected. Figure 4 shows how HADA computes some patient pathways.



HESID	SPEC	REQDATE	DATE	EVENTTYPE
2839180	110	20/09/2002	20/09/2002	GP
2839180	110	20/09/2002	28/06/2004	OP2
2839180	110	20/09/2002	09/08/2004	OP2
2839180	110	20/09/2002	13/08/2004	OP2
2839180	110	20/09/2002	23/08/2004	OP2
2839180	110	20/09/2002	18/10/2004	OP2
2839180	110	22/11/2004	22/11/2004	GP
2839180	110	22/11/2004	17/01/2005	OP1
2839513	110	15/12/2003	15/12/2003	GP

Figure 4: Pathway Details Output Screen.

Based on the pathways table, HADA calculates the waiting times, or the delays between events. This is written to an MS Access table (Figure 5). Analyzing waiting times data is not done by HADA and is left to the users to do it externally, e.g. copying GP_OP1 column to Excel to draw a histogram.



ID	HESID	SPEC	REQDATE	GP_OP1	OP_OP	OP_IPDC	OP_IPOR	OP1_IPDC	OP1_IPOR	OP1_END	OP1_DADC	OP1_DAOR	DA_IPDC	DA_IPOR
16036	2472675	100	26/10/2004	21	63					63				
16037	2472675	130	31/08/2004	0	3					3				
16038	2472779	110	04/07/2002											
16039	2472779	130	18/03/2002											
16040	2472779	300	23/01/2000											
16041	2473509	300	07/01/2003		280									
16042	2473509	300	29/01/2004											
16043	2473797	110	24/11/2003			176							25	
16044	2473797	110	15/10/2004	48						0				
16045	2473797	502	19/04/2004	7	266	17		283			266		17	
16046	2473980	110	17/12/2001											
16047	2474338	330	18/02/2005	14										
16048	2474485	110	15/05/2003		126									157
16049	2474485	300	22/10/2004	59										
16050	2474534	300	27/06/2002											
16051	2474534	303	24/06/2004	26										
16052	2474534	502	27/04/2001											

Figure 5: Waiting Times Details Output Table in MS Access.

HADA's final output is the RTT State Transition Matrix. This is produced for every specialty and for all specialties separately. An example is given in Table 4. These matrices can be used for evaluating day-case and ordinary surgery rates, or follow-up and end of treatment percentages.

Table 4: General Surgery RTT State Transition Matrix for Specialty Code 100.

	GP	OP1	OP2	END	IPDC	IPOR	POP
GP	0	1	0	0	0	0	0
OP1	0	0	0.391	0.538	0.034	0.037	0
OP2	0	0	0.457	0.492	0.011	0.040	0
END	0	0	0	0	0	0	0
IPDC	0	0	0	0.601	0	0	0.399
IPOR	0	0	0	0.379	0	0	0.621
POP	0	0	0	0.487	0	0.105	0.408

5. CONCLUSIONS

We have discussed a generic software package for analyzing hospital activity data for a patient-flow hospital simulation model (DGHPSim). This software, HADA, can be used by analysts who wish to investigate a hospital's past performance, thus becoming an aiding software for better decision making, and for estimating input parameters for DGHPSim and other patient flow simulations. At this time, HADA has only been tested with DGHPSim, but future research includes tests with other simulation models.

HADA is intended to significantly reduce the time required by simulation users to set up input parameters. Hence, the use of this software makes patient-flow hospital simulation models more reusable.

ACKNOWLEDGMENTS

The DGHPSim project is funded by the Engineering and Physical Sciences Research Council of UK, under grant EP/C010752/1.

Iván Castilla is being supported by an FPU grant (ref. AP2005-2506) from the Ministerio de Educación y Ciencia of Spain.

REFERENCES

Banks, J. and Carson II, J.S., 1984. *Discrete-event system simulation*. Engelwood Cliffs, New Jersey: Prentice-Hall, inc.

Gunal, M.M. and Pidd, M., 2007a. Moving from Specific to Generic: Generic Modelling in Health Care. *Proceedings of the 2007 INFORMS Simulation*, July 5-7, INSEAD, France

Gunal, M.M. and Pidd, M., 2007b. Interconnected DES Models of Emergency, Outpatient, and Inpatient Departments of a Hospital. *Proceedings of the 2007 Winter Simulation Conference*, pp. 1461-1466. December 9-12, Washington, D.C. (USA).

Harper, P.R. and Gamlin, H.M., 2003. Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. *OR Spectrum*, 25(2): 207-222.

Jurishica, C.J., 2005. Emergency department simulations: medicine for building effective models. *Proceedings of the 2005 Winter Simulation Conference*, pp. 2674-2680. December 4-7, Orlando, FL (USA).

Pidd, M., 2002. *Tools for thinking: modeling in management science*. 2nd ed. Chichester: John Wiley & Sons Ltd.

Katsaliaki, K., Brailsford, S., Browning, D. and Knight, P., 2005. Mapping care pathways for the elderly. *Journal of Health Organization and Management*, 19(1): 57-72.

Sciomachen, A., Tanfani, E. and Testi, A., 2005. Simulation models for optimal schedules of operating theatres. *International Journal of Simulation: Systems, Science and Technology*, 6(12-13): 26-34.

The NHS Information Centre, 2008. *HES Online Data Dictionaries*. Available from: http://www.hesonline.org.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=289 [Accessed 8 July 2008]

BIOGRAPHIES

IVÁN CASTILLA was born in La Laguna, Tenerife and went to the University of La Laguna, where he studied Engineering Computer Science and obtained his degree in 2004. He is currently developing his PhD with the department of Systems Engineering and Automation at the same university. His interests include discrete event simulation and business intelligence. Email to <ivan@isaatc.ull.es>.

MURAT GUNAL is pursuing his PhD at Lancaster University. He received his MSc degree from the same university in 2000. He is an experienced simulation modeller and his current research interest is to investigate the use of simulation on performance measurement in public services such as hospitals. Email to <m.gunal@lancaster.ac.uk>.

MIKE PIDD is Professor of Management Science at Lancaster University where his work spans simulation model-ling and the development of improved simulation methods utilising current developments in computing hardware and software. His current application work focuses on simulation modelling for improvement in public services, especially in policing and healthcare. He is known for three books: Computer simulation in management science (in its 5th edition), Tools for thinking: modelling in management science (in its 2nd edition) and Systems modelling: theory and practice – all published by John Wiley. Email to <m.pidd@lancaster.ac.uk>.

ROSA M. AGUILAR received her MS degree in Computer Science in 1993 from the University of Las Palmas de Gran Canaria and her PhD degree in Computer Science in 1998 from the University of La Laguna. She is associate professor in the Systems Engineering and Automation Department at the University of La Laguna, Canary Islands, Spain. Her current research interests are decision-making based on discrete event simulation systems and knowledge based systems, intelligent agents, and intelligent tutorial systems. Email to <r.aguilar@ull.es>