Operations Research in Healthcare: a survey

Abdur Rais$^a$ and Ana Viana$^{a,b}$

$^a$INESC Porto, Campus da FEUP, Rua Dr. Roberto Frias 378, 4200-465 Porto, Portugal,
$^b$ISEP — Instituto Superior de Engenharia do Porto, Rua Dr. António Bernardino de Almeida 341, 4200-072 Porto, Portugal

E-mail: arais@inescporto.pt [Rais]; aviana@inescporto.pt [Viana]

Received 16 September 2009; received in revised form 21 January 2010; accepted 21 January 2010

Abstract

Optimisation problems in Healthcare have received considerable attention for more than three decades. More recently, however, with decreasing birth rates in nearly all of the developed countries and increasing average longevity globally, optimisation issues in Healthcare have become noticeably important and attract keen interest from the Operations Research community. Over the years, attention has gradually expanded from resource allocation and strategic planning to include operational issues such as resource scheduling and treatment planning.

This paper surveys several applications of Operations Research in the domain of Healthcare. In particular, the paper reviews key contributions addressing contemporary optimisation issues in this area. It highlights current research activities, focusing on a variety of optimisation problems as well as solution techniques used for solving the optimisation problems.

Keywords: healthcare; operations research; optimisation; survey

1. Introduction

How to decide on the best locations of medical clinics and emergency vehicles for providing maximum Healthcare coverage to a given population? How many base locations of medical ambulances are needed if the total distance from the locations to the hospitals must be less than a given number? How should radiation treatment be planned for minimising the treatment time of a cancer patient? How should the nurses in a trauma centre be scheduled and re-rostered to maintain an adequate service level even in the worst-case scenario? Many problems like these need to be addressed in Healthcare, and Operations Research (OR) provides numerous methodologies and solution techniques for tackling them.

Although optimisation problems in Healthcare have received considerable attention during the last few decades, many issues are becoming much more important and relevant now because of...
the growth in ageing population from decreasing birth rates in nearly all of the developed countries and increasing longevity globally. Moreover, enormous public and private funds required to cover the rapidly escalating Healthcare costs also necessitate a much closer scrutiny for cost-saving measures. For instance, the United States spends more than $2 trillion or 16% of its GDP on Healthcare even though 50 million Americans do not have health insurance and another 25 million remain underinsured (Catlin et al., 2008). Without a doubt, Healthcare is currently one of the main social and economic problems in the United States.

This paper surveys selected research work in OR applied exclusively to the problems in Healthcare. OR is being utilised much more now to address day-to-day hospital management, resource-constrained operations or treatment planning aspects in Healthcare (Royston, 2009). Current research focus has led to the discovery of many new optimisation issues and the development of many new solution approaches. Key healthcare optimisation issues include service planning, resource scheduling, logistics, medical therapeutics, disease diagnosis and preventive care. In particular, this survey addresses more recent and relevant works that are not known to have been covered in other survey papers. It also provides the reader with information on seminal survey papers, books and research centres addressing relevant research in this area.

The current section introduced the objectives of this paper. Planning aspects of healthcare are surveyed in Section 2. Works related to healthcare management and associated concerns are discussed in Section 3. Healthcare practice is addressed in Section 4, followed by specialised and preventive care in Section 5. Section 6 gives some additional relevant information while Section 7 concludes with a few remarks.

2. Healthcare planning

The importance and significance of planning in healthcare can hardly be overemphasised today when providing proper and adequate service continues to be a key concern of most countries. For growing longevity and ageing population amidst dwindling birth rates, many countries are increasingly hard-pressed for the extra budget and resources to meet the healthcare needs. Many have turned to OR for optimisation and cost-control measures. Some of the key healthcare issues considered in OR today include estimation of future demand for services in order to build enough capacity, selection of hospital locations for covering a target population and design of the emergency facilities for efficient handling of patients. A powerful tool for planning is Simulation (cf. Jun et al., 1999; Fone et al., 2003; Ashton et al., 2005; Eldabi et al., 2007; Hall, 2006; Gunal and Pidd, 2009, 2010). But several other techniques have been successfully used, as shown in the following sections.

2.1. Demand forecasting

Accurate demand forecasting is essential in healthcare planning, its results providing the input to several optimisation problems. While forecasting methods can be qualitative or quantitative, most research work has focused on quantitative analysis because of better accuracy. However, the availability of good, historical data is essential for quantitative methods.
Finarelli and Johnson (2004) gave a detailed, nine-step, quantitative demand forecasting model for healthcare services, while Cote and Tucker (2001) discussed four common methods for forecasting demand for healthcare services: percent adjustment, 12-month moving average, trendline and seasonality. The accuracies of various forecast methods were evaluated by Jones et al. (2008). They used data from daily patient arrivals at the emergency departments of three different hospitals and considered the following methods: time series regression, exponential smoothing, seasonal autoregressive integrated moving average and artificial neural network models.

Beech (2001) derived market-based healthcare service forecasting from a broad range of available data for estimating the future demand. The data sets involve primary as well as secondary service areas, service-area populations by various demographic groupings, discharge utilisation rates, market size and market share by service lines. They found that market dynamics can allow a variety of explicit assumptions and trends for developing scenarios of potential future demand.

A two-step approach for forecasting future demand along with capacity needs was given by Myers and Green (2004). In particular, their approach develops a facility master plan incorporating the projected capacity as well as physician requirements.

Xue et al. (2001) analysed the continued growth of the end-stage renal disease population in the United States. They forecasted up to the year 2010 using historical data with stepwise autoregressive and exponential smoothing models.

2.2. Location selection

A large body of publications in healthcare addressed various issues concerning capacity management and location selection, both for healthcare services and medical material (see, e.g., Ndiaye and Alfares, 2008; Araz et al., 2007; Bruni et al., 2006; Rauner and Bajmoczy, 2003; Verter and Lapierre, 2003).

Daskin and Dean (2004) reviewed location set covering, maximal covering and P-median models for addressing the location planning issues in healthcare. They presented a novel application of the set covering model for analysing cytological samples. Additional review of location models in healthcare can be found in Smith-Daniels et al. (1988). Review of location and allocation models restricted to developing countries can be found in Rahman and Smith (2000).

2.2.1. Healthcare centres

Where to set healthcare centres for maximising accessibility has been extensively studied by OR practitioners.

Location problems arising in developing countries have been discussed by several authors. Smith et al. (2009) studied planning of sustainable community healthcare in the rural areas of developing countries. They considered both top-down and bottom-up hierarchical location models for the efficient planning of community health schemes and proposed a Mixed Integer Program for determining the locations of maximal number of sustainable facilities.
Murawski and Church (2009) consider the problem of improving health service accessibility by upgrading links to the existing facility locations of the transport network to all-weather roads. Their integer-programming model is adequate for rural areas of underdeveloped countries where, during bad weather conditions, accessibility is diminished because of the lack of all-weather roads. Their model addressed a real-world problem scenario in Ghana.

A slightly different problem is solved in Ndiaye and Alfares (2008). They study the problem of selecting public service locations for nomadic population groups that seasonally change their locations. The authors have presented a binary integer-programming model to determine the optimal number along with the locations of the primary healthcare units that can satisfy seasonally varying demands.

Rahman and Smith (1999) addressed the problem of finding additional suitable sites for healthcare facilities in rural areas and modelled the problem as a maximal covering location model. They solved the model using heuristic methods and data from Bangladesh.

Finally, Hodgson et al. (1998) used the covering tour model to plan mobile health services such that a sufficient number of facilities are geographically accessible as well as sustainable. Their integer linear-programming model minimises mobile facilities travel for serving all the population centres within a certain feasible range. They used exact as well as heuristic methods to solve the model and gave computational results using data from Ghana.

Additional location selection problems are discussed in Aaby et al. (2006a). This work presented a simulation tool for planning the locations of mass vaccination clinics in case of an influenza pandemic. Preventive healthcare facilities location is also studied by Verter and Lapierre (2003). They addressed this location problem so that participation in the prevention programs is maximised. They gave a mathematical programming formulation and demonstrated the results using data from Georgia, USA and Montreal, Canada.

Cote et al. (2007) addressed traumatic brain injury problems involving active duty military personnel. As a timely treatment is critical for achieving recovery, being near a medical centre with specialised treatment services can be a critical factor. They developed a Mixed Integer Program for locating new treatment units while minimising the sum of patient costs (treatment, lodging, travel) along with the penalties associated for foregone treatment revenue and excess capacity utilisation.

The organisational structure of a transplant system in Italy is the focus of the work by Bruni et al. (2006). They proposed a location model based on mathematical programming for optimising such a structure. In particular, they considered the critical role of time in the transplant process and the spatial distribution of the transplant centres. The objective of the model was to allocate transplantable organs across various regions with the strategic goal of achieving regional equity in healthcare.

Marianov and Taborga (2001) address location planning for public healthcare centres having the goal of maximising coverage for the low-income population within a pre-specified distance and Harper et al. (2005) are concerned with service planning when the geographical locations of the services, as well as the patients needing services, are important altogether. A stochastic, geographical simulation model is developed to capture the underlying process of patients travelling to the providers while accounting for the necessary resource capacities, variability in patient needs and travel considerations.
Joint capacity planning and location selection for determining locations and sizes of the medical departments in a hospital network is addressed by Stummer et al. (2004). They proposed a multiobjective decision-support system having mathematical-programming and real-world data from hospitals in Germany. Four objectives are considered: minimise total travel costs incurred by the patients; minimise total costs associated with location-allocation in the hospital plan; minimise the number of patients who would have to be rejected as a consequence of low service capacities (shortage of beds); and minimise the number of unit moves necessary to restructure the current allocation.

Joint location allocation is also addressed by Chu and Chu (2000) to study the planning issues of hospital location and service allocation for new service distribution as well as existing service redistribution. They developed a general modelling framework for supply and demand matching and solved their goal-programming model using data in Hong Kong.

More recently, the work by Zhang et al. (2009b) addressed preventive healthcare facilities network design, regarding the number of facilities and their location, with the aim of maximising the participation of the population in preventive healthcare programmes.

Pierskalla and Brailer (1994) and Pierskalla (2004) gave a comprehensive overview of blood banking supply chain and addressed various questions concerning blood banking functions and locations. In particular, they analysed how many community blood centres should be in a region and where they should be located; how supply and demand should be coordinated; and how donor areas should be covered by community blood centres. They discussed tactical as well as operational issues confronting the collection of blood, deciding on inventory levels, allocating blood to hospitals and delivery to different sites.

### 2.2.2. Emergency vehicles

Although on a smaller scale, location of emergency vehicles is also addressed in the literature, as follows.

A facility location model for ambulances that minimises the number of units needed for performing at pre-specified service levels was developed by Ingolfsson et al. (2008). They incorporated uncertainties and randomness into their convex optimisation model using actual data from Edmonton, Canada. They found the model to be tractable with general-purpose optimisation solvers for cities with population of up to one million.

Araz et al. (2007) addressed the problem of determining the best base location of a limited number of emergency vehicles for optimising service levels. Three objectives measure the service levels: maximisation of the population covered by one vehicle; maximisation of the population with backup coverage; and minimisation of the total travel distance from locations at a distance bigger than a pre-specified distance. They used a fuzzy, multiobjective model for solving the problem.

Henderson and Mason (2004) discussed a simulation and analysis software tool for ambulance service planning. The decision support tool has various advanced features for operations management at all levels (including strategic and tactical). It currently helps several organisations in Australia, New Zealand and the United States for providing efficient ambulance services.

Finally, a trauma resource allocation model involving both hospitals and ambulances for maximising coverage for severely injured patients is discussed by Branas et al. (2000). They solved their integer-programming model by exact and heuristic methods using data from Maryland, USA.
2.3. Capacity planning

Hospital capacity planning is full of challenging problems for OR practitioners. Green and Savin (2008), for example, used OR-based analyses to address the increasingly critical hospital capacity planning decisions. They used a queuing model formulation and gave examples of how OR models can be used for deriving important insights and operational strategies.

A cooperative solution approach for hospital capacity to treat emergency patients in the Netherlands is proposed in Litvak et al. (2008). They presented a mathematical method inspired by the overflow model of phone calls in telecommunication and used simulation for validating the results. Emergency departments are also studied in Mayhew and Smith (2008). They used a queuing model to evaluate the emergency departments of hospitals in United Kingdom. Per government mandated targets, they focused on completing and discharging 98% of patients within four hours. Finally, Lovejoy and Li (2002) studied the problem of deciding whether the capacity of operating rooms in a hospital should be met by building new operating rooms or extending working hours in the current ones. They developed a multiobjective model with three objectives for addressing this issue: (1) minimise wait to get on schedule, (2) maximise schedule procedure start time reliability and (3) maximise hospital profit. They focused on the trade-offs between objectives and validated the model using simulation.

At the tactical level Adan et al. (2009) addressed the problem faced at a cardiothoracic surgery centre for optimising resource utilisation. They modelled it as a Mixed Integer Program having stochastic lengths of stay. Their results suggest that, for realising a given target of patient throughput, master surgical schedules with better performance on target utilisation levels of resources can be generated by considering a stochastic version of the deterministic problem.

Estimation of unavoidable hospital costs per unit of measurable output and identification of the sources of the inefficiencies is studied by Oliveira and Bevan (2008). They developed a stochastic multilevel model and applied it to hospitals in Portugal. Various inefficiencies arising from bed distributions, incentives and economies of scale were highlighted. The model could capture cost variations at different kinds of hospitals.

Discrete event simulation was used by van Berkel and Blake (2007) to analyse the waiting lists for surgery in Nova Scotia, Canada. They developed a model for capacity planning decisions and analysed performance measures for enhancement as well as embellishment.

Chapter 2 of Brandeau et al. (2004) is entirely devoted to capacity planning and management in hospitals. A considerable amount of work has also been carried out on bed capacity planning (see Smith-Daniels et al., 1988; Kokangul, 2008; Harper and Shahani, 2002; Ridge et al., 1998).

3. Healthcare management and logistics

Patient scheduling, resource scheduling and logistics in healthcare are likely the most extensively referenced management problems in traditional OR journals. In particular, nurse scheduling has deserved special attention (see Burke et al., 2004; Cheang et al., 2003).
3.1. Patient scheduling

An optimised patient–staff/patient–facilities schedule can lead to a considerable cost reduction and an increase in service quality.

The problem of scheduling patients for surgery at a hospital in Sweden given certain medical, economic and time constraints is addressed by Persson and Persson (2009). To prevent long queues, patients are allowed to be scheduled for surgery at other hospitals according to the Swedish health policy. They developed a hybrid simulation and integer-programming approach for solving the problem.

Patient–staff scheduling is studied in Ogulata et al. (2008). They developed hierarchical mathematical models to generate the schedules and considered three sub-problems: (1) selection of patients; (2) assignment of patients to the staff; and (3) scheduling of the patients throughout the day. The objectives are to maximise the number of selected patients, to balance the workloads of physiotherapists, and to minimise waiting times of the patients on their treatment days. They tested their models with real-world data from hospitals.

Dynamic generation of patient schedules, having varying priorities, to public health-care facilities are modelled by Patrick et al. (2008) as a Markov decision process. The model is solved using Approximate Dynamic Programming and the quality of solutions is analysed through simulation. Simulation is also used in Patrick and Puterman (2007) to improve resource utilisation for diagnostic services through flexible inpatient scheduling.

Green and Savin (2008) addressed last-minute cancellations of patient appointments and modelled the problem as a single-server queue. They derived stationary distribution of the queue size and gave extensive computational results as well as guidance for relevant future work.

In another line, Jiang and Giachetti (2008) addressed patient flow in outpatient facilities and considered parallel activities for the reduction of cycle times. They improved on existing multi-class, open queueing, network models and demonstrated that parallelisation can indeed reduce the cycle times for the patients needing multiple diagnostic or treatment procedures. They compared the results with those obtained through simulation. Kaandorp and Koole (2007) considered outpatient appointment scheduling and derived a local search procedure that converges to the optimal schedule when the objective is a weighted average of expected waiting times of patients, idle time of the doctor and tardiness (lateness). They found that, for a certain combination of the parameters, the Bailey–Welch rule gives an optimal appointment schedule. A review on outpatient scheduling is provided by Cayirli and Veral (2003).

Genetic Algorithms were used by Chien et al. (2008) for scheduling patients and increasing the service quality through the reduction of patient waiting times and the increase of therapy equipment utilisation. The problem is modelled as hybrid shop scheduling. A Mixed Integer Program is used as a benchmark for validating the solution quality. Empirical data from a medical centre in Taiwan are used to validate the approach.

Rauner et al. (2008) developed an internet game, based on discrete event simulation, to illustrate the economic and organisational decision-making processes in hospitals. Managements from up to six hospitals can play for the patients of various disease categories and budgets; players can assess options for capacity planning and patient scheduling through an analysis of resources, processes and financial results.
In Goddard and Tavakoli (2008) a queuing model to evaluate patient welfare under different systems is used to manage the waiting lists for public health services. They analysed various rules applied to the queueing model and discussed patient welfare as well as service efficiency aspects.

Queuing models are also used by Shmueli et al. (2003). They are based on the probability distribution of the number of occupied beds and aim at optimising admission to the intensive care unit. They validated the model using data from a hospital in Israel with the objective of maximising the expected incremental number of lives saved by operating at the unit.

Challenges and research opportunities for appointment scheduling are discussed in Gupta and Denton (2008).

3.2. Resource scheduling

Resource scheduling has been an active area of research in healthcare for increasing capacity utilisation, cost control measures and improving tactical as well as operational efficiencies of services and facilities. Special attention has been given to nurse scheduling but several other issues have also been addressed.

3.2.1. Nurse scheduling

In general, the nurse scheduling problem is that of assigning shifts to nurses having different skills while satisfying as many soft constraints and personal preferences as possible. A nurse schedule will typically strive to meet the required personnel coverage over a predefined planning period. Recent publications on nurse scheduling include Burke et al. (2006, 2008); Bard and Purnomo (2005b).

In a recent paper, Grano et al. (2009) developed a two-stage approach for nurse scheduling that considers both nurse preferences and hospital requirements. An optimisation model first awards shifts to the nurses with highest bids while satisfying all hospital constraints; subsequent optimisation then allocates the remaining shifts to all the nurses for completing the schedule. They applied the approach to a case study involving the emergency department of a hospital in Pennsylvania, USA. The problem is also addressed in Burke et al. (2006), which gave a real-world nurse-rostering methodology that is more flexible compared with the traditional fixed period-based approach. Burke et al. (2008) combined heuristic ordering with Variable Neighbourhood Search and Sundaramoorthi et al. (2009) developed a simulation approach to evaluate nurse–patient assignments. They used real data sets from a hospital in Texas to demonstrate their results.

In certain situations, patients may need to be assigned to specific nurses having special skills. Punnakitikashem et al. (2008) gave a stochastic programming approach for such nurse-to-patient assignments and Mullinax and Lawley (2002) gave mathematical programming and heuristic approaches for assigning nurses to critically ill newborn infants.

A problem deserving special attention is the nurse rerostering problem. It arises when at least one nurse announces that she will be unable to undertake the tasks assigned previously to her. Pato and Moz (2008) solved this problem using a Genetic Algorithm while Moz and Pato (2005) tackled it as a bi-objective problem and solved it using Goal Programming.
The problem of minimising the number of nurses that visit a patient (also known as “nurse–patient loyalty”) is addressed by Steeg and Schröder (2007). For the solution approach, they proposed a hybrid method combining Constraint Programming with a large neighbourhood search metaheuristic. Along a similar line of research, Bertels and Fahle (2006) considered the problem of home healthcare and nursing; i.e., visiting and nursing patients in their homes. Besides finding the nurse schedules, the associated routing problem is also considered: each nurse must be able to visit patients in a given roster using public transport. The goal is to minimise transportation costs and maximise satisfaction of both patients and nurses. A combination of Linear Programming, Constraint Programming and Metaheuristics is used to solve the problem. 

Bard and Purnomo (2005b) addressed a nurse preference scheduling problem having both hard and soft constraints and solved using a combination of heuristic and integer-programming methods. They extended the work further in Bard and Purnomo (2005a) and improved the quality of schedules by incorporating a downgrading option. They also presented a model in Bard and Purnomo (2007) that combines cyclic as well as preference scheduling by formulating an Integer Program and decomposing it via Lagrangian relaxation.

The model presented in Felici and Gentile (2004) maximises staff satisfaction using a polyhedral approach. The authors described a particular sub-problem whose solution can be used to determine strong cuts for the complete problem and design branching rules to break the symmetries in the solution space. They demonstrated that these rules strongly impact the efficiency of the method. 

Multiobjective personnel rostering is discussed in Berrada et al. (1996). The authors associated a slack variable to several soft constraints and considered objectives that minimise the slack values in a nurse scheduling problem. They used three different approaches for solving the problem: Sequential Technique, Equivalent Weights Technique, and Tabu Search. See Steuer (1985), Sherali (1982) and Glover and Laguna (2002), respectively, for details on the methodologies.


Some healthcare problems addressed personnel and facilities scheduling simultaneously; for instance, Belien and Demeulemeester (2008) integrated nurse scheduling with operating room scheduling and solved the problem using Linear Programming with Column Generation.

Personnel scheduling (mainly nurse scheduling or rostering) is extensively analysed in the survey papers by Ernst et al. (2004), Burke et al. (2004) and Cheang et al. (2003). Early research in nurse rostering is given in Pierskalla and Rath (1976) and Warner et al. (1990).

3.2.2. Operating room and physician scheduling

Although of much interest, the problem of scheduling physicians and/or operating rooms in healthcare institutions has not received as much attention as nurse scheduling (cf. Carter and Lapierre, 2001).
Operating room scheduling is discussed in Blake et al. (2002) and Blake and Donald (2002), among others. Their models can determine how much operating room time each surgeon should be assigned but, in most cases, cannot give detailed information on the timing of the assignments. Similarly, heuristics based on partial branch-and-bound are used in Beaulieu et al. (2000) for scheduling physicians in emergency rooms such that the sum of penalties associated with “deviation” constraints is minimised.

A different set of objectives is discussed in Ozkarahan (2000), who proposed a Goal Programming procedure for minimising idle time and overtime of operating rooms while simultaneously maximising satisfactions of the surgeons, the patients and the staffs.

Lamiri et al. (2008) proposed a stochastic model for operating room planning with two types of demand for surgery: elective and emergency. Elective surgery can be planned ahead and have patient-related cost depending on surgery date. Emergency surgery arrives randomly and must be performed on the day of arrival. The planning problem consists of assigning elective cases to different periods over a planning horizon in order to minimise the sum of the elective patient-related costs and overtime costs of the operating rooms. They proposed a method combining Monte Carlo simulation with mixed Integer Programming.

An Integer Programming model is developed in Santibanez et al. (2007) to schedule surgical blocks for each specialty into operating rooms, considering operating room time availability and post-surgical resource constraints. The model was used by the hospitals in a British Columbia Health Authority. An Integer Programming model, together with Simulation, is also used in Zhang et al. (2009a) for allocating operating room capacity to specialties on a weekly basis. Minimisation of inpatients’ cost (measured by their length of stay) is expected.

A new approach for surgical case scheduling in hospitals through the extension of a Job Shop scheduling problem was developed by Pham and Klinkert (2008) while Green (2004) studied the problem of adjusting staff in emergency rooms based on patient arrival rates, so that timely care of the patients is optimised. He evaluated the effectiveness of queueing models to identify the staffing patterns for reducing the fraction of patients who leave without being seen.

The problem of optimising the assignment of surgeries and sufficient planned slacks to the operating days such that the risk of working overtime is minimised, no surgeries are cancelled and operating room utilisation is improved is addressed by Hans et al. (2008). They proposed several constructive and local search heuristics.

Resident on-call scheduling concerns work-nights of residents considering departmental staffing needs, skill requirements and resident preferences. The problem is solved in Wang et al. (2007) using Genetic Algorithms. In Sherali et al. (2002), the resident on-call scheduling problem is modelled as a Mixed Integer Program and solved by heuristic procedures that exploit the network structure embedded in the model.

Molema et al. (2007) analysed part-time work options for doctors. They illustrated through a simulation of two case studies that introduction of part-time work can indeed improve service offerings and system design.

Concerns on bed occupancy are reflected in the work by Belien and Demeulemeester (2007). They addressed the problem of building surgery schedules with levelled resulting bed occupancy and considered two types of constraints: (1) demand constraints to ensure that each surgeon (or surgical group) obtains a specific number of operating room blocks, and (2) capacity constraints...
to limit the available blocks each day. Heuristics based on mixed Integer Programming and a
metaheuristic are developed to minimise the expected total bed shortage.

The schedule of emergency medical residents is constrained by a large number of rules: limits on
number of consecutive work hours, number of day and night shifts that should be worked by each
resident, resident staffing requirements according to seniority levels for the day and night shifts,
restrictions on the number of consecutive day and night shifts assigned, vacation periods, weekend
time-off requests and a fair distribution of responsibilities among the residents. Because some of
these rules are conflicting, Topaloglu (2006) studied the problem from a multiobjective perspective.

Finally, Carter and Blake (2004) discussed four different problem instances in healthcare for
which simulation can prove to be a good analysis tool: inpatient surgical scheduling; emergency
room waiting time; emergency room scheduling; and medicine ordering. They discussed obstacles
as well as lessons learned from the simulation modelling techniques used.

Additional work on the scheduling of operating rooms can be found in Puente et al. (2009),
Guinet and Chaabane (2003) and Dexter and Traub (1999). A review on operating room planning
and scheduling is provided by Cardoen et al. (2010).

3.3. Logistics

Logistics in healthcare is concerned, among others, with the definition of medicine stock levels
and allocation of resources.

Inventory needs are studied by Little and Coughlan (2008). They developed a constraint-
programming optimisation model for the problem of inventory needs in hospitals constrained by
the importance associated with various product and service levels. Their model can determine
optimal stock levels of products restricted by the space, delivery and criticality of the items. It was
validated using sterile and bulk items in an intensive care unit of a hospital in Ireland.

Lapiere and Ruiz (2007) addressed hospital logistics by coordinating the procurement and
distribution operations while respecting inventory capacities. Jacobson et al. (2006) calculated
paediatric vaccine stockpile levels necessary for avoiding vaccination interruptions and Swaminathan
(2003) proposed a multiobjective optimisation heuristic for allocating scarce drugs to clinics, having
taken into account efficiency, effectiveness and equity of the drug-allocation process.

Rauner and Bajmoczy (2003) addressed the allocation of medical materials (semi-automated
ey early defibrillators) to different regions of Austria. The problem considered acquisition and
maintenance costs of defibrillators; training costs for emergency medical technicians; hospitalisa-
tion costs for patients with sudden cardiac arrests; future healthcare costs for surviving patients
discharged from hospitals; and improved survival as well as quality-of-life benefits from using the
defibrillators. They developed a decision support system based on integer-programming models.

van de Klundert et al. (2008) showed that up to 20% cost reduction is possible in the
Netherlands through the optimisation of the flow of sterile instruments between sterilisation
departments and hospital operating theatres along with streamlining of the processes and
standardisation of the materials. They showed the general instance of the problem to be NP-hard
and gave an integer-programming formulation of a dynamic, non-deterministic model. They
specifically addressed optimisation of the flow between the central sterilisation departments and the
operating theatres by developing a cost minimisation model involving transportation and

© 2010 The Authors.
International Transactions in Operational Research © 2010 International Federation of Operational Research Societies
inventory. For this particular instance of the problem, they used Dynamic Programming to find a solution in polynomial time.

3.4. Others

Several researchers have looked into manufacturing and management areas of healthcare. This section discusses a selected few.

Aktin and Ozdemir (2009) developed a two-stage approach for coronary stent manufacturing from a one-dimensional cutting stock problem. The first stage calculates the number of patterns to be cut and the second stage finds the cutting plan. The problems were subject to demand, material availability, regular as well as overtime working hours and due date constraints. They used two integer-programming models: one for solving the problem of minimising trim loss and the other for minimising the total cost associated with material inputs, number of set-ups, labour hours and overdue time.

Haijema et al. (2007) discussed production and inventory management of platelets at a blood bank. They combined Markov Dynamic Programming with simulation and applied the approach to a real-world case of a Dutch blood bank.

Flessa (2003) developed a linear-programming model to analyse the optimum allocation of budget to a set of healthcare resources (Prevention, Dispensaries, Health Centres, District Hospitals and Regional Hospital) in Tanzania. They addressed five different objectives: minimisation of death cases, minimisation of years of life lost, minimisation of incidence, minimisation of prevalence, and minimisation of loss of quality of life.

Kongnakorn and Sainfort (2004) reviewed health outcome modelling in general and the detailed theoretical background for a Quality-Adjusted Life Years model grounded in OR and Utility Theory. They analysed various measures of health outcomes and gave comprehensive account of current issues and research directions.

Richter (2004) focused on clinical trials and discussed using the best currently available information for decision making in order to bridge the gap until better information becomes available. They showed that the combination of OR and Health Economics has an enormous potential for providing practical information to the decision makers.

Blood bank management policies, decisions and theories of perishable inventories are reviewed in Prastacos (1984) and Nahmias (1982). Although this appeared to be a promising area of research, OR/MS research in this area decreased considerably between the mid-1980s and 2004 as pointed out in Pierskalla (2004). Recently, Kopach et al. (2008) discussed management of red blood cell supplies based on two demand rates: urgent and non-urgent. They addressed the new supply needs and multi-tier control policies in view of extended expiration dates.

4. Healthcare practice

Besides addressing various other healthcare management aspects, OR researchers have made significant contribution to drug treatment planning, infectious disease prevention and control, pandemic preparedness, emergency response and organ donation.
4.1. Disease diagnosis

The use of optimisation techniques for disease diagnosis is extensively discussed in Lee and Wu (2009). A review on the application of the Analytic Hierarchy Process to numerous problems in medical decision making is also provided in Liberatore and Nydick (2008). They summarised 50 different articles from seven categories: diagnosis, patient participation, therapy/treatment, organ transplantation, project and technology evaluation and selection, human resource planning and healthcare evaluation and policy.

Abnormal brain activity is investigated in Chaovalitwongse et al. (2003). They applied global optimisation and dynamical systems for predicting epileptic seizures. Brain activity is also addressed by Chaovalitwongse et al. (2006) and Chaovalitwongse et al. (2008). The former applied optimisation-based data mining techniques to classify the brain’s normal and epilepsy activities using intra-cranial electroencephalogram. The results of their studies suggest the possibility of the design and development of efficient seizure warning algorithms for diagnostic and therapeutic purposes. The latter uses a new classification scheme for multidimensional time-series data. They used a mathematical-modelling framework for minimising errors in classification (or maximising accuracy).

Cancer diagnoses are the focus of the work by Bortfeld et al. (2008) and Sofer et al. (2003). Bortfeld et al. (2008) addressed uncertainties in radiation therapy for cancer patients. They gave a robust optimisation formulation of the problem that generalises mathematical-programming formulations. They also discussed extensive computational results using clinical data. Sofer et al. (2003) proposed an optimisation model to determine optimal biopsy protocols; that is, the zones to be biopsied in order to maximise the probability of cancer detection.

Paltiel et al. (2004) introduced “Asthma Policy Model” as a Markov state-transition simulation to forecast asthma-related symptoms, acute exacerbations, quality-adjusted life expectancy, healthcare costs and cost effectiveness. They gave extensive discussions of the results from simulation on a 10-year horizon.

Finally, coronary risk prediction is addressed by Alexe et al. (2003). The authors introduced the Logical Analysis of Data and discussed how it can be used for disease prediction.

4.2. Treatment planning

Within Medical Therapeutics, an area that has been receiving special attention is radiation therapy (Preciado-Walters et al., 2006; Holder and Salter, 2005). Still, several other areas of planning and intervention have been the subject of study as reflected in the papers by Paltiel et al. (2004), Lee et al. (2008) and Zenios (2004).

Recently, Holder (2004) gave a comprehensive discussion of linear- and non-linear programming models for Intensity Modulated Radiotherapy Treatment (IMRT). In particular, they focused on the linear models in detail. Furthermore, Ehrgott et al. (2008) gave an extensive survey of the optimisation models, methods and theories concerning IMRT design.

IMRT, as well as three-dimensional conformal radiotherapy (3DCRT), are addressed in Ferris et al. (2004). The problems are described in quadratic, linear, piece-wise linear and non-linear
formulations. The authors gave details on how to generate practical solutions and discussed several optimisation tools and programming environments.

Minimisation of the total treatment time in cancer radiotherapy using multileaf collimators is studied by Wake et al. (2009). The approach considers a Mixed Integer Program that happens to be a modification of a cutting-stock problem formulation. Maillart et al. (2008) developed a Markov chain model for investigating the proper frequency of mammography screening. They analysed a broad range of screening policies and discussed computational results.

Kidney dialysis therapy initiation for evaluating cost and effectiveness is investigated in Lee et al. (2008). They used Approximate Dynamic Programming and Simulation to determine an optimal therapy and a strategy for maximising patient welfare.

Lee and Zaider (2004) described a clinical decision support system for treatment planning in Brachytherapy (placement of radioactive seeds inside a tumour) and used Mixed Integer Program for optimisation.

Davies and Brailsford (2004) described a simulation model for screening complications related to diabetes. In particular, they considered the natural history of eye disease in patients with diabetes and discovered the trade-offs between screening frequency, screening sensitivity and patient compliance.

5. Specialised and preventive healthcare

Increase in longevity, escalating health-care costs and the emergence of new diseases in recent years have forced medical decision-makers to focus strongly on preventive and specialised measures. Many researchers have used stochastic as well as discrete optimisation to analyse the propagation of infectious diseases, evaluate cost-benefit aspects of the screening tests for susceptible groups in the population and transplant as well as donation of vital organs. Terrorism as well as the potential for bio-terrorism have added much more to the complexity and challenges of preventive care.

5.1. Organ donation and transplant

Most research work on organ transplantation focuses on policies for allocating donated organs to the waiting patients (Bruni et al., 2006), liver and especially kidney transplants being the most monitored problems.

Zenios (2002) addressed the mix of direct and indirect exchanges of organs that maximise the expected total discounted quality-adjusted life years of the candidates in the participating pairs. To capture the trade-off, they developed a double-ended queuing model for an exchange system with two types of donor-candidate pairs. They found direct exchanges to be preferable because the candidate receives a living donor organ instead of an inferior cadaverous organ in an indirect exchange. However, the latter involves a shorter waiting time. Zenios (2004) specifically addressed the problem of kidney transplantation and discussed the critical issues in vital organ transplant procedures. They analysed allocation, queuing and simulation models as the core OR methodologies for medical decision making. Su and Zenios (2006) presented the kidney allocation
problem as a sequential stochastic assignment model with the objective of determining an organ allocation policy that maximises the total expected reward. The stochastic model not only gives the patients a greater choice of their positions on the transplant waiting list but also aims to create a more efficient and equitable system. Previously, Su and Zenios (2005) analysed kidney allocation considering the society’s as well as an individual patient’s perspectives. Zenios et al. (2000) also addressed a dynamic kidney allocation problem with three objectives: maximising the quality-adjusted life expectancy of transplant candidates; minimising a linear function of the likelihood of transplantation of the various types of patients; and minimising a quadratic function that quantifies the differences in mean waiting times across patient types. They modelled with linear differential equations and an approximate analysis of the optimal control problem in simulation using a large data set from transplants in the United States. Six demographic groups based on sex, race and age groups were considered.

A software solution for optimising kidney pair donation based on maximum edge-weight matching algorithm is presented in Segev et al. (2005b). The algorithm first constructs a graph in which each node is an incompatible donor–recipient pair and each edge is a potential match between the two connecting nodes. The graph can be constructed by entering the blood type, antibody and antigen information for each pair of nodes. The matching algorithm then determines which pairs are compatible (possible matchings) and draws edges between the corresponding pairs. Finally, it compares all possible combinations of the matchings to determine the best possible solution for the given criteria. The software has already been used by many hospitals; a detailed description can be found in Segev et al. (2005a).

Roth et al. (2004) addressed the problem of having a donor in a noncompatible patient–donor pair offering a kidney to some other patient in the waiting list and Sandiki et al. (2008) addressed privacy concerns of the patients waiting for liver transplant and modelled the problem as a Markov decision process. They discussed conditions under which optimal solutions can exist. They also gave extensive details of the numerical studies conducted.

Liver transplants are analysed by Alagoz et al. (2004). They studied how to optimally time liver transplant to maximise the patient’s total reward, namely quality-adjusted life expectancy. The problem is modelled as a Markov decision process in which the state is described by patient health. Later, Alagoz et al. (2007) suggest that much of the research work on the optimal allocation of organs focused on designing an optimal allocation system for maximising the society’s welfare, such as: mean expected quality-adjusted life years and the average 1-year graft survival probability. They addressed a decision problem faced by the patients with end-stage liver disease: whether to accept or decline an organ of a given quality. The problem is formulated as a Markov decision process in which the state is described by the patient state and organ quality. Stahl et al. (2005) developed an integer-programming framework for determining a configuration of the regions in the United States that maximise liver transplant allocation efficiency and geographic parity.

Shechter et al. (2005) designed a biologically based discrete-event simulation to test changes in allocation policies for liver diseases. Another simulation model which allows a comparison of different policies for allocating donor hearts on pre-transplant outcomes, is presented in van den Hout et al. (2003).

for maximising patient’s benefit. A Game-theory-based approach in which the players are the patient–donor pairs is given in Biro and Cechlarova (2007).

5.2. *Prevention of diseases*

Optimisation problems related to the prevention of diseases concern mostly vaccine selection. The vaccine selection algorithm has been extensively studied as it was first introduced in Jacobson et al. (1999). Sewell and Jacobson (2003) extended the selection algorithm to cover the entire immunisation schedule. They developed an integer-programming model to assess the economic premium that exists in having combination vaccines available. The same authors in Jacobson and Sewell (2002) used the Monte Carlo simulation in conjunction with reverse engineering to determine the probability distributions of four hypothetical combination vaccines. They found that such information is useful for vaccine manufacturers to determine the fraction of the market share they may be able to attract for certain price ranges of the vaccine.


Hall et al. (2008) addressed a vaccine formulary problem for generic childhood immunisation schedules that minimises the weighted sum of the costs for immunising a child and the amount of extra-immunisation. The problem is shown to be NP-hard unless vaccines, schedule parameters, or the diseases are restricted. They formulated an integer-programming model to find the maximum number of vaccines that can be administered without extra-immunisation, which has been a major concern in the United States for childhood immunisation programs. An exact Dynamic Programming algorithm along with a randomised heuristic for the integer-programming model is developed to solve the model. Computational effectiveness and limitations are discussed.

Finally, Wu et al. (2005) formulated the annual vaccine-strains selection problem as a stochastic dynamic program. They used the theory of shape space that maps each vaccine and epidemic strain to a point of a multidimensional space.

In another line, Earnshaw et al. (2007) addressed a resource allocation problem for HIV prevention and developed a linear-programming model for improving on past allocation strategies. Their results suggest that the model can serve as a tool for fund allocation in other prevention activities. Allocation of resources is also addressed in Harris (2006), Zaric and Brandeau (2001) and Zaric and Brandeau (2002). Harris (2006) used a non-linear optimisation model to determine resource allocation in a multiple-site needle exchange programme to achieve the largest possible reduction in new HIV infections at a minimum cost. Optimal allocation of HIV prevention funds among three types of prevention programmes are analysed by Zaric and Brandeau (2001). They also investigated the impact of different allocation methods on health outcomes. Their goal is to maximise the life years gained or HIV infections averted over a specified time horizon. The same authors, in Zaric and Brandeau (2002), developed a dynamic resource allocation model in which limited budget for epidemic control is allocated over multiple time periods. Brandeau et al. (2003) used non-linear optimisation techniques combined with...
epidemic modelling to determine the optimal allocation of limited resources for epidemic control in multiple, non-interacting populations. Brandeau (2004) extended the work to various related problems and gave a comprehensive overview of the exact as well as non-exact OR techniques used for solving the problems.

Markov models are used in Sanders et al. (2005) and Harper and Shahani (2003) to deal with distinct problems. Sanders et al. (2005) used a Markov model of costs, quality of life and survival associated with a HIV-screening programme to analyse benefits and effectiveness of screening for HIV during the highly active antiretroviral therapy. Harper and Shahani (2003) developed a Markov operational model of patient care and used simulation to predict future patient numbers as well as associated healthcare costs at different AIDS clinics in India. Markov models are also used in Siebert et al. (2003) as a decision-analytic tool for economic evaluation and cost-effectiveness of human papilloma virus testing in cervical cancer screening.

Contact tracing or partner notification can help to control infectious disease spread but determining the optimal level of investment required is challenging. Armbruster and Brandeau (2007) presented a simulation-based methodology to evaluate cost the and effectiveness of different levels of contact tracing and extended the model to optimal budget allocation. They discussed how a simulation model can be used as a policy-making tool.

Influenza is the subject of the work by van Genugten et al. (2003). They considered influenza pandemic situations and used scenario analysis on potential effects of intervention options for making the policy decisions on the design and planning of outbreak control management. Chick

Table 1
OR in Healthcare: Research Centres

| Medical Operations Research Laboratory (morLAB), University of Toronto, Canada |
| Center of Excellence in Mathematical Sciences (MITACS), University of British Columbia, Canada |
| Center for O.R. in Medicine and HealthCare at GeorgiaTech, GeorgiaTech, USA |
| Centre for Operational Research and Applied Statistics, University of Salford, UK |
| Health Management Consortium, McGill University, Canada |
| Operations Research Manhattan, Cornell University, USA |
| Health Policy Institute, Boston University, USA |
| Regenstrif Center for Healthcare Engineering, Purdue University, USA |
| Centre for Research in Healthcare Engineering, University of Toronto, Canada |
| Operations Research for Improved Cancer Care, University of British Columbia, Canada |
| Center for Health Care Management, University of British Columbia, Canada |
| Health Systems Research Center, University of Lancaster, UK |

Table 2
OR in Healthcare: working groups within OR associations

| INFORMS Health Applications Section |
| EURO Working Group on OR in Computational Biology, Bioinformatics and Medicine |
| EURO working group on OR applied to health services |
| Health and Social Services, The OR Society of UK |
et al. (2008) studied supply chain issues related to influenza vaccination programmes. They developed a variant of the cost-sharing contract, which provides incentives for the supply chain to achieve global optimisation with an improved supply of vaccines. Nigmatulina and Larson (2009) considered mitigation strategies for global pandemic influenza and used spreadsheet-based models to report on non-pharmaceutical intervention strategies. Infection spread between heterogeneous communities is studied using Monte Carlo simulations, having incorporated into the model a feedback parameter to represent the changes in human behaviour.

Kaplan and Wein (2004) discussed the various aspects of bioterror preparedness and evaluated the consequences of different attack scenarios paired with alternative preparedness and response policies. They used Smallpox in a case study and reviewed the US vaccination policy of 2002 in a decision analysis framework.

Aaby et al. (2006b) used discrete-event simulation, capacity-planning and queuing-system models for the planning of emergency mass dispensing and vaccination clinics. The same

Table 3
OR in Healthcare: Governmental entities

| Department of Health and Human Services, USA |
| National Institute of Health, USA |
| Office of Disease Prevention and Health Promotion, USA |
| Agency for Healthcare Research and Quality, USA |

Table 4
OR in Healthcare: Books

*Operations Research in Health Care: A critical analysis*
  Shuman et al. (1975)

*Operational Research applied to health services*
  Boldy (1981)

*Modeling the AIDS epidemic: Planning, policy, and prediction*
  Kaplan and Brandeau (1994)

*No time to lose: Getting more from HIV prevention*
  Ruiz et al. (2001)

*Quantitative evaluation of HIV: Prevention programs*
  Kaplan and Brookmeyer (2002)

*Operations Research and Health Care. A handbook of methods and applications*
  Brandeau et al. (2004)

*Quantitative Neuroscience: Models, algorithms, diagnostics, and therapeutic applications*
  Pardalos et al. (2004)

*Patient flow: Reducing delay in healthcare delivery*
  Hall (2006)

*Optimization in Medicine*
  Alves et al. (2008)

*Optimization in Medicine and Biology*
  Lim and Lee (2007)

© 2010 The Authors.
International Transactions in Operational Research © 2010 International Federation of Operational Research Societies
methodology is used in Mark et al. (2002) for devising a guidance for the prevention of coronary heart disease.

Diabetes and Hepatitis are also addressed by some authors. Harper et al. (2003) discussed the prevention and treatment of diabetic retinopathy. Earnshaw et al. (2002) designed a linear-programming model to select interventions for preventing complications of Type-2 diabetes and maximizing the quality-adjusted life years subject to budget and equity constraints. The interventions considered are intensive glyceamic control, intensified hypertension control, cholesterol reduction and smoking cessation. Singer and Younossi (2001) used Markov decision analysis models to estimate the cost effectiveness of screening for Hepatitis C in asymptomatic, average-risk adults and Stein et al. (2004) used sensitivity analyses in spreadsheet-based models to estimate the cost utility of testing for Hepatitis C in drug users in the United Kingdom.
6. Additional relevant information

The large number of recent publications cited in this survey paper is certainly indicative of the importance of healthcare to OR professionals. Various OR-related research centres, professional societies and networking groups are known to be actively engaged in healthcare issues. Selected relevant information, as well as information sources that can be important to researchers but typically not found in publications, are presented here in Tables 1–5.

7. Concluding remarks

Papageorgiou (1978) presented one of the earliest surveys of OR applied to problems in healthcare, highlighting how design, operation and management of hospitals can be improved. Since then, OR methodologies have been successfully applied to a variety of optimisation problems arising in healthcare.

During the last decade, OR research communities have become increasingly more interested in tackling numerous challenging problems in healthcare. By addressing the underlying optimisation issues, both deterministic and non-deterministic models have been used to capture the real-world needs and provide a systematic framework for analysis and evaluation. Many new problems have been solved and many previously known “best” solutions as well as solution techniques have been improved. They helped considerably to improve planning, delivery and management of healthcare services.

Simulation and related non-deterministic research in OR cover around 15% of the recent publications in healthcare applications (see Brandeau et al., 2004; Pardalos et al., 2004). Researchers have addressed mostly the optimisation problems associated with hospital admission, hospital services, patient recovery, resource planning, facility utilisation, logistics, supply-chain coordination, vaccination, bioterrorism and emergency response.

The majority of current publications in the deterministic side of OR involve mathematical-programming models and address a wide range of solution techniques including column-generation, shadow pricing, primal-dual, branch-and-cut, branch-and-bound and interior-point methods. Dynamic programming models have been used for surgery and emergency room scheduling, vaccine formulary, as well as capacity planning. Key application areas include: treatment selection and radiotherapy treatment planning (Floudas and Pardalos, 2009; Preciado-Walters et al., 2006; Ferris et al., 2003); emergency-related resource/budget allocation, location planning for medical services, staff and shift scheduling (Oliveira and Bevan, 2006; Shuman et al., 1975); and epidemic modelling (Kahn et al., 1998).

Many problems in healthcare are quite understandably very complicated and cannot be captured using linear relationships; Quadratic Programming, Gradient Projection and Convex Programming have been used for radiotherapy planning (Fox et al., 2006; Ferris et al., 2003). Non-linear Dynamics have been used for designing seizure warning algorithms (Pardalos et al., 2004).

An interesting growing area of current research involves combining different OR methodologies for solving certain complex classes of problems in healthcare. Neural Networks have been used in a linear-programming model for optimising radiotherapy planning (Wu et al., 2000). Simulated
Annealing has also been used as a heuristic optimisation approach for radiotherapy planning (Webb, 1991; Shu et al., 1998).

Another area of research explored techniques that have been successfully used in certain application domains and examined whether such techniques are also suitable for healthcare. Naseer et al. (2009) have made parallels with the defence sector. Practices and strategies successfully used in the manufacturing sector (Lean Manufacturing, Six Sigma, Theory of Constraints) are addressed by Young et al. (2004), Umble and Umble (2006) and Joosten et al. (2009).

Considerably less work appears to have been carried out with potentially promising methodologies such as scenario planning, robust optimisation and reliability modelling. In particular, adaptability and reliability aspects are important for healthcare delivery systems to perform well despite capacity limitations or facility closures (Daskin and Dean, 2004).

Tackling the wide range of optimisation problems in healthcare will certainly require a considerable amount of research work. Identifying the optimisation issues and capturing relevant parameters in mathematical models can be challenging. Finding appropriate solution techniques or formulating new methodologies for solving the models can be tricky. Although OR researchers have already dealt with many new problems and solution methodologies, much still remains to be investigated and solved in this enormous as well as complex problem domain.

References


Blake, J.T., Donald, J., 2002. Mount Sinai Hospital uses integer programming to allocate operating room time. Interfaces 32, 2, 63–73.


Department of Health and Human Services, USA. Available at: http://www.hhs.gov/ (accessed September 9, 2009).


© 2010 The Authors.
International Transactions in Operational Research © 2010 International Federation of Operational Research Societies


Office of Disease Prevention and Health Promotion, USA. Available at: http://odphp.osophs.dhhs.gov/ (accessed September 9, 2009).


© 2010 The Authors.


© 2010 The Authors. International Transactions in Operational Research © 2010 International Federation of Operational Research Societies


