Document generated on 04/20/2024 12:22 a.m.

Algorithmic Operations Research

Multiobjective Optimization in Health Care Management. A metaheuristic and simulation approach.

Cristina Azcárate, Fermín Mallor and Aurora Gafaro

Volume 3, Number 2, Fall 2008

URI: https://id.erudit.org/iderudit/aor3_2art07

See table of contents

Publisher(s)

Preeminent Academic Facets Inc.

ISSN

1718-3235 (digital)

Explore this journal

Cite this article

Azcárate, C., Mallor, F. & Gafaro, A. (2008). Multiobjective Optimization in Health Care Management. A metaheuristic and simulation approach. *Algorithmic Operations Research*, *3*(2), 186–202.

Article abstract

This paper describes a methodology which combines elements of statistics, probability, mathematical programming, simulation, multiobjective optimization and metaheuristics, to analyze management problems in a health care context. We apply this approach to a staffing problem in a primary care center, taking into account both cost and service quality criteria. We illustrate our approach with a case study.

All rights reserved © Preeminent Academic Facets Inc., 2008

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Algorithmic Operations Research Vol.3 (2008) 186–202

Multiobjective Optimization in Health Care Management. A metaheuristic and simulation approach.

Cristina Azcárate, Fermín Mallor and Aurora Gafaro

Department of Statistics and Operations Research, Public University of Navarre, Spain

Abstract

This paper describes a methodology which combines elements of statistics, probability, mathematical programming, simulation, multiobjective optimization and metaheuristics, to analyze management problems in a health care context. We apply this approach to a staffing problem in a primary care center, taking into account both cost and service quality criteria. We illustrate our approach with a case study.

Key words: Multiobjective optimization, Health-care management

1. Introduction and literature review

In this paper we show how simulation can be used in combination with other statistical and optimization tools to analyze health care management problems, and obtain the "best configuration" when several objectives are simultaneously considered (cost and patient satisfaction measures). We apply this analysis framework to a real case study of a primary health care center (HCC) in the city of Pamplona, Colombia.

The paper is organized as follows. The rest of this section is devoted to a review of past literature on simulation studies and optimization studies in health care and an introduction to multi-criteria decision problems. Section 2 presents the formulation of the problem and the methodology proposed to solve it. Section 3 describes the case study. Finally, some suggestions for future research and concluding remarks are given in section 4.

Operations Research methods are an important and effective tool for handling a wide range of health care problems. The article by Bailey [7], published in 1952, which used queuing theory to analyze waiting times and appointments in hospital outpatient departments, is considered the first OR work to be applied to health care management.

In the last few decades, many journals and books with a combined health-mathematics profile have released studies relating to the management and operation of health care systems and medical decision making. These problems have been tackled mainly with mathematical programming techniques, heuristics, decision models, queuing theory and discrete event simulation (DES).

1.1. *Simulation in heath care*

Much of the research on medical decision making involves the modeling of disease processes in order to evaluate strategies for prevention, treatment and other interventions. Although Markov processes and decision models are often used to evaluate these medical procedures, DES has become very popular [21]. Simulation, for example, is used to evaluate the cost-effectiveness of treatments for osteoporosis in women [50], for prevention of mother-to-child HIV transmission [44], for coronary heart disease prevention strategies [6], in the prevention and treatment of diabetic retinopathy [33], in a helicobacter pylori screening program [20], for the early detection and treatment of colorectal cancer [31], etc.

Nevertheless, DES has been more widely used to tackle problems relating to health care system management and operation, which are typically characterized by an uncertain working environment and limited human and material resources. Since the first paper that used simulation modeling for hospital facilities [25], a wide range of problems, such as patient flow modeling, bed capacity, management of waiting lists, health care center design, emergency facilities, etc., have been addressed by means of this OR technique. A review of the application of DES to issues arising in health care clinics can be found in [37].

One of the most widely studied problems is hospital resource needs and capacity planning. For example, the variability in hospital bed occupancy makes bed availability planning difficult. The need to cover peak

© *2008 Preeminent Academic Facets Inc., Canada. Online version: http://journals.hil.unb.ca/index.php/AOR. All rights reserved.*

demands and avoid congestion, while also achieving a good average occupancy level is a hard problem for health administrators. Simulation is often used to plan effective and efficient bed capacity. In [18] a DES is used to balance bed utilization in an obstetrics hospital with more than 200 beds. [34] presents a simulation model for hospital bed capacity planning and management, with various types of patient flows. Emergency bed requirements are analyzed in [3]. Peak demands are considered in [53]. A DES is proposed in [45] for planning intensive care units, one of the most expensive hospital departments.

Simulation is also used to determine the appropriate level of material and human resources needed in a health center, as well as to analyze how existing resources could be used in a more efficient way. This kind of problem arises both in the design of a new center and in the reorganization of an existing one. Examples can be found in a family practice health center [51], in a physician clinic within a physician network [52], in a laparoscopic surgery [49], or in relation to the number of nurses needed in an intensive care unit [29].

Several works analyze health care systems operation. There is a rather extensive literature on appointment systems and waiting list management. Waiting lists are the main cause of patient dissatisfaction in developed countries and a problem to which health managers are drawing attention. A simulation model is proposed in [24] as a decision-support instrument for the scheduling of patients waiting for elective surgery in a public hospital system. The model can be used as a tactical and operational decision-support system to schedule the flow of elective surgery patients to appropriate hospitals, besides exploring different ways of using existing or additional resources. A simulation model is built in [54] to compare different hospital admission systems, in [13] to analyze the effect of patient classification in scheduling appointments for ambulatory care services, and in [38] to analyze an intensive care admission and discharge process. Related works can be found in [1], [5] and [46].

Special attention is paid to organ transplant waiting lists. Unfortunately, many patients in Europe have died while waiting for a donor organ. In [42], a simulation model is applied to evaluate the cost-efficiency of different allocation policies in a liver transplant center. Nine alternative policies, depending on clinical severity, time spent on waiting list, age, blood group and estimated chance of survival are considered. This topic is also studied in [48].

Simulation has already been used in other health care applications: Red Cross bloodmobiles [12], a hospital lift system [17], the geographical planning of health centers [32], in modeling the public health response to bioterrorism [35], or in robotic courier deliveries [47]. Problems associated with the incorporation of human behavior into simulation models are considered in [11].

1.2. *Multicriteria Optimization in heath care*

The available literature on optimization techniques for planning health care resources is extensive. One of the problems most frequently addressed is the allocation of resources in hospitals with special emphasis on staff planning. Examples of this type of problem, and a list of further references, can be found in [8] and [16]. While these studies, and many others, consider only one objective function, in this paper we are interested in problems simultaneously involving several objectives.

Multicriteria Decision Making deals with problems entailing multiple and conflicting objectives. The conflicting nature of the objective functions makes a global optimal solution unfeasible. Thus, a compromise solution must be found.

In this context, optimality must be replaced by efficiency. A solution is said to be *efficient* (or non-dominated or Pareto optimal) if any objective function value can be improved without jeopardizing at least one of the remaining objective values. The set of all efficient solutions is called the efficient set or the Pareto frontier. The solution process usually requires the participation of a human decision-maker, who must inform about her preferences in relation to the conflicting objectives. In the search for the final solution, only the efficient solutions must be considered.

Since the seventies, several methods have been proposed to handle this problem. They can be classified on the basis of different criteria. If the number of feasible solutions is small, the problem is called a multi-attribute problem and can be solved with different kinds of methods: multi-attribute utility, outranking methods (like ELECTRE or PROMETHEE) and the Analytic Hierarchy Process. When the number of feasible solutions is large or infinite, the problem is said to be a multi-objective optimization problem, which is an extension of the mathematical programming problem. Multiobjective programming methods can be divided into three classes according to the role played by the decision-maker in the solution process: generating methods (or *a posteriori* methods), *a priori* methods and interactive methods. *Generating methods* are methods for generat-

ing the set of efficient solutions. The two most common approaches to characterizing efficient solutions are the weighting method and the ϵ-constraint method. Among all the existing literature, we suggest the reading of [14], for the classical methods, and the more recent [26], for a review of the state of the art.

The models proposed for the analysis of health care management problems are frequently used to evaluate different performance measures, which can be subdivided into economic performance and service quality measures. Economic performance measures mainly consider the total cost or profit of the health center configuration and the resource usage level. The research has defined several service quality indicators, such as average waiting-in-line times and throughput times, the percentage of patients turned away, patient throughput, over-utilized time, treatment and prevention effectiveness, etc. A detailed discussion of health care performance measurement can be found in [40].

The conflicting nature of these performance measures makes it difficult to determine the optimal configuration of the system. For instance, there is a trade-off between minimizing the cost of the system configuration and maximizing patient satisfaction.

Many studies address these multiobjective situations by using the model to evaluate only a small number of scenarios, showing the trade-off between the performance measures considered, for an example, see [54]. Others authors (as in [51]) narrow their focus to a single budgetary objective. They estimate patient and medical staff satisfaction in monetary terms, by considering economic sanctions for failing to meet waiting time and other service quality objectives. Studies dealing with the evaluation of prevention, treatment and other strategies often handle this problem by defining ratios to compare the cost-effectiveness of different strategies [43]. We have also found applications of multi-criteria decision making techniques incorporating the use of mathematical programming or other optimization techniques (metaheuristic) (for example, see [30], [9] or [41])**,** but have found only one in a simulation context [10].

1.3. *Optimization with simulation in heath care*

Nevertheless, only a few works use an integrated simulation and optimization approach. In [22] a DES model is combined with nonlinear programming and neuronal networks to determine the optimal configuration of a transfusion center. The method determines the numbers of reception staff, nurses or doctors for the hemoglobin test, doctors for the medical examination and venipuncture beds (the decision variables). The authors consider two objectives: minimization of the configuration cost and minimization of the average total time spent in the system. Their proposed model integrates simulation, neural networks and nonlinear integer optimization as follows. The simulation model is used to obtain the sample for the estimation of a functional relation ($y = f(x)$) between the average total time spent in the system, y, and the decision variables, x. This function, estimated by means of the neural network, is used in the formulation of two optimization problems, one for each objective considered, to obtain the optimal system configuration. The simulation model is also used to validate the solutions proposed by the optimization problems. Our research differs from [22] in the way optimization and simulation is combined, as will be explained in section 2. Recently, Brailsford et al. published a study, [10], for the optimal choice of screening policy for diabetic retinopathy, by embedding discrete event simulation in an ant colony optimization model. They consider two objectives (cost and effectiveness) although only one function, the cost-effectiveness ratio, is used to compare solutions. In future research they plan to improve their model by considering multi-criteria objective functions for the ant colony algorithm.

In our case we consider several objective functions, using the ϵ-constraint method to estimate the Pareto frontier in combination with a scatter search method to find solutions and a simulation model to evaluate them and check their feasibility.

2. Problem formulation and methodology

2.1. *Problem statement*

We consider a general primary care center, in which patients arrive without a scheduled appointment. Patients first have to visit an administrative office to obtain an appointment with medical staff (pediatricians, doctors, nurses) if there is remaining medical capacity, they can then visit the medical rooms.

This problem can be modeled as a network queue, with several service facilities. Patients can take different paths through this network. Each service is performed by one of a range of service providers (administrative staff, medical staff, beds…).

From a queuing theory point of view, the study tackles the problem of optimizing the system configuration, that is, determining the number of service providers in each

facility and their time-schedule, in order to optimize certain criteria.

This decision generally identifies two types of criteria, economic and service quality-related.

The problem has random components: the patient arrival process, medical condition, time spent in the doctor's room, and so on. This random environment makes it difficult to define clear objective functions and/or constraints (for example, service quality criteria as the average total time spent in the system, or waiting in line, the percentage of patients turned away, resource usage, etc.).

2.2. *Simulation model*

The uncertainties and complexities of this healthcare system led us to choose simulation as the basic analysis instrument. In constructing the simulation model, the following steps were considered.

1.- Structural modeling

By structural modeling we mean identifying all the important elements in the system, describing them in mathematical terms, and establishing their logical relationships. The nature of the problem makes it straightforward to model the health system as a queuing network with several service facilities. Thus, for each service facility, the usual elements in a queuing model must be considered: the number of service providers, the waiting room, queue discipline, service times, etc.

2.- Data modeling

The input data needed to run the model are the arrival pattern, branching probabilities and service durations.

2.1- Arrival pattern

The patient arrivals are usually seen as a discrete event process that can be described by using appropriate stochastic point processes. A reasonable choice is the Poisson Point Process, due to its characterization which is as follows:

A stochastic patient arrival process $\{N(t), t \geq 0\}$ is a *Poisson Process* if:

(1) Patients arrive one at a time.
(2) The number of arrivals in the time interval $(t,\ t + s]$, $N(t + s) - N(t)$, is independent of the number and times of arrivals taking place from 0 until time t. That is, it is independent of the variable set $\{N(u),\ 0 \leq\ u \leq\ t\}$.
(3) The distribution of $N(t + s) - N(t)$ is independent of t for all t, $s \geq\ 0$.

Properties 1 and 2 can be interpreted as follows. Patients arrive at the hospital on an individual basis, knowing nothing about the patients that have arrived before them (or whatever they know has no influence in their decision about when to go to the health care center) and without anyone coordinating the arrivals of patients according to a pre-established plan. Condition 3 sets the *homogeneity* of the process through time. This condition is more difficult to assume in this kind of arrival process, because arrivals usually peak several times throughout the day. If this third condition is removed from the definition, we get a *non-homogeneous* Poisson Process.

We introduce a new function $\Lambda(t)$ defined as the expected number of arrivals until time t, that is, $\Lambda(t) = E[N(t)]\ (t \geq 0)$. When $\Lambda(t)$ can be derived, its derivative is called the arrival ratio function $\lambda(t) = \Lambda'(t)$ which can be interpreted as the instantaneous expected number of arrivals per unit time at time t. In a non-homogeneous Poisson Process this instantaneous expected mean varies through time.

Either of the two functions, $\lambda(t)$ or $\Lambda(t)$, completely determines de Poisson Process. We propose a procedure for the estimation of these functions from the patient arrival data in appendix A. The method involves some straightforward spreadsheet calculations, an ordinary statistical regression analysis and the derivative of the fitted function. Its main advantage is that it provides a smooth function for the arrival rate function instead of a step function as the usual modeling method does. For more on the importance of correct patient arrival modeling and the use of non-homogeneous Poisson processes, see Alexopoulos et al. [2].

2.2 - Branching probability estimates

After visiting one service facility, a patient can take different paths. Branching probabilities can be estimated from data and/or from verbal reports provided by the staff.

2.3 - Service time estimates

Service times in each of the service facilities can also be estimated from data and/or from verbal consultation with staff.

Doctors' service time usually depends on the patient's sex, age and medical condition. Each combination of age, sex and illness defines a different group of patients.

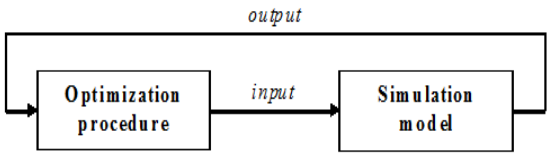

Fig. 1. Combination of Simulation with optimization techniques

Service times for each group of patients and for each type of medical service must be estimated, again from real data and/or from expert opinion.

2.3. *Combining optimization and simulation.*

The simulation model only allows us to test the performance of a given configuration for the HCC through the statistical analysis of a set of output performance measures. Therefore, to determine the best configuration for the system, an optimization procedure must be introduced and linked with the simulation model (see [4]) in the following way.

The optimization procedure determines a system configuration or solution (that is, a value for the decision variables of the problem). This system configuration is simulated. The output of this simulation is used in the optimization procedure to evaluate the random objectives and/or constraints. The optimization procedure, with this information and its search method, decides the next solution to be evaluated (Figure 1). This process goes on until stopping conditions of the optimization method are met.

Metaheuristic approaches are frequently used as an optimization engine, when combining simulation with optimization techniques. For a discussion of metaheuristics, see [28] . At this point, many different metaheuristics can be chosen. As an example, in what follows we present the approach used to analyze our case study: the scatter search, as proposed by Laguna and Martí [39] . The main steps are outlined in Figure 2.

A scatter search was implemented to solve the following type of optimization problem:

$$\text{Min } F(x) \quad \text{subject to} \begin{cases} Ax \le b \\ g_l \le G(x) \le g_u \quad (2) \\ l \le x \le u \quad (3) \end{cases} \tag{1}$$

Where x is the decision variables vector and $F(x)$ may be any mapping from a set of values x to a real value.

1. Apply a diversification generation method to generate a starting set of solutions. Designate a subset of the best points (judged by quality and diversity) to be used as reference points.
   While (*stopping criteria are not met*)
   {
2. Form combinations of subsets of the current reference points to create new points. The combinations are (a) chosen to produce points both inside and outside the convex region spanned by the reference points, and (b) modified by generalized mapping processes to yield feasible points according to the constraints in the problem (both linear and integrality constraints).
3. Update the reference set by selecting points that can improve the quality and/or diversity of the set.
   If (*no new combinations can be explored in the current set*)
   {
   Extract a collection of the best points in the current reference set to be used as starting points for a new application of the diversification generation method.
   }
   }

Figure 2. *The scatter search optimization method (Laguna and Martí, 2002)*

The feasibility of a solution x depends on a set of linear constraints (set 1 in formula (1)) and bounds for the variables (set 3 in formula (1)). Since both types of conditions are known *a priori*, the feasibility of a solution, according to these conditions, can be tested before sending it to the simulator for evaluation. The feasibility of a solution also depends on a set of requirements represented by functions $G(x)$ which are not necessarily linear (set 2 in formula (1)) and can only be checked after the simulation process, because they are expressed in terms of a set of performance measures estimated by means of the statistical analysis of simulation output. The coefficient matrix A, the vector b, and the bounds l, u, g_l and g_u must be known.

The scatter search method implemented in this research begins by generating a starting set of diverse points. Once this initial reference set has been created, new solutions are generated by linear combinations of reference solutions. Before sending a solution to the simulator to obtain its performance measures, a feasibility test is carried out by checking constraints and bounds. If a solution x is not feasible then the following linear programming is solved to find the closest feasible solution x^* to x.

$$\text{Min} \quad d^- + d^+ \quad \text{subject to} \begin{cases} Ax^* \le b \\ x - x^* - d^- + d^+ = 0 \\ l \le x^* \le u \\ d^-,\ d^+ \ge 0 \end{cases} \tag{2}$$

In (2) , d^- is the negative deviation and d^+ is the positive deviation from the feasible solution x^* to the infeasible reference solution x. The mapped solution is sent to

the simulator to obtain a set of performance measures. One of these measures is used as the objective function value *F(x)*, which provides the way to distinguish good solutions from bad ones. Each requirement $G(x)$is also evaluated and compared with its bounds to check the feasibility of the solution. Infeasible solutions are not discarded but handled with a composite function $P(x)$ that penalizes violations of the requirement. The penalty is proportional to the degree of violation and does not remain static throughout the search.

If the reference set does not change because the new solutions lack sufficient quality to be included in it then a diversification step is required, in which the reference set is rebuilt to create a balance between solution quality and diversity. More details about this and other steps in search procedure can be found in [39].

2.4. *Multiobjective optimization*

In our problem we consider both economic and service quality criteria. Most applications in economic or industrial contexts estimate quality criteria in economic terms. Nevertheless, in the health care context, monetary measures of quality aspects can be difficult to obtain and may prove inaccurate. As a result, both kinds of criteria must be considered, and because of their conflictive nature, the problem must be solved by means of a multi-criteria decision model. Thus, the purpose of the analysis is to determine the Pareto Frontier associated with the multiobjective problem. Our proposal is to use multiobjective mathematical techniques to generate the set of efficient solutions. From all the possible generating methods, we have chosen the ϵ-constraint method.

The *ϵ-constraint method* optimizes one of the k objective functions and incorporates the other objectives by means of bound constraints:

$$\begin{array}{ll} \text{Min} & F_i(x) \\ \text{subject } to & \begin{cases} F_j(x) \leq \varepsilon_j \forall j \neq i, j = 1, ..., k \\ x \in S \end{cases} \end{array} \qquad (3)$$

It can be proved (see [14]) that x^* is an efficient solution if, and only if, x^* solves the above problem for every i = 1,..., k. Besides, if x^* is a unique solution of the ϵ-constraint problem, x^* is an efficient solution.

Each of the problems generated by the *ϵ-constraint method* is solved by the scatter search optimization procedure which calls on the simulator to evaluate the explored solutions. Each iteration provides an efficient solution. The methodology is outlined in figure 3.

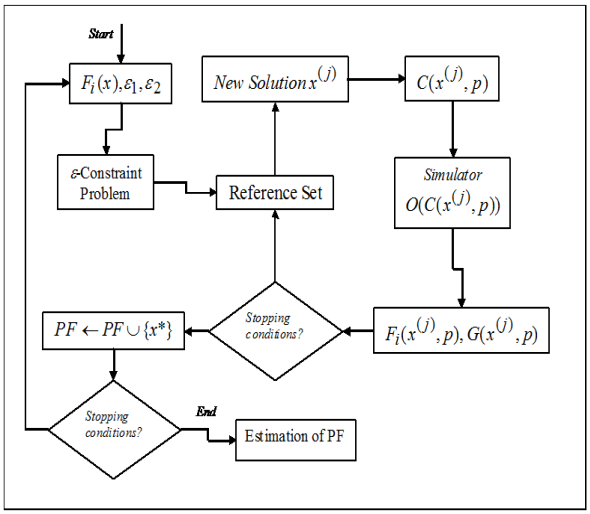

Fig. 3. Methodology: combining ϵ-constraints with scatter search and simulation.

The procedure begins by defining an optimization problem according to the *ϵ-constraint method* that is, by choosing one of the objective functions, $F_i(x)$, to be optimized and two aspiration levels, ϵ_1 and ϵ_2, for the two remaining objective functions. This optimization problem is sent to the scatter search procedure, which builds an initial set of solutions called the *reference set*. Each of these solutions, $x^{(j)}$, together with a set of parameters p (arrival rates, costs, service times, etc.), defines a configuration, $C(x^{(j)}, p)$, of the HCC. This configuration is simulated to obtain a set of performance measurements from the statistical analysis of the simulation output, *O(C(x$^{(j)}$,p))*. In particular we estimate the functions *F_i(x$^{(j)}$,p)* and *G(x$^{(j)}$,p)* which allow us to check the feasibility and quality of the solution $x^{(j)}$. If the stopping conditions of the scatter search are not met, then a new solution is taken from the *reference set* to be evaluated and sent to the simulator. If stopping conditions are met, then the best solution in the reference set, *x**, is considered an optimal solution of the ϵ-constraint problem and, also, therefore an efficient solution to be included in the Pareto Frontier *PF*. While each of the three objective functions are not being optimized for an appropriate set of aspiration levels for the other two objective functions, a new optimization problem is defined (changing values ϵ_1, ϵ_2 and/or function F_i to be optimized) and the preceding iteration repeated.

However, if the decision-maker has a clear idea of how to compare them, then an *a priori* method would be more appropriate to handle the multiobjective problem.

The most representative *a priori* method is Goal Programming (GP), proposed by Charnes and Cooper [15] . The decision-maker must specify his/her aspiration levels, a_i, for each objective function F_i, *i=1,...,k*. By introducing positive and negative deviation variables (d^+,d^-), a goal is built as the relation between the objective function and the aspiration level proposed by the decision-maker. In GP models, the deviations are minimized. A general structure for a GP problem is:

$$\text{Min} \quad g(d^+, d^-)$$
$$\text{subject } to \begin{cases} F_i(x) + d_i^- - d_i^+ = a_i \quad \text{i} = 1,...,\text{k} \\ d_i^-, d_i^+ \geq 0 \quad \text{i} = 1,...,\text{k} \\ x \in S \end{cases} \tag{4}$$

There are different GP approaches. In the weighted approach the decision-maker must also specify positive weights and the sum of the weighted deviation variables is minimized. That is, $g(\mathrm{d}^+, d^-) = \sum_{i=1}^{k} \left(w_i^- d_i^- + w_i^+ d_i^+\right)$

3. A case study

3.1. *Problem description*

We applied the described methodology to the primary health center of the Hospital San Juan de Dios in the city of Pamplona (Colombia), which provides medical services to an economically and culturally disadvantaged population with a high level of poverty and illiteracy. These conditions are the main reason for the special operation characteristics of the center that might appear strange to anyone living in a more developed region.

Patients arrive at the center without a scheduled appointment, from Monday to Friday. From 5 a.m. onwards they start to arrive at the invoicing office, where two people attend patients from 7 a.m. to 9 a.m. Patients are not allowed to arrange appointments by telephone. People without the correct pay-documents are rejected and have to return another day. The receptionist assigns each patient to one of the two doctors until the daily medical capacity is fully allocated, at which point, the patients waiting in the queue are turned back and have to try another day, reinitiating the process at the invoicing office. In exceptional cases, certain special patients may be given an appointment even if capacity is fully allocated. Some patients have their own preferences regarding choice of doctor and this is taken into account when assigning appointments.

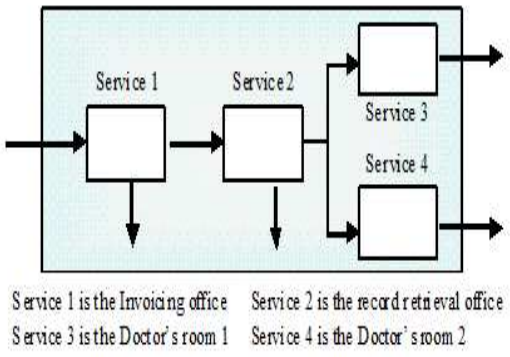

Fig. 4. Network model for the primary health care center.

Next, accepted patients must visit the record retrieval office (open from 7 a.m. to 11 a.m.), where two administrative workers have to retrieve their case history, or, if necessary, open a new patient case history, which adds to service time. Any patients whose documents are not in order are turned away from the record retrieval office and thus exit the system. After this second stage, patients must wait their turn in a waiting room. Doctors see patients from 7 a.m. to 1 p.m.

The purpose of our study was to improve the performance of this care center, taking both economic and patient satisfaction measures simultaneously into consideration. To reach this goal, we developed a model to be analyzed with a combination of simulation and multi-criteria optimization techniques.

3.2. *Simulation model*

3.2.1. *Structural modeling*

From the description of the problem presented in the previous section, it is straightforward that this health system can be modeled as a network queue with 4 service facilities (figure 4). For each service facility we have considered the usual elements in a queuing model: number of service providers, the waiting room, queue discipline, service times, etc.

Input modeling

To check for time-dependence in the arrival pattern, daily patient arrival time data for the period January to September 2004 were analyzed.

Although no monthly pattern emerged, time-dependence in day of the week and hour of the day were detected. A non-*homogeneous* Poisson Process was

considered for each day of the week. The Friday arrival rate estimate is presented in appendix 1 as an example.

Branching probabilities estimates

The first branching in patient flow occurs in the invoicing office, where some patients continue in the care system, while others exit because they are unable to present the required documents or daily medical capacity is already fully allocated. Daily capacity can only be exceeded to attend priority patients. The branching probabilities of a patient without the required documents and the percentage of special patients were derived from the data. These estimations have been validated by invoicing office staff.

The small percentage of patients that exit the system at the record retrieval point was also estimated from data.

Service time estimates

Admission times

The percentages of patients with and without the required documents and the percentage of special patients were estimated from data supplied by the invoicing office staff. The different service times for these three types of patients in the invoicing office were also estimated from the same source.

Record retrieval times

Service times for the three types of patients in record retrieval (patients with a case history, patients without a case history and patients without the required documents) were also estimated from staff-supplied data.

Doctors' service times

As already stated, this service time depends on the patient's sex, age and medical condition. The hospital considers 6 age levels and 21 categories of illness. Different patient-groups are defined by combining age, sex and medical condition data. The fact that not all combinations are possible means that there are only 69 different groups of patients.

Two different data sources were used to estimate service times. On the one hand, day-to-day administrative records are kept of the sex, age and medical condition of each patient seen by each doctor. Thus, we know how many patients from each group are seen each day by the doctors. On the other hand, the service times for each of the patients that entered a doctor's consultation room were also recorded for a period of nine months. Unfortunately, we only know the sequence of times spent by the successive patients inside the doctor's room, but nothing of their personal characteristics such as sex, age or medical condition. So, there was no link between these two data sources, and we were therefore unable to distribute service times across the different patient groups.

A third source of service time data was the doctors themselves, who reported on the minimum, maximum and most usual time needed to see a patient in each of the sex/age/medical condition groups. Furthermore, they agreed to use triangular distributions for the service time in each patient group.

Consequently, we were able to obtain data for global time spent in doctors' rooms and the triangular distributions estimated from the doctors' reports. Doctors' estimates were validated by means of the following statistical analysis.

Let *TT* be the total doctor service time needed to attend all the patients for whom we have recorded data (9,732 patients over the nine months).

We know, from the administrative records, how many patients there are in each group. Then, we have $TT = \sum_{j=1}^{69} \sum_{i=1}^{n_i} T_{ij}$, where $T_{ij}$is the time taken to serve patient i of group j, and n_i is the number of patients in group i. Observe that $T_{ij} = T_j, \forall i$ have a triangular distribution with known parameters (those estimated by the doctors).

By using Liendeberg's central limit theorem, the random variable *TT* can be approximated to a normal distribution, with the following mean and variance:

$$\left.\begin{array}{l} E[TT] = \sum_{j=1}^{69} n_j E[T_j] = 1,858.88 \\ V[TT] = \sum_{j=1}^{69} n_j V[T_j] = (27.77)^2 \end{array}\right\} \quad (5)$$

According to this normal distribution there is a probability of less than 0.001 of taking a value less than or equal to 1,769.93, the sum of the 9,732 service times recorded in hours. We can therefore conclude that, in general, doctors overestimate service time.

Because of the discrepancy between these two information sources, we considered modifying the doctors' probability distributions. To do this, we propose a maximum likelihood classification method, which uses the aforementioned triangular distribution estimates and linear programming, as shown in appendix 2.

As a result of this classification method, we assign a patient group to each service time recorded. Thus, we have samples of service times for each group. With

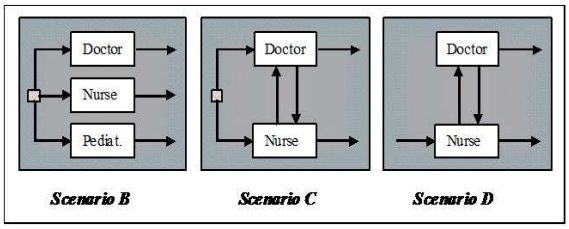

Fig. 5. Network model for the alternative scenarios.

these samples, new service time distribution estimates can be obtained. Note that these new distributions take into account the data supplied the doctors and more closely match the observed times.

3.3. *The optimization problem*

In line with the request of the hospital managers, we considered four different operation scenarios for this health care center and studied the staffing problem for each one of them.

The four scenarios are as follows: scenario A (see figure 4) represents the current configuration of the system, composed of 2, 2, and 2 service providers (invoicing office, record retrieval, doctor rooms, respectively). In scenario B, C, D different operation configurations are proposed (figure 5).

More specifically, scenario B considers three kinds of patient-health needs and the possibility that one or more nurses may attend adult patients with minor ailments, one or more doctors may attend adult patients with severe medical conditions and one or more pediatricians may attend children (up to fourteen years of age).

Scenario C considers the possibility of separate rooms for doctors (D) and nurses (N). Depending on their medical condition, patients may need to visit a doctor only, a nurse only or both, each case resulting in a different flow: $D-N$, $N-D$, $N-D-N$ or $D-N-D$.

Scenario D differs from the previous scenario in that all patients must visit a nurse; this determines patient flow in each case, thus, the allowed flows are N, $N-D$, $N-D-N$.

For each of the above scenarios, we consider the three following **objective functions**:

(1) $F_1(x)$: Minimize cost.
(2) $F_2(x)$: Minimize the percentage of patients turned away.
(3) $F_3(x)$: Maximize a quality factor of doctor service.

The first objective function is an economic criterion that takes into account all economic costs associated with the HCC, the bulk of which is staff salaries; the second objective is to properly dimension the HCC by minimizing the percentage of patients turned away due to the limited daily capacity of doctors/nurses/pediatricians; and the third captures the quality of doctor service to patients by means of a quality measure that depends on the time spent by doctors/nurses/pediatricians on each group of patients. For this purpose we define the variable *"quality factor"*, δ, as a factor to multiply the service times for doctors. So, the simulated doctor service times are the corresponding estimated distributions multiplied by the quality factor. Then, δ*=1* leaves the service times unchanged; δ*=1.20* increases the service times by 20%, and δ*=0.9* decreases the service times to 90%.

The following **decision variables** are considered in the optimization problem

- δ : the quality factor of doctor service which is a continuous bounded variable.
- XI_i, XR_i, *i=1,...,5*: the number of service providers in the invoicing office and in record retrieval, respectively, during the working day i. They are 10 integer variables.
- XD_i, XN_i, XP_i, *i=1,...,5*: the number of doctors, nurses and pediatricians, respectively, during the working day i. They total 10 or 15 integer variables, depending on the scenario.
- TD_i, TN_i, TP_i, *i=1,...,5*: the number of overtime hours for each medical staff type, that is, doctors, nurses and pediatricians, respectively, during the working day i. They total 10 or 15 variables, depending on the scenario.

Each combination of values for these variables constitutes a solution for the decision problem. Given a solution, we need to determine the daily capacity of medical staff, that is, how many patients per day can be appointed to medical staff.

We calculate this number of patients as the maximum value, verifying that the sum of their time spent in the doctor's room is less than the total available doctor service time (*WT*) with a probability equal to or greater than 0.95.

Let C_i denote the number of appointed patients during working day i and T_j the random variable that describes the service time spent on patient j. Then

$$C_i = \max\left\{s/P\left(\sum\nolimits_{j=1}^{s} T_j \leq WT_i\right) \geq 0.95\right\} \quad (6)$$

Denoting by *E[T]* and *V[T]* the mean and variance of a service time T for a generic patient, straightforward

calculations (detailed in appendix C) give us:

$$C_i = \lfloor \left[2\delta\, E[T]\, WT_i + (1.65)^2 \delta^2\, V[T] - \sqrt{(1.65)^4 \delta^4\, V[T]^2 + 4(1.65)^2 \delta^3 E[T]\, V[T]\, WT_i} \right] / (2\delta^2 E[T]^2) \rfloor \quad (7)$$

3.4. *Embedding the simulation in the optimization problem*

From the analyst's point of view, the decision problem is solved by determining the Pareto Frontier for the multiobjective optimization problem associated with each scenario.

The Pareto optimal frontier for each scenario is estimated following the methodology presented in section 2. For ease of notation, suppose that the three objective functions *(j=1,2,3)* represent quantities to be maximized. First, we formulate the set of optimization problems required by the ϵ-constraint method.

$$\text{Maximize } F_i(x) \quad \text{subject to } \begin{cases} F_j(x) \geq \varepsilon_{j,k} & j \neq i;\ k = 1, \ldots, r \\ x \in S \end{cases} \quad (8)$$

S denotes the prior feasibility set for the vector x, that is, in our case, the set of bounds on the variables. For example, we could consider overtime to be bounded at two hours per day. It would also be possible to include linear restrictions on the variables, by requiring, for example, that the number of nurses in scenario C and D be greater than or equal to the number of doctors: $XD_i - XN_i \leq 0$.

The bounds for the non-optimized objectives are also considered as constraints in the case of functions F_1 and F_3 and as a requirement in the case of function F_2. Simulation is needed for the evaluation of the number of patients turned away.

Varying the bounds $\epsilon_{j,k}$ within a range of values decided in conjunction with the managers, we obtain a set of problems, each of which will provide an efficient solution. By exchanging the objective function to be optimized we obtain three such sets of problems.

Each of these optimization problems is solved using the scatter search metaheuristic.

To evaluate a solution the simulator is needed. Note that other performance measures (expected patient time in system, expected time in queue, expected number of patients in queue, resource usage, etc.) are also estimated through simulation. Although these performance measures are not considered as objective functions in our optimization problem, they are evaluated and used as additional, complementary information to present to decision-makers.

Several recently launched commercial discrete-event simulation software packages incorporate an optimization module allowing some kind of optimization to be performed, usually, with the help of metaheuristic techniques. In our research, we have used ARENA simulation software to build our simulation model, which contains the OptQuest optimization module in which the scatter search optimization method is implemented. Comprehensive examples of the use of OptQuest can be found in [4] and a detailed explanation of the optimization procedure implemented in this research can be read in [39].

3.5. *Simulation experimental design and output analysis*

We consider an endpoint simulation model, starting from an empty and idle system state. The simulation run length for each replication is one week of system operation. We use the sequential sampling method to determine the number of replications needed to obtain the accuracy considered in the estimations. That is, our simulation model incorporates a program to check the accuracy of some quality performance estimates after each simulation replication. In more specific terms, at the end of each replication, it evaluates the confidence-interval half width of the performance measures considered. If these half widths are sufficiently accurate, the simulation stops; otherwise, one more replication is run.

We use a variance reduction technique to reduce output randomness. That is, we use the common random numbers technique, to synchronize the use of random numbers across the model alternatives.

As an illustrative example, figure 6 shows the estimated optimal Pareto frontier for scenario C. This figure plots the optimization results for the three objective functions: cost (Z axis), percentage of patients turned away, (Y axis), and service quality factor (X axis). The numerical values of the estimated Pareto optimal frontier plotted in figure 6 are shown in the table below.

Note that this Pareto frontier is an estimation of the true Pareto frontier because one of the three objective functions is evaluated by simulation and because an

Table 1

Solution	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Cost	626	742	756	800	873	1206	656	735	687	660	693	623	665	626
% Turned away	12.27	4.94	2.69	1.68	1.52	0.00	8.95	4.94	7.18	8.36	3.38	9.34	7.18	12.3
Quality	1.50	1.50	1.50	1.50	1.50	1.5	1.43	1.34	1.32	1.30	1.30	1.28	1.27	1.25
Solution	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Cost	1136	642	636	1118	985	947	664	716	616	616	913	695	648	757
% Turned away	0.00	7.18	8.99	0.00	0.00	0.01	2.09	1.01	7.61	7.18	0.00	1.01	1.01	0.00
Quality	1.24	1.24	1.17	1.15	1.12	1.12	1.07	1.05	1.02	0.96	0.86	0.78	0.75	0.75

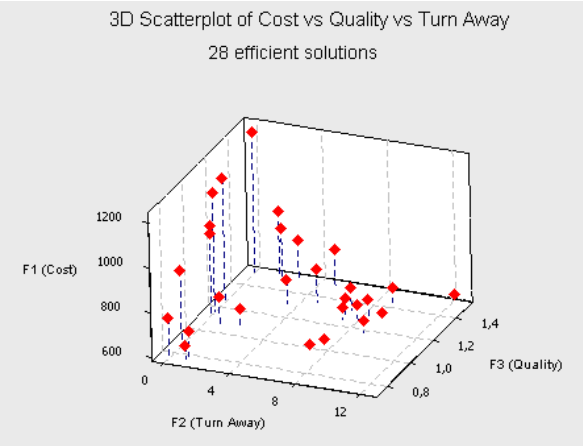

Fig. 6. Example of efficient solutions obtained for the scenario C.

approximating optimization method is used; that is, with no guarantee of obtaining the optimal solution.

From a mathematical point of view, a multiobjective problem is solved by finding the Pareto frontier. From experimental results, we have observed that the usual aspects of the quality of the estimated Pareto frontier, such as proximity, diversity and pertinence are fulfilled:

- Proximity: the estimated Pareto frontier is close to the real one.
- Diversity: a good distribution of solutions in the three objective functions considered.
- Pertinence: the solutions set should contain configurations within the decision-maker's area of interest.

Observe that diversity and pertinence are guaranteed by the use of the ϵ-constraint method; proximity is due to the quality of the metaheuristic used in the resolutions of the ϵ-constraint problems.

Once this efficient set has been generated, it is presented to decision-maker, who must choose the preferred solution. This could be complicated in the event of a Pareto Frontier with many points. The decision-maker should first identify the area of interest to compare solutions by making a kind of trade-off between pairs of objectives. An example of the type of reasoning that might be used by the decision-maker is as follows: Suppose that we want to keep the budget controlled and then relatively close to the minimum (616) in the sense that increments of less than 20% (around 720) are desirable. We also want to keep down the number of patients turned away, say below 2%. These two conditions would define the pertinence area, which is given by solutions 22, 26 and 27. These solutions provide the same percentage of patients turned away and have costs of 716, 695 and 648, respectively, and quality measures of 1.05, 0.78 and 0.75, respectively. Which is the best choice? The decision-maker will probably discard solution 26 first, because the quality is similar to that of 27 but the cost is higher. Then, 22 or 27? Solution 22 seems more appropriate because a 10% decrease in the budget would result in a reduction of more than 25% in quality, which may be too drastic.

So, the process of choosing one solution involves comparing the different objectives. However, if the decision-maker has a clear idea of how to compare them, then an *a priori* method would be more appropriate to handle the multiobjective problem .

We intend to present the results of this study to the managers of the HCC because in the initial stages of the research they provided us with the necessary data to build the model and they also showed great interest in the results. Nevertheless, to be of use to them, our study is still lacking in one final step, that is, to integrate the analysis in user-friendly software to provide the decision-maker with a useful means to define the scenarios and obtain usable reports.

4. Conclusions and future research

In this paper we have shown how simulation, combined with statistics and optimization techniques, is a valuable decision-making tool in the health care field. We have illustrated the proposed methodology by mod-

eling and analyzing the primary health center located in the Hospital *San Juan de Dios* of Pamplona (Colombia). Although the system is not very big in terms of infrastructure and human resources, etc., it has all the characteristics of a complex decision problem, such as randomness in the number and arrival times of the patients; random service times; simultaneous consideration of many performance measures, some of which are difficult to calculate; and many possible configurations for the system, making it impossible to evaluate all of them in order to select the "best one". We were able to deal with all these system characteristics.

Furthermore, the simulation software offers animation and graphical outputs which reassure managers (decision-makers) regarding the credibility of the model and the reliability of the analysis. They are able to visualize the effect of different management or operation policies, which is very important when decisions have to be taken by persons not specialized in the use of these quantitative research techniques.

Our main contribution is methodological in nature, having built a method that draws on statistics, probability, mathematical programming, simulation, multiobjective optimization and metaheuristics. Furthermore, as far as we know, such methodology has never been used in the analysis of real cases in health care settings.

Another, theoretical, contribution is the procedure to obtain the maximum likelihood classification of doctor service times in groups of patients. This classification is used to validate expert opinions and obtain the distribution of the service time using a form of Bayesian approximation in which triangular distributions play the role of *a priori* distributions. Furthermore, from the assignment it is possible to estimate the order in which patients entered the office. It can be used to test the patient arrival pattern for group-independence. In our model we assumed the independence hypothesis.

The advantage of using a generating method to handle the multiobjective problem is that it yields the full efficient set (or at least a good approximation). The disadvantages are the considerable computational effort required and, from a practical point of view, the difficulty for the decision-maker to select one solution from a very large set of efficient solutions. Cluster and filter techniques are used to help the decision-maker by reducing the set of solutions presented. When the decision-maker is well acquainted with the structure of the problem and is able to compare the different objective functions then he or she must specify his/her opinion or preferences before the solution process begins. These preferences could be incorporated into the mathematical model by using an *a priori* multiobjective method.

In our research we have used a multiobjective mathematical technique, the ϵ-constraint method, to estimate the Pareto frontier, but this simulation and optimization model could also incorporate other metaheuristic techniques. In this sense several methods have been proposed in recent years to solve real multiobjective problems ([27] , [36]). Among them, one of the easiest to implement and combine with simulation are evolutionary algorithms. See the books by Deb [23] and Coello et al. [19] for a deeper explanation.

Acknowledgement: The authors are grateful to two anonymous referees for having provided some additional and interesting references and for their comments, which have helped us to improve the presentation of this paper.

References

[1] Aharonson-Daniel, L., Paul, R., Hedley, A. (1996). Management of queues in out-patient departments: the use of computer simulation, Journal of Management in Medicine, 10, 50-58.

[2] Alexopoulos, C.; Goldsman, D.; Fontanesi, J.; Kopald, D.; Wilson, J. W. (2008). Modeling patient arrivals in community clinics. Omega, Vol. 36, 33-43.

[3] Altinel, I.K., Ulas, E. (1996). Simulation modelling for emergency bed requirement planning. Annals of Operations Research, 67, 183-21

[4] April, J., Glover, F., Kelly, J., Laguna, M (2003) Practical introduction to simulation optimization. Proceedings of the 2003 Winter Simulations Conference, S. Chick, P.J. Sánchez, D. Ferrin and D.J. Morrice, eds.

[5] Ashton, R., Hague, L., Brandreth, M., Worthington, D., Cropper, S. (2005). A simulation-based study of a NHS Walk-in Center, Journal of the Operational Research Society, 56, 153-161.

[6] Babad, H., Sanderson, C., Naidoo, B., White, I., Wang, D. (2002). The development of a simulation model of primary prevention strategies for coronary heart disease, Health Care Management Science 5, 269-274.

[7] Bailey, N.T.J. (1952). A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting times, Journal of the Royal Statistical Society, 14, 185-199.

[8] Beaulieu H, Ferland JA, Gendron B, Michelon P. (2000). A mathematical programming approach for scheduling physicians in the emergency room. Health Care Management Science, 3, 193-200.

[9] Blake, J., Carter, M. (2002). A goal programming approach to strategic resource allocation in acute care hospital, European Journal of Operational Research, 140, 541-561.

[10] Brailsford, S. C.; Gutjahr, W.; Rauner, M. S.; Zeppelzauer, W. (2007). Optimal screen policies for diabetic retinopathy using a new combined discrete-event simulation and ant colony optimization approach. Computational Management Science, Vol. 4, N0 1, 59-83.

[11] Brailsford S., Schmidt B. (2003). Towards incorporating human behaviour in models of health care systems: an approach using discrete event simulation. European Journal of Operational Research, 150, 19-31.

[12] Brennan J.E., Golden B.L., Rappoport H.K. (1992). Go with the flow: improving Red Cross bloodmobiles using simulation analysis. Interfaces, 22, 1-13.

[13] Cayirly, T., Veral E., Rosen H. (2006). Designing appointment scheduling systems for ambulatory care services, Health Care Management Science, 9, 47-58.

[14] Changkon, V., Haimes, Y.Y. (1983). Multiobjective Decision Making Theory and Methodology. Elsevier Science.

[15] Charnes, A., Cooper, W.W. (1961). Management models and industrial applications of linear programming. Wiley.

[16] Cheang B, Li H, Lim A, Rodrigues B. (2003). Nurse rostering problems, a bibliographic survey, European Journal of Operational Research, 151, 447-460.

[17] Chu, S., Lin, C.K., Lam, S.S. (2003). Hospital lift system simulator: a performance evaluator-predictor, European Journal of Operational Research, 146, 156-180.

[18] Cochran, J.K, Bharti, A. (2006) Stochastic bed balancing of an obstetrics hospital, Health Care Management Science, 9, 31-45.

[19] Coello, C. A.; Van Veldhuizen, D. A.; Lamont, G. B. (2002). Evolutionary algorithms for Solving Multiobjective Problems. Kluwer, New York.

[20] Davies, R., Crabbe D., Roderick P., Goddard J.R., Raftery J., Patel P. (2002). A simulation to evaluate screening for Helicobacter pylori infection in the prevention of peptic ulcers and gastric cancers. Health Care Management Science, 5, 249-58.

[21] Davies, R., Roderick, P., Raftery, J. (2003). The evaluation of disease prevention and treatment using simulation models. European Journal of Operational Research, 150, 53-66.

[22] De Angelis, V., Felici, G., Impelluso, P. (2003). Integrating simulation and optimisation in health care center management, European Journal of Operational Research, 50, 101-114.

[23] Deb, K. (2001). Multiobjective Optimization using Evolutionary Algorithms. Wiley, Chichester.

[24] 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.

[25] Fetter, R.B., Thompson, J.D. (1965). The simulation of hospital systems, Operations Research, 13, 689-711.

[26] Figueira, J., Greco, S., Ehrgott, M. (Eds.) (2005) Multiple Criteria Decision Analysis: State of the Art Surveys, in Series: International Series in Operations Research & Management Science, Vol. 78.

[27] Gandibleux, M, Sevaus, K, Sorensen, V, T'kindt (Eds.) (2004). Metaheuristics for multiobjective optimization. Lectures Notes in Economics and Mathematics Systems 535, Springer.

[28] Glover, F., Kochenberger, G. (2003), Handbook of metaheuristics, International Series in Operations Research & Management, Vol. 57, Kluwer Academic Publisher.

[29] Griffiths J.D., Price-Lloyd N., Dmithies M., Williams J.E. (2005). Modelling the requirements for supplementary nurses in an intensive care unit. Journal of the Operational Research Society, 56, 126-133.

[30] Gutjarhr, W. J., Rauner, M. S. (2007). An ACO algorithm for a dynamic regional nurse-scheduling problem in Austria. Computers & Operations Research, Vol. 41, No. 3, 642-666.

[31] Harper, P.R., Jones, S.K. l (2005). Mathematical models for the early detection and treatment of colorectal cancer, Health Care Management Science, 8, 101-109.

[32] Harper, P.R., Phillips, S., Gallagher, J.E. (2005). Geographical simulation modelling for the regional planning of oral and maxillofacial surgery across London, Journal of the Operational Research Society, 56, 134-143.

[33] Harper, P.R., Sayyad, M.G. et al (2003). A system modelling approach for the prevention and treatment of diabetic retinopathy, European Journal of Operational Research, 150, 81-91.

[34] Harper, P.R., Shahani, A.K. (2002). Modelling for the planning and management of bed capacities in hospitals, Journal of the Operational Research Society, 53, 19-24.

[35] Hupert N., Mushlin A.I., Callahan M.A. (2002). Modelling the public health response to bioterrorism: using discrete event simulation to design antibiotic distribution centers. Medical Decision Making, 22, 17-25.

[36] Jones, D.F., Mirrazavi, S.K., Tamiz, M. (2002), Multiobjective meta-heuristics: an overview of the current state of the art,, EJOR, 137, 1-9.

[37] Jun, J.B., Jacobson, S., Swisher, J. (1999). Application of discrete-event simulation in health care clinics: a survey, Journal of the Operational Research Society, 50, 109-123.

[38] Kim, S-C., Horowitz, I., Young, K.K., Buckley, T.A. (1999). Analysis of capacity management of the intensive care unit in a hospital, European Journal of Operational Research, 115, 36-46.

[39] Laguna, M., Martí, R. (2002). The OptQuest Callable Library. In Optimization Software Class Libraries, Stefan

Voss and David L. Woodruff (eds.), Kluwer Academic Publishers, Boston, pp. 193-218.

[40] Perrin, E. B.; Durch, J. S.; Skillman, S. M. (1999). Health performance neasurement in the Public Sector. National Academic Press.

[41] Petrovski, A., McCall, J. (2001). Multiobjective optimization of cancer chemotherapy using evolutionary algorithms, in Zitler, Deb, Thiele, Coello Eds., Evolutionary Multicriterion Optimisation, Lecture notes in Computer Science, n[a]1993, 531-544.

[42] Ratcliffe, J., Eldabi, T, Burroughs, A. et al. (2001). A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation, Health Care Management Science, 4, 117-124.

[43] Rauner M, Bajmoczy N. (2003). How many AEDs in which region? An economic decision model for the Austrian Red Cross. European Journal of Operational Research, 150, 3-18.

[44] Rauner, M.S., Brailsford, S.C., Flessa, S. (2005). Use of discrete-event simulation to evaluate strategies for the prevention of mother-to-child transmission of HIV in developing countries, Journal of the Operational Research Society, 5, 222-233.

[45] Ridge, J.C., Jones, S.K., Nielsen, M.S., Shahani, A.K. (1998). Capacity planning for intensive care units, European Journal of Operational Research, 105, 346-355.

[46] Rohleder, T., Klassen, K. (2002). Rolling horizon appointment scheduling: a simulation study, Health Care Management Science, 5, 201-209.

[47] Rossetti M.D., Felder R.A., Kumar A. (2000). Simulation of robotic courier deliveries in hospital distribution services. Health Care Management Science, 3, 201-13.

[48] Schaefer, S., Bryce, C., Alagoz, O., Kreke, J. et al. (2005). A clinically based discrete-event simulation of end-stage liver disease and the organ allocation process, Medical Decision Making, 25, 199-209.

[49] Stahl, J. E., Rattner, D., Wiklund, R., Lester, J, Beinfeld, M., Gazelle, S. (2004). Reorganizing the System of Care Surrounding Laparoscopic Surgery: A Cost-Effectiveness Analysis Using Discrete-Event Simulation, Medical Decision Making, 24, 461-471.

[50] Stevenson, M.D., Brazier, J.E., Calvert, N.W., Lloyd-Jones, M., Oakley, J.E., Kanis, J.A. (2005). Description of an individual patient methodology for calculating the cost-effectiveness of treatments for osteoporosis in women, Journal of the Operational Research Society; 56, 214-221.

[51] Swisher, J.R., Jacobson, S. (2002). Evaluating the design of a family practice healthcare clinic using discrete-event simulation, Health Care Management Science, 5, 75-88.

[52] Swisher, J.R., Jacobson, S.H., Jun, J., Balci, O. (2001). Modelling and analyzing a physician clinic environment using discrete event (visual) simulation, Computers and Operations Research, 28, 105-125.

[53] Vasilakis C, El-Darzi E. (2001). A simulation study of the winter bed crisis. Health Care Management Science, 4, 31-6.

[54] Vissers, J., Adan I., Dellaert, N. (2006). Developing a platform for comparison of hospital admission systems: an illustration, European Journal of Operational Research, (in press, can be read in www.science direct.com).

h

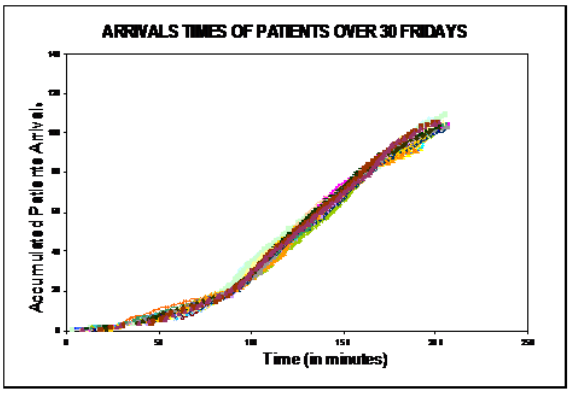

Fig. A.1. Daily accumulated number of patient arrivals. Data for Fridays.

Appendices

A Input modeling

As an example, the estimation of the *non-homogeneous* Poisson Process defined by arrival rate for Fridays is presented in this appendix.

Let us consider the patient arrival times from the 30 recorded Fridays. Time is measured in minutes starting at 5 a. m.

We estimate the arrival rate function in the following 4 steps.

Step 1.-*Observed accumulated number of arrivals over the 30 Fridays*
Figure A.1 shows the function $\Lambda_i(t) = \sum_{j=1}^{n_i} 1_{\{t_{ij} \leq t\}}$, that is, the accumulated number of arrivals against the time for the 30 Fridays.

Step 2.- *Aggregated accumulated number of arrivals*
We estimate the expected accumulated arrival functions by dividing the aggregated function by the aggregated number of days (cf. Figure A.2)

Step 3.- *Polynomial regression of the aggregated accumulated* number of arrivals

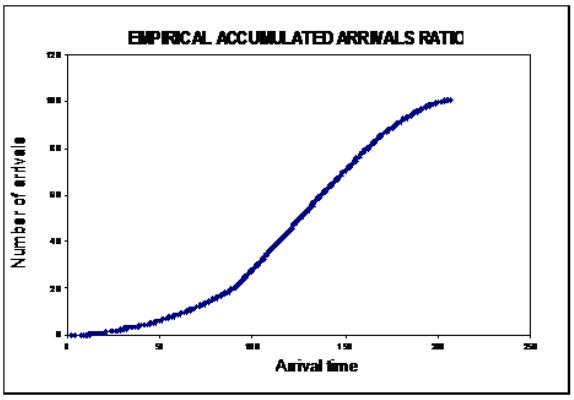

Fig. A.2. Average accumulated number of patient arrivals. Data for Fridays.

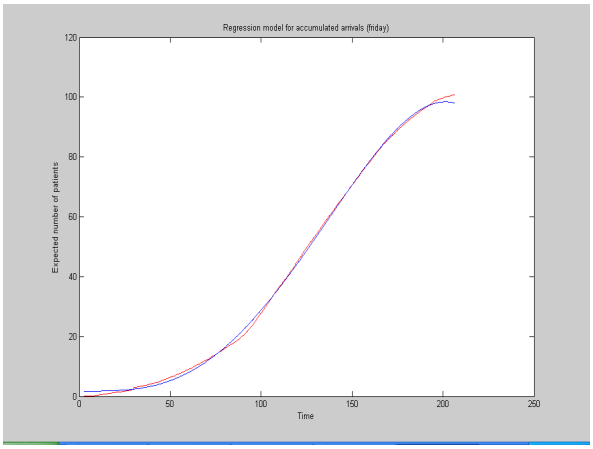

Fig. A.3. Regression model for the average accumulated number of patient arrivals. Data for Fridays.

We have considered a polynomial regression model. Figure A.3 shows the graphical result.
As result of the regression analysis we have
$\hat{\Lambda}(t)$= -20.6667 10^{-8} t^4+ 60.0043 10^{-6} t^3- 14.0187 10^{-4} t^2+ 1.7186 10^{-2}t+ 1.6728

Step 4.- *Estimated arrival ratio function*
We derive the polynomial function fitted in step 3 to obtain the estimation of the arrival ratio function $\lambda(t)$ (cf. figure A.4).
$\tilde{\lambda}(t)$ = -8.2667 10^{-7} t^3+ 18.0013 10^{-5} t^2- 2.8037 10^{-3} t+ 1.7186 10^{-2}

B Maximum likelihood classification

This method is based on the resolution of an assignment/transportation problem, which is a particular type

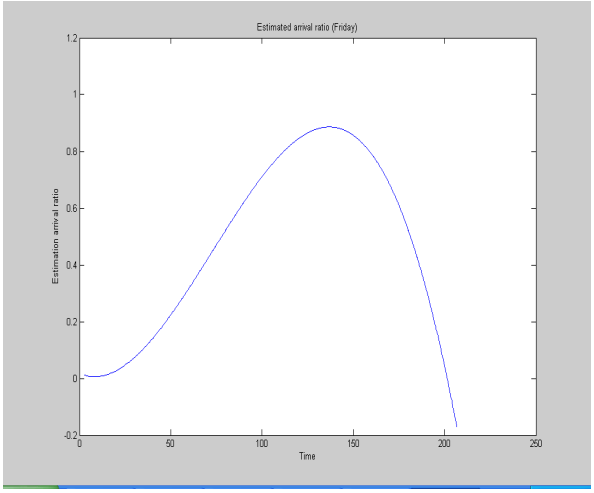

Fig. A.4. Estimated arrival rate for Fridays.

of integer linear programming problem. Given two sets, each with an equal number of items, the problem is to find an optimal full matching between the items of the two sets. Each pair formed by one item each set has associated a cost. The total cost of a full matching is the sum of the costs of each of the pairs involved. Thus, the optimal matching is the one with the minimum cost associated. In our case one set of items is formed by the set of service times recorded during a day and the other set is made up of the groups to which the patients seen by the doctors that day belong.

Consider the following notation: $P_1, \ldots, P_k$ are the k patients seen by the doctors on a certain day. Their service times, denoted by $t_1, \ldots, t_k$, respectively, have to be classified into r groups, $g_1, \ldots, g_r$, with frequencies $n_1, \ldots, n_r$, respectively. G represents the set of these r groups. We also denote by $l_j = \{(i, g(i)) \mid i = 1, \ldots, k;\ g(i) \in G\}$ one of the s different possible classifications, $g(i)$ being the group to which P_i has been assigned.

$$s = \frac{k!}{n_1! \times \cdots \times n_r!} \tag{B.1}$$

For each patient group g_i the density function $f_i(t)$ of the doctor service time is known (triangular distribution given by expert opinion). Then, the likelihood function $L(l_j)$ (or the "probability" of the sample of recorded times) is

$$L(l_j) = \prod_{i=1}^{k} f_{g(i)}(t_i) \tag{B.2}$$

The log-likelihood function is

$$\log L(l_j) = \sum_{i=1}^{k} \log f_{g(i)}(t_i) \qquad \text{(B.3)}$$

The assignment of patients to groups l_j^* maximizing the above function is the maximum likelihood classification. The solution to this maximization problem is obtained by solving an assignment/transportation problem when the cost of assigning a patient P_i to group g is $C(i, g) = \log f_g(t_i)$. In this case we have a maximization problem.

Observe that, when a recorded time t is beyond the positive variation range of a density function, then its associated assignment cost is -∞. This means that the patient with this time can not be assigned to the group described by this density. But, if there are many pairs associated with an infinite cost, then the problem could be infeasible. That is, there is no possible assignment in which all the pairs have a finite cost. Obviously, infeasibility will never arise if all the recorded times t have been drawn from the densities provided by the experts. Thus, an infeasibility situation is a symptom of poor estimation in the density functions. But, if our primary objective is to provide the "best" classification according to the expert opinions, then we should slightly modify the cost definition to avoid the infeasibility results.

A modified maximum likelihood assignment

We are going to modify the triangular densities $f(x)$ to prevent them from taking value zero. The idea, as shown in Figure B.1, is to substitute the triangular density in intervals $(0, L^+)$ and (U^-, ∞) with another function taking a value greater than zero at all points of both intervals.

We organize the procedure in the following steps.

- ***Step 1***. *Calculation of the cut points L^+ and U^-.*

These two points are determined by the probability α we want to leave in the tails. This probability is fixed *a priori* and should be assigned a low value to avoid overmodifying expert opinion. For the right side we have:

$$\alpha = (U - U^-)\frac{2(U - U^-)}{(U - m)(U - L)}\frac{1}{2} = \frac{(U - U^-)^2}{(U - m)(U - L)} \qquad \text{(B.4)}$$

Deriving U^- as function of α, we obtain

$$U^- = U - \sqrt{\alpha(U - m)(U - L)} \qquad \text{(B.5)}$$

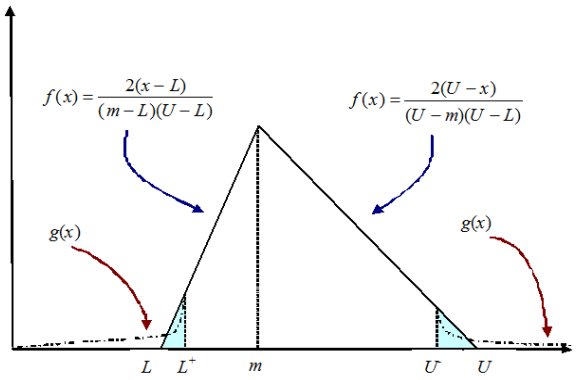

Fig. B.1. Modified Triangular Distribution.

For the left side we have:

$$\alpha = (L^+ - L)\frac{2(L^+ - L)}{(m - L)(U - L)}\frac{1}{2} = \frac{(L^+ - L)^2}{(m - L)(U - L)} \qquad \text{(B.6)}$$

Deriving L^+ as function of α, we obtain

$$L^+ = L + \sqrt{\alpha(m - L)(U - L)} \qquad \text{(B.7)}$$

Acceptable values for α could be those in the interval (0.01, 0.05). In any case the value α must be in the interval (0, (U - m) / (U - L)), for the right tail case, and in (0, (m - L) / (U - L)), in the left tail case.

- ***Step 2***. *Obtaining the new density function g(x) in* (U^-, ∞).

We consider the following exponential family of functions depending on two parameters: $g(x) = k\lambda e^{-\lambda x}$. Parameters k and λ are determined when we impose the following two conditions.

- ► The area under the curve in the interval (U^-, ∞) is α:

$$\alpha = \int_{U^-}^{\infty} g(x)\, dx = \int_{U^-}^{\infty} k\lambda e^{-\lambda x}\, dx = k e^{-\lambda U^-} \qquad \text{(B.8)}$$

- ► Continuity: functions $g(x)$ and $f(x)$ take the same value in $x = U^-$:

$$f(U^-) = \frac{2(U - U^-)}{(U - m)(U - L)} = k\lambda e^{-\lambda U^-} = g(U^-) \qquad \text{(B.9)}$$

Solving this two-equation system we obtain

$$\lambda = \frac{f(U^-)}{\alpha} \text{ and } k = \alpha e^{U^- f(U^-)/\alpha} \qquad \text{(B.10)}$$

Summarizing, we find

$$g(x) = f(U^-)\, e^{-f(U^-)(x-U^-)/\alpha} \qquad \text{(B.11)}$$

- ***Step 3***. *Obtaining the new density function* $g(x)$ *in* $(0, L^+)$.

We consider the following family of functions depending on two parameters: $g(x) = ae^{-b}$. Parameters a and b are determined when we impose the following two conditions:

- The area under the curve in the interval $(0, L^+)$ is α :

$$\alpha = \int_0^{L^+} g(x)dx = \int_0^{L^+} a\, x^{-b} dx = \frac{a}{1-b}(L^+)^{1-b} \qquad \text{(B.12)}$$

- Continuity: functions $g(x)$ and $f(x)$ take the same value in $x = L^+$:

$$f(L^+) = \frac{2(L^+ - L)}{(m-L)(U-L)} = a\, L^{-b} = g(L^+) \qquad \text{(B.13)}$$

Solving this two-equation system we obtain

$$b = 1 - \frac{L^+}{\alpha} f(L^+)$$

and

$$a = f(L^+) \times (L^+)^{\left(1-\left[\frac{L^+ f(L^+)}{\alpha}\right]\right)}$$

Thus, the density function is

$$g(x) = f(L^+) \left(\frac{x}{L^+}\right)^{-(1-(L^+ f(L^+)/\alpha))} \qquad \text{(B.14)}$$

Received 15 March 2007; revised 28 June 2007; accepted 17 October 2007

C Capacity Determination.

The capacity C_i for day i is defined as

$$C_i = \max\left\{s/P\left(\sum_{j=1}^{s} T_j \le WT_i\right) \ge p\right\} \qquad \text{(C.1)}$$

Then the p value represents the probability that the doctors' working time will be enough to serve C_i patients. Typically the p value will be 0.95.

All random variables T_j describe the service time for a generic patient j and are considered to be independent and identically distributed: $T_j \stackrel{d}{=} T$

Thus, it is verified that $\sum_{i=1}^{s} T_i \to N(s\delta E[T], s\delta^2 V[T])$.

By standardizing and operating, we obtain:

$$P\left[\sum_{j=1}^{s} T_j \le WT_i\right] \ge p$$

if and only if

$$p\left[z \le \frac{WT_i - s\delta E[T]}{\delta\sqrt{sVar[T]}}\right] \ge p$$

with Z the standard normal variable. Then, C_i is the value s verifying

$$\frac{WT_i - s\delta E[T]}{\delta\sqrt{sVar[T]}} \ge z_p \qquad \text{(C.2)}$$

From this expression, by squaring both sides of the above equation and solving the second order equation, it follows that:

$$C_i = \lfloor [(2\delta\, E[T]\, WT_i + z_p^2\delta^2\, V[T] - \sqrt{z_p^4\delta^4\, V[T]^2 + 4z_p^2\delta^3 E[T]\, V[T]\, WT_i}\,)\,] / 2\delta^2 E[T]^2 \rfloor$$

Let us note that, because $z_p = -z_{1-p}$, the two solutions correspond to

$$P[Z \le \frac{TW_i - s\delta E[T]}{\delta\sqrt{sVar[T]}}] \ge p$$

and

$$P\left[z \le \frac{wT_i - s\delta E[T]}{\delta\sqrt{sVar[T]}}\right] \ge 1 - p$$

Therefore, only the negative root should be considered.