Alma Mater Studiorum Università di Bologna

DEI - Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione
"*Guglielmo Marconi*"

Dottorato di Ricerca in Automatica e Ricerca Operativa
Ciclo XVI

Settore concorsuale di afferenza: 01/A6 - RICERCA OPERATIVA

Settore scientifico disciplinare: MAT/09 - RICERCA OPERATIVA

Operational Research applied to Regional Healthcare System

Paolo Tubertini

Coordinatore
Prof. Daniele Vigo

Relatore
Prof. Andrea Lodi

Final Exam 2014

Contents

Keywords

- *Healthcare*
- *Operations Research*
- *Optimization*
- *Mixed-Integer Programming*
- *Discrete Event Simulation*
- *System Dynamics*
- *Workforce planning and forecasting*
- *Operating Theater planning*
- *Kidney Exchange*
- *Emergency Department planning*

Acknowledgements

First of all, I would like to deeply thank Professor Andrea Lodi that supported and advised me during these three years of joint work. His open minded and brilliant approach to problem solving has been fundamental when the cultural barriers of health managers were too thick and high to be tackled by myself.

I should thank Doctor Roberto Grill for his lasting belief that Operations Research should be applied to Emilia-Romagna Regional Health system.

A special thank goes to Francesca Senese that guided me through the meanders of Health Systems like Dante's Virgil.

I am very glad to have shared this path together with my friends and colleagues Tiziano Parriani and Eduardo Álvarez Miranda. Our daily discussions have been essential to strengthen potential ideas and to dispel emerging doubts.

My gratitude also goes to Professor Ana Viana and Nicolau Santos, it has been a pleasure to work with them on kidney exchange problem, and to Gianmaria Leo that has been fundamental for the success of the collaboration with the Dipartimento di Epidemiologia del Lazio.

A final thank goes to all members of LabOR group and in particular to Professor Alberto Caprara, a magnificent scientist with which I sadly had the opportunity to work just for a short period of time.

Chapter 1

Introduction

Throughout the world, in different ways, the management of health services and the improvement of health care at any level are considered as primary goals for national administrations. Many governments consider health care as one of the most fundamental rights that has to be guaranteed to each citizen. As a consequence the effective and efficient management of national health systems is of paramount importance, not only for its economic sustainability, but also for the huge impact that these activities have on people's life. Such very general considerations are valid not only for the Italian context but also for most of western countries health systems and shall be considered as fundamental in a context of welfare state crisis where traditional funding and management techniques are getting less and less sustainable from an economic point of view. As stated in the 2013 annual dossier of Emilia-Romagna Regional Health System, it is especially in these moments of crisis that health care, and the welfare system in general, can mitigate the social impact and support system recovery. The economic resources and energies devoted to the improvement of health and well-being conditions must then be considered as an investment rather than as an expense. Those resources, like any investment, in order to be fully fruitful should be directed towards the most effective solutions and exploited in the most efficient way. The present and the future of the National and Regional Health Services are conditioned by two fundamental elements: the ability to adequately respond to the needs and expectations of the population, and the ability of the system to be economically sustainable with respect to the allocated resources. It is a fact that the structural factors that influence the trend in health care costs will continue to increase more than the produced wealth. As a consequence, the system will have to deal with an inevitable increase of health costs, even under the hypothesis of a maximum resource utilization efficiency.

The redefinition of Regional and National budgets allocated to health systems is just an element of evaluation in terms of system redesign. It is therefore necessary that the organization of the public health sector develops in parallel with measures that minimize the effects of the economic and financial rationing, controlling costs in the short term, and rationalizing the system in order to improve its economic and financial sustainability while simultaneously preserving the quality of service in the long term. If in the short term the system will not be equipped by the appropriate tools to support a decade that will be dominated by a strong technological innovation and its related increased costs, then it will be difficult to pursue in the long term the continuous improvement required not only because of increased population need of care but also in terms of economic efficiency that can be achieved by technological and organizational innovations.

1.1 Health care systems: characteristics and challenges

There are many and heterogeneous ways in which each national health system is structured in order to deliver services to the population. In any case, required performances and the critical nature of provided services are generally comparable. Health Systems performances affect in a direct way people health and their evaluation is usually linked to indicators such as mortality rates or life expectancy recorded on organizations catchment areas. Absence or lack of prevention, patients acceptance delays and technical or technological inability to deal with certain medical conditions may have a direct impact on population life expectancy, increasing its mortality rate. The ability to monitor system activity and to identify and quickly apply effective innovations is therefore crucial for any health service. The monitoring of population health needs is a difficult task due to the fact that health systems are highly complex and difficult to standardize in a set of procedures. The behavior and the choices of patients and professionals are difficult to translate into repetitive patterns. In other words, even if several care pathway guidelines are defined, patients and professionals discretionary decision-making strongly affect health system services utilization. The uncertainty does not have to be considered as a lack of control over the system but as the impossibility of predicting in an exact way the behavior of the environment in which the health system operates.

It is therefore easy to understand how difficult is to manage such complex organizations trying to minimize waste and maximize efficiency. The identification of the set of inefficient activities and the evaluation of the impact that organizational and technological innovations may have on the whole system dynamics are targets for both primary health care and in-hospital services. The adoption of technological or organizational innovations can also cause initial loss of efficiency and the ability to predict these distortions, and thus somehow to reduce them, is a matter of particular importance in a context in which this can result in damages to the health of the people.

Health care management entails a wide number of planning problems. In Brandeau et al. [2004] the authors suggest that health care challenges should be distinguished among health care planning and health care delivery problems. Health care planning and organizing is strongly related to high-level strategical political issues. As an example, health care regional and national managers have to define which is the set of services that the health system should provide to the population and how the budget has to be allocated among them trying to define future cost patterns (see Bertsimas et al. [2008]). Resource allocation should be split among different components of the health system such as primary care and hospital services. Financial resources should be dedicated to human resources rather than technical equipment and new technologies. Urban versus rural financing should be managed in order to guarantee equal accessibility to services to all citizens. Resources must also be allocated among screening versus treatment programs.

The delivery of health services can be considered as a problem strictly related to system operations at a strategical, tactical and operational level. At a strategical level this means to define the design of services in terms of the set of care activities that will be provided in each health care facility as well as to design the health care supply chain in terms of number and characteristics of hospitals, outpatient clinics, laboratories and primary care structures. Other strategic planning problems include demand and capacity planning and forecasting. Health operations issues arise also at a tactical and operational level. Workforce sizing and rostering, operating theater planning and scheduling, bed management are just a subset of problems that should be managed with hospital facilities. Health care managers should also tackle inventory problems related to drugs, blood and general medical supplies availability that can strongly affect the daily provision of care and, in some cases, the ability of the care facility to properly treat the patient. Hospitalization services are just a component of healthcare delivery operations. Primary care and public health services such as screening and home care delivery are elements of increasing importance in terms of volume of activities and the health managers will face an increasing pressure to effectively organize and monitor those programs.

1.2 Health care systems: a brief overview of Operations Research applications

To provide the best health care given the amount of limited resources available, policy makers need effective methods for planning, prioritization, and decision making, as well as for management and improvement of health care systems. In recent years Operations Research (OR) has been increasingly applied to support healthcare strategic, tactical and operational planning problems by exploiting both optimization and numerical simulation techniques. The range of applications involving OR is so wide that it is difficult to define a complete overview of the covered problems. Nevertheless, several attempts have been made to summarize the improvements and the success stories achieved till now. In Brandeau et al. [2004] the authors, after a first definition of challenges in health care, classify Operations Research applications as: (i) health care operations management, (ii) public policy and economic analysis, and (iii) clinical applications.

Among the cited problems it is possible to identify the subset of the most studied ones. Health related applications can be split over hospital versus primary and home health care. Hospital management includes nurse rostering problems reviewed in Burke et al. [2004], appointment scheduling problems in Gupta and Denton [2008] and general capacity planning problems analyzed in Green [2004]. Operating theater management, due to the complexity and the economic value of the involved resources, is another important problem for which several applications have been developed and an in depth review of the state of the art in this field can be found in Cardoen et al. [2010].

It is clear that primary care and home health care are planning sectors of rising importance due to the general perception that inpatient activities should be more and more reduced. An overview of applications related to primary care has been conducted in Balasubramanian et al. [2013], while home health care problems are studied in Lanzarone et al. [2012] and in Mankowska et al. [2014]. Health facility location is then a problem of paramount importance in terms of impact on population health. A general review of the applications in this field can be found in Daskin and Dean [2004]. In Brotcorne et al. [2003] an in-depth analysis of ambulance location and relocation models is reported.

So far the cited applications analyze health care problems from an optimization perspective, it is important to point out that a wide branch of studies focus instead on numerical simulation techniques. In Seila and Brailsford [2009] and Mielczarek and Uzialko-Mydlikowska [2012] a wide analysis of simulation-driven applications is reported classifying the most relevant contributions for strategical, tactical and operational planning. Most of the problems are related to capacity planning analysis, waiting time and length of stay reduction and care pathway analysis (see Segev et al. [2012]).

1.3 Thesis aim and objectives

Emilia-Romagna Regional Health system constantly faces most of the challenges that were previously cited. As an example the increase of the resident population (+0.6%) and its aging (people aged over 65 years exceeded one million: 1,004,450 units equal to the 22.5% of the population) reported in the 2013 annual dossier of Emilia-Romagna Regional Health System put an high pressure on the importance of system redesign. The aging of the population is even a more dramatic factor considering that the regional phenomenon is higher than the national one (20.3% at December 31, 2010), and has been continuously rising for more than two decades. One of the consequences is that the number of home cared patients nearly doubled from 2001 (55,000). These phenomena lead to an increase in terms of overall system expense that raised from 7,627,534 euros in 2007 up to 8,676,661 euros in 2012. These elements, together with the enduring economic crisis and the subsequent need to reduce the budget dedicated to public facilities through political reforms (spending review), push regional health managers to look for new methods and techniques in order to reduce system inefficiencies by, in turn, maintaining the quality of care provided to the population.

In this context "Agenzia Sanataria e Sociale Regionale dell'Emilia-Romagna", a regional center for organizational and technological innovation in health care, in 2010 decided to evaluate how Operations Research techniques could be exploited as decision support system tools for planning problems arising in the public regional health sector. The project aims to assess how numerical simulation and combinatorial optimization models, customized to the healthcare environment, can guarantee planning effectiveness through the integration with the set of techniques already used to support decision processes.

The decision support system tools that we will present in the following chapters should not be considered as an attempt to impose quantitative outputs as out-of-the-box answers to solve health planning problems. Qualitative evaluations of expert health managers and clinicians are unavoidable and fundamental factors during policy making processes in health systems. The idea of the thesis is that informed and quantitative-driven decision-making processes can lead to an in-depth analysis of system dynamics and criticality and, as a consequence, to more efficient and effective final solutions. The collaboration lead to a three-year analysis to deal with some of the main challenges and organizational problems that can be effectively studied and solved applying operations research techniques. Three case studies have been developed for the Regional Health Sector.

1.3.1 Emilia-Romagna case studies

Chapter 2 In Chapter 2 a long term strategic planning model related to workforce planning is presented. The problem is the forecast and fund allocation of medical specialty positions. The Regional healthcare authority each year has to define how many graduated physicians will be trained in medical specialization schools. In order to do this it has to negotiate the number of scholarships that will be funded by national government and to define how many extra ones will be financed by the Regional health care system itself. Our working hypothesis relies on the assumption that future Emilia-Romagna Region (ERR)-Human Resources in Health (HRH) requirements and the regional grants allocation have to be defined in correlation with the current shortages/surpluses and demographic and service utilization changes. The approach we decided to follow is a *demand-based* one and it relies on the assumption that health care services are driven by population demand of care. We developed a System Dynamics model that represents the human resources of regional medical specialists. The model calculates the flows of medical specialists trained in the ERR since 1999 to employment positions in the regional healthcare labor market. Each component of the labor market as well as the training is defined as a stock element of a System Dynamics model and flow components are used to characterize retirement and dropout rates as well as hiring and graduating ones. Then, the model ages the human resources population by projecting to 2030 its topology. Once the requirement and supply projections have been defined two kind of gaps have been analyzed in order to identify future shortages and surpluses, namely the occupational and the training ones. The former defines employment shortages due to dropout and retirement issues compared to requirement projections, the latter defines shortages and surpluses between requirement projections and medical specialists available on the employment market. In order to optimally manage training policies an Mixed Integer Programming (MIP) model has been implemented. The MIP model allocates on each year a number of grants to medical specialties that show the higher shortages. The model, designed in order to support a strategic decision process, minimizes the total shortage over a 20-years time horizon. The case study has been submitted to the EURO Summer Institute (ESI) on OR applied to health in a modern world (ESI XXXI), discussed during the "WORLD SUMMIT ON BIG DATA AND ORGANIZATION DESIGN" held in Paris in May 2013, presented at "XVIII Convegno Nazionale Associazione Italiana di Economia Sanitaria" held in Trento in November 2013. Finally, the Regional working group has been invited to join the European Joint Action for workforce planning and forecasting as a component of Work Package 5 on methodology definition and pilot studies.

Chapter 3 In Chapter 3 a planning problem related to care pathway evaluation and capacity planning is presented. The aim of this work is to study how two different Discrete Event Simulation (DES) software packages can be effectively applied to support tactical and operative decision-making processes. In order to show the potential of DES modeling, a breast-screening pathway has been studied to demonstrate how numerical simulation techniques can be used to evaluate lead-time performance under different capacity settings. The case study was focused on 45-49 and 70-74 women age bands inclusion that took place in 2010 and has shown how simulation can help in understanding how many resources the screening program needs in order to face an increased demand of services. The screening program is analyzed under two perspectives (local and regional) implementing a simulation model with two software packages in order to show how different conditions in terms of available datasets and modeling detail can help stakeholders that operate at different decision levels. The case study has been accepted for publication in the proceedings of the European Conference on Modelling and Simulation, ECMS 2014.

Chapter 4 In Chapter 4 an operational planning model related to Operating Theater management is presented. In collaboration with the orthopedic department of a local hospital we analyze, by means of a MIP model, how to effectively plan the assignment of operating room time slots to surgeons and the corresponding identification of which patients have to be admitted to peri-operative activities. The orthopedic department has a set of operating rooms that have to be daily assigned to surgeons. Each assignment identifies, for every operating room and for every session (morning or afternoon), the surgeon, the class of surgical cases (prosthetic or not) and the related subset of patients admitted. Our assignment is driven both by patient Key Performance Indicators, such as clinical priority, length of stay in waiting list and pre-defined deadlines and by surgeon peculiarities, such as minimum and maximum desired number of weekly sessions assigned and length of the associated waiting list. Two models are presented in the chapter: a general one, based on Emilia-Romagna regional guidelines, which includes all possible resources that restrict operating room planning (Intensive Care Unit, ordinary and day hospital stay beds, surgical teams and pre-operative assessment activities), and a tailored one that considers only the resources that characterize the orthopedic department. The work has been presented at "INFORMS annual meeting" held in Phoenix in October 2012.

1.3.2 National and International case studies

In addition to Regional-driven case studies two external collaborations have been activated.

Chapter 5 In Chapter 5 a simulation-optimization model for the policy evaluation of the so-called Kidney Exchange Problem is presented. In recent years several countries have set up kidney exchange policies between living pairs in parallel with deceased donors transplantation. These exchanges may occur when a patient that needs a kidney has an incompatible donor, so he is willing to exchange it with another pair, or more than one, in order to perform a transplant. This problem can be modeled and solved via integer programming models. Because the problem has a dynamic arrival of pairs (patient, donor) within the waiting list, it is suitable to study different policies via a simulation-optimization approach. In collaboration with INESC TEC, an R&D center in Porto, Portugal, we implemented a simulation-optimization tool that gives the possibility, to clinicians and policy makers, to test different configurations regarding matching frequency, matching characteristics and pool characteristics. Discrete Event Simulation has been applied to model exchange pool dynamics. We defined an input infrastructure that easily allows the modification of pool characteristics so as to build a tool that can be used to test different policies for different catchment areas. We collected a set of data in order to test the proposed approach under different configurations in terms of pool characteristics and of matching policy. Finally, we compared the simulation results quantifying the increase of number of transplants with the inclusion of altruistic donors and compatible pairs as well as with the increase of the exchange length in the number of pairs involved.

Chapter 6 In Chapter 6 an allocation model for the patient flow management of patients to Regional Emergency Departments is presented. The assignment of service requests to Emergency Rooms is of paramount importance both from a life-threatening and an economical viewpoints. In the process of a more general project that aims at defining optimal allocation policies of patients to regional hospital network facilities, the Department of Epidemiology of the Regional Health Service of Lazio, Italy was interested in obtaining a completely offline picture of the effect of an optimal assignment of requests to Emergency Rooms so as to be able to evaluate both the state of the art and future reorganization ideas. We have implemented and tested with real-world data of all service requests of 2012 a MIP model that computes such an optimal request allocation by minimizing travel and waiting times and penalize workload unbalance among emergency rooms in the region. Within the development process we have studied special cases and relaxations of the complete model showing interesting mathematical properties that are, in turn, useful from a practical viewpoint.

Bibliography

Hari Balasubramanian, Ana Muriel, Asli Ozen, Liang Wang, Xiaoling Gao, and Jan Hippchen. Capacity allocation and flexibility in primary care. In Brian T. Denton, editor, *Handbook of Healthcare Operations Management*, volume 184 of *International Series in Operations Research & Management Science*, pages 205–228. Springer New York, 2013. ISBN 978-1-4614-5884-5. doi: 10.1007/978-1-4614-5885-2_8. URL http://dx.doi.org/10.1007/978-1-4614-5885-2_8.

Dimitris Bertsimas, Margrét V. Bjarnadóttir, Michael A. Kane, J. Christian Kryder, Rudra Pandey, Santosh Vempala, and Grant Wang. Algorithmic prediction of healthcare costs. *Operations Research*, 56(6):1382–1392, 2008.

Margaret L. Brandeau, François Sainfort, and WilliamP. Pierskalla. Health care delivery: Current problems and future challenges. In MargaretL. Brandeau, François Sainfort, and WilliamP. Pierskalla, editors, *Operations Research and Health Care*, volume 70 of *International Series in Operations Research & Management Science*, pages 1–14. Springer US, 2004. ISBN 978-1-4020-7629-9. doi: 10.1007/1-4020-8066-2_1. URL http://dx.doi.org/10.1007/1-4020-8066-2_1.

Luce Brotcorne, Gilbert Laporte, and Frédéric Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147(3):451–463, 2003. URL http://dblp.uni-trier.de/db/journals/eor/eor147.html#BrotcorneLS03.

Edmund K. Burke, Patrick De Causmaecker, GreetVanden Berghe, and Hendrik Van Landeghem. The state of the art of nurse rostering. *Journal of Scheduling*, 7(6):441–499, 2004. ISSN 1094-6136. doi: 10.1023/B:JOSH.0000046076.75950.0b. URL http://dx.doi.org/10.1023/B%3AJOSH.0000046076.75950.0b.

Brecht Cardoen, Erik Demeulemeester, and Jeroen Beliën. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921 – 932, 2010. ISSN 0377-2217. doi: http://dx.doi.org/10.1016/j.ejor.2009.04.011. URL http://www.sciencedirect.com/science/article/pii/S0377221709002616.

Mark S. Daskin and Latoya K. Dean. Location of health care facilities. In Margaret L. Brandeau, François Sainfort, and WilliamP. Pierskalla, editors, *Operations Research* *and Health Care*, volume 70 of *International Series in Operations Research & Management Science*, pages 43–76. Springer US, 2004. ISBN 978-1-4020-7629-9. doi: 10.1007/1-4020-8066-2_3. URL http://dx.doi.org/10.1007/1-4020-8066-2_3.

Linda V. Green. Capacity planning and management in hospitals. In MargaretL. Brandeau, François Sainfort, and WilliamP. Pierskalla, editors, *Operations Research and Health Care*, volume 70 of *International Series in Operations Research & Management Science*, pages 15–41. Springer US, 2004. ISBN 978-1-4020-7629-9. doi: 10.1007/1-4020-8066-2_2. URL http://dx.doi.org/10.1007/1-4020-8066-2_2.

Diwakar Gupta and Brian Denton. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9):800–819, 2008. doi: 10.1080/07408170802165880. URL http://www.tandfonline.com/doi/abs/10.1080/07408170802165880.

E. Lanzarone, A. Matta, and E. Sahin. Operations management applied to home care services: The problem of assigning human resources to patients. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42(6):1346–1363, Nov 2012. ISSN 1083-4427. doi: 10.1109/TSMCA.2012.2210207.

DorotaSlawa Mankowska, Frank Meisel, and Christian Bierwirth. The home health care routing and scheduling problem with interdependent services. *Health Care Management Science*, 17(1):15–30, 2014. ISSN 1386-9620. doi: 10.1007/s10729-013-9243-1. URL http://dx.doi.org/10.1007/s10729-013-9243-1.

Bożena Mielczarek and Justyna Uzialko-Mydlikowska. Application of computer simulation modeling in the health care sector: A survey. *Simulation*, 88(2):197–216, February 2012. ISSN 0037-5497. doi: 10.1177/0037549710387802. URL http://dx.doi.org/10.1177/0037549710387802.

Danny Segev, Retsef Levi, Peter Dunn, and Warren Sandberg. Modeling the impact of changing patient transportation systems on peri-operative process performance in a large hospital: insights from a computer simulation study. *Health Care Management Science*, 15(2):155–169, 2012. URL http://EconPapers.repec.org/RePEc:kap:hcarem:v:15:y:2012:i:2:p:155-169.

Andrew F. Seila and Sally Brailsford. Opportunities and challenges in health care simulation. In Christos Alexopoulos, David Goldsman, and James R. Wilson, editors, *Advancing the Frontiers of Simulation*, volume 133 of *International Series in Operations Research & Management Science*, pages 195–229. Springer US, 2009. ISBN 978-1-4419-0816-2. doi: 10.1007/b110059_10. URL http://dx.doi.org/10.1007/b110059_10.

Chapter 2

Forecast and fund allocation of medical specialty positions: Emilia-Romagna Region case study [1]

2.1 Introduction

The labor market and the demand for medical doctors are extremely adaptable to new technologies, societal demand and organizational models, therefore, planning human resources for health (HRH) is a logistical task of great complexity (Barber and Lopez-Valcarcel [2010]). In Italy the perceived shortages of medical specialists led to increase medical schools intake in 2010 (+29%), although, imbalances and appropriateness of current medical supply and distribution to health care needs have not been assessed. Planning HRH requires to model the supply of specialists but, most noticeably, to forecast the changing needs of the population and the emerging care pathways. Italian Regions aim at correcting perceived imbalances through the funding of supplementary grants for Medical Specialists Schools, namely Emilia-Romagna's funding covered almost 20% of total grants in 2011.

The need for HRH long-term planning and the availability of different methodological approaches (supply, requirements, needs-based) have been stressed by World Health Organization in WHO [2010] and by European Union in CE2 [2008]. HRH availability is crucial to pursuit high performances since HRH imbalances are associated with financial consequences (e.g. increased cost of labor, supply induced demand), poor responsiveness to patient expectations due to burn-out and waiting times, and

[1] This chapter is based on Technical Report OR 14-5 (see Lodi et al. [2014]) submitted to ESI XXXI EURO Summer Institute on Operational Research applied to Health in a Modern World

poor services quality and safety in case of under-staffing. Models for doctors supply and requirement forecasting have been built in several countries such as USA (PSR [2005]), France (Doan [2004] and Coste and Doan [2003]), Spain (Barber and Lopez-Valcarcel [2010]), Belgium (Roberfroid et al. [2009]) and Australia (Joyce and McNeil [2006], Schofield [2007]). Our working hypothesis relies on the assumption that future Emilia-Romagna Region HRH requirements and prioritization of regional grants have to be defined in correlation with current shortages/surpluses and demographic and service utilization changes. Therefore, the development of a demand-based approach to forecast HRH will prevent future imbalances in the regional health labor market, considering different demand scenarios that can be driven both by population evolution and by regional and national guidelines.

The chapter is organized as follows. In Section 2.2 we report a literature review of current forecasting models. In Section 2.3 we present an HRH forecasting model for Emilia-Romagna Region. In subsection 2.3.2 we discuss the characteristics of the simulation component and in subsection 2.3.4 we propose an Integer Programming model for the allocation of medical funded grants to specialties.

2.2 Literature review

The interest in the development of quantitative models to estimate the needs of human resources in health care is evident in many countries and international organizations including the World Health Organization (WHO), the Organization for Economic Co-operation and Development (OECD) and the European Union (EU). The need to improve Human Resources in Health planning is motivated by several factors, such as the regulation of the number of grants funded for health area, the imbalance evaluation (excess and defect) between needed and employed HRH both in the public and private sectors and the perception, at least in OECD countries, that the current workforce is made up mainly of baby-boomers who will soon leave the service.

The idea that the determination of volumes of health professionals training had to be mainly linked with the need of the public sector is a conceptual simplification that, however, facilitates the analysis of the, current and future, alignment between the supply and the demand of health personnel. Supply and demand dynamics for human resources in health are complex to analyze, as a consequence a forecasting approach through a correct model of future staffing needs, is suitable in order to plan in a timely manner the development of public facilities and of university education.

An agreement on measures of HRH needs satisfaction, or an estimate of the "right number" of professionals is the basis for the definition of correct forecasting models.

It is fundamental to avoid that HRH policies leads to inadequate satisfaction of population needs, due to poor quality of services (staff shortages), inappropriate services application and rising costs (excess staff) see Duckett [2000].

While considering these crucial elements the literature review will be an overview of the dominant HRH forecasting models highlighting their methodological assumptions and the main variables and constraints that had been taken into consideration. The aim of this literature review is to provide a methodological guidance on options for achieving HRH forecast, by taking into account and analyzing the opportunity to integrate the approaches and to bring the financial forecast as part of the cycle of strategic planning of human resources.

2.2.1 Supply models

Irrespective of the interest motivating the HRH analysis (occupational policies, training policies, geographical distribution of services), all models that we reviewed firstly take into account the supply of professionals. Supply models have been developed in the USA, Belgium (Roberfroid et al. [2009]), Australia (Joyce and McNeil [2006], Schofield [2007]), Canada and France (Doan [2004] and Coste and Doan [2003]). Supply forecasts take into account the main input and output variables that determine the availability of personnel as the number of active professionals, the newly licensed physicians and the volume of drop outs per year due to retirement, death, disability, migration, or unknown causes. A particular emphasis is given to professional demography whereas age and gender are considered the basic variables to predict supply future behavior, since working productivity or full time equivalent (FTE) can vary a lot due to sex or age characteristics. As an example in PSR [2005] it appears that the peak of productivity for both men and women is in 50-55 age band followed by a gradual decrease until retirement. In the same study it is said that working hours can be 30-36 % higher for men if compared to women. A similar analysis was conducted in Roberfroid et al. [2007] for Belgian physicians. Given a complete description of the baseline stock, inflow and outflow forecasts define supply model gaps, between remaining professional supply and initial staffing levels, that must be filled with additional training grants. As an example the Physician supply model in PSR [2005] analyzes the national health workforce and projects its evolution by defining how many professionals will be needed in the future in order to maintain a pre-defined ratio between resident population and staffing level. In this case the model assumes that the number of graduates and the productivity of physicians will remain constant over the years and estimates the relationship between supply and current population applying FTE information and assuming that the number of required physicians will expand or contract with respect to population demographic trends.

Supply models are usually criticized because they assume as adequate the actual system configuration, defined as level of activity, skill mix and physician-to-population ratio. Supply model scenarios can be defined as predictive because they are based on observed trends and their forecasts generally overestimate future staffing needs. This is because both population and professional aging phenomena are not balanced with technological or care pathway innovations, which are proper of industrialized countries. Despite being the most severely criticized, supply models have a good internal consistency and can support long-term predictions on possible future shortfalls of professional stock. Forecasts to 2050 in Doan [2004] on the flow of incoming and outgoing doctors in France is an example of how this kind of models, despite their limits, can provide useful insights for nationwide training policies.

2.2.2 Physician requirement or demand models

Physician supply analysis can be considered as the side of the problem that analyzes only the evolution of professionals that were active at the beginning of the analysis. It is clear that this is just one of the multiple dimensions of the HRH planning problem. Models that takes into account both the supply and demand sides are known as physician requirement models or demand models. Most of the demand models take into account changes in care pathways such as the utilization of primary, outpatient or inpatient services due to organizational choices and epidemiological factors. In order to project the need of new medical doctors it is then fundamental to define what are the elements, or the drivers that influences their requirement over time. According to Greenberg and Cultice [1997] model, the segmentation of the population, the knowledge of care pathways and the consumption of services by age groups are fundamental. Roberfroid et al. [2009] suggest that requirement models may use two different approaches in order to define future demand of services. Current service utilization ratios can be used as a proxy of real demand as an alternative a reference ratio between providers and target population can be defined before applying demographic trends. The predictive model proposed in Nooney and Lacey [2007] for nursing staff as well as the one proposed in O'Brien-Pallas et al. [2001] for human resources in health suggest that future needs of staff should be estimated taking into account health system infrastructures. The availability of hospitals, clinics or nursing homes can be considered as active constraints to effective service delivery, therefore employment requirements cannot be separated from an evaluation of structural constraints. Taking care of the population in an adequate way does not simply mean to adjust the number of professionals to population increasing needs, but to make sure that human resources properly fit with Health System facilities. The limitations of the demand models arise from their implicit assumptions:

- Current delivery of services respond adequately to the demand;

- Demand of services responds to population health need;
- Age and sex specific segments of the population won't change their consumption behavior;
- The population will follow the demographic changes predicted by the model.

Most radical criticism of this approach come from Birch et al. [2003] and Tomblin Murphy et al. [2009] who argue that age is not a good proxy for the consumption of services and for the demand of services. These authors discourage the idea to plan services according to population dynamics because a retrospective analysis of health status shows that different generations (cohorts), at the same age, require different type and amount of services.

2.2.3 Needs-based models

Weaknesses of supply and requirement models have highlighted the need to guide the planning of human resources through the detection of population real needs. Models that aim at defining the need for services are recent and have been created to fill the gaps of supply and requirement models that tend to systematically overestimate the staffing needs leading to potentially unsustainable systems from the economic point of view. Need-based approach assumes that current staffing levels are not necessarily coherent with population needs and may be adapted in order to improve system efficiency and responsiveness. Birch et al. [2007], WHO report of 2001 and CHSRF report of 2007 suggest that HRH staffing levels should be based on population health needs that can be periodically inspected and defined through sample surveys and epidemiological observations. Needs-based models, also called epidemiological approaches, are not based on demand-model drivers such as service utilization or population ratio but they rely on medical experts opinions and generally assume that:

- Health needs must and can be satisfied through a proper resource resizing;
- An a priori definition of need of care can be convenient since it evolves less rapidly than population demand of services;
- Demand appropriateness is defined by specialists and not by users;
- Needs can be derived by risk assessment analysis and evidence-based morbidity data.

Central to needs-based modeling of HRH is the resizing of human resources through basic and continuous training ensuring an efficient deployment of resources. The cornerstone of Birch et al. [2007] approach is the estimation of the health needs integrated in a continuous planning cycle.

The main limitation of need-based approaches is the lack of population epidemiological data. Persaud and Narine [1999] underline that the lack of proper data is usually overcome using DRG utilization or insurance refunds data, in order to approximate population health needs. Not surprisingly, studies that promote needs-based approaches can rely on a consistent volume of epidemiological historical data.

2.2.4 Final considerations

The reviewed studies confirm that, although the estimate of needs is the ultimate goal, an accurate and integrated analysis of at least three factors must to be taken into account: training, employment and the demand for health care personnel.

The debate on predictive models in recent years underlined the presence of factors that are difficult to quantify such as the impact that technological innovations will have on work organization, the freedom of choice of professionals regarding retirement decisions and the role played by financial incentives of both public and private health organizations. Another phenomenon of increasing importance is the highly specialized curriculum of professionals able to produce specific outcomes. Health care is a labor-intensive sector, since it provides services that are highly personalized and characterized, in some cases, by an important relational component. Higher education, training and productivity of health professionals are then the heart of healthcare industry. It is, therefore, difficult to find clear instructions on how to conduct a reliable forecast of health care personnel demand. It is more common to identify lists of warnings or issues that have to be taken into account, among which it is worth remembering some practical policy recommendations (see WHO [2010]), such as:

- Utilization of updated and reliable data;
- Definition of a clear and comprehensible model to facilitate the involvement of different actors within the decision process;
- Model verification on the most critical areas (e.g. professional mobility and the pressure of aging population).

Analyzed models can be roughly distinguished into descriptive and forecasting studies.

Most descriptive studies are limited to the quantification of the current stock of professionals distinguished by specialty, age, gender and assess whether the number of professionals in training will be enough to balance the output of the staff due to retirement and other dynamics of turnover or not.

HRH forecasting models provide, instead, a dynamic cycle of analysis, including an evaluation of the current system setting and a forecast of its future evolution (see Gavel [2004]).

2.3 Emilia-Romagna case study

2.3.1 Objectives and methodology

Each year Emilia-Romagna health authority's managers have to negotiate the number of medical specialization grants that will be financed by the national government and to define the amount of additional grants that will be funded by the regional budget. The final goal of this study is to provide a requirement indicator for each medical specialty within Emilia-Romagna territory over a twenty year planning horizon. Based on these indicators a model will compute potentially effective assignments of medical specialties grants considering future population utilization of public health services, programmed needs expressed by Local Health Trusts and the role of commissioned private organizations. In order to meet the objectives, the decision-making model that will support regional planners has to be separated in two main components. A simulation model, which describes the supply and demand behavior over time and an optimization model that evaluates the imbalances that emerged from the simulation model and suggests an optimal funding strategy in order to reduce the gap between HRH availability and requirements.

2.3.2 A system dynamics simulation model

The most appropriate simulation technique to represent the dynamic behavior of multiple interacting populations (interns, physicians, enrolled patient population, etc.) is System Dynamics (SD). System Dynamics is a modeling methodology and a simulation technique to monitor, understand and evaluate organizational problems. Originally developed in the fifties to help business managers to improve their understanding of industrial processes, SD is currently used in both the public and private sectors to support strategic decision-making. The objective of SD models is to understand how the structure of a system determines its behavior. This understanding normally produces a framework for determining what actions can improve the system performances or fix its problems. In a system dynamics model, the simulations are essentially time-step simulations. SD models complex systems using a stock and flow representation. A stock is a container that varies over time due to inflows and outflows. An inflow is a component that is used to increase the initial level of a stock component, an outflow instead is a component used to reduce the level of the upstream stock. Inflows and outflows can be then considered as the elements that influence the stock levels and can be characterized by complex rules that can take into account various elements (stocks, flows, variables, etc.) of the systems, simultaneously influencing their behavior. Figure 2.4 shows the implemented SD model for Emilia-Romagna. In the following subsections we describe its supply and demand components.

2.3.2.1 The supply representation

A supply representation has to take into account all the steps that represent physicians work-life cycle starting from specialization school, up to retirement or dropout because of death or move. The work-life begins with admission to specialization schools that can last up to six years; once graduated a trained specialist becomes available for hiring in the public or private regional sectors. Each employment sector competes with the others to hire trained specialist. Working life can finish because of two main factors that are retirement or move to another region.

It is easy to understand that regional supply cannot be considered as a monolithic element because several classes of health providers can be distinguished by their role or their organizational setting. We defined the working population of the study as the set of medical specialists working in Emilia-Romagna region. Since the objective of the study is to provide a guideline on future training needs of medical specialists, we decided to study employment dynamics at the regional level, including in and out flows from the private sector, accredited profession and underemployment conditions. It is possible to identify six main components of the supply side:

- Public sector (Servizio Sanitario Regionale SSR);
- Private hospital sector (AIOP);
- Private Outpatient Sector;
- Public self-employed ambulatory specialists (Sumai);
- General Practitioners;
- District Pediatricians “Pediatri di Libera scelta”.

A proper modeling of HRH forecasting in Emilia Romagna has to deal with a screening of available data sets. Data collection is complicated because, even if multiple data sources are available and accessible, stored information are usually not homogeneous. Historically data sources that collected information on HRH were implemented in order to answer to specific organizational requirements that did not take into account the possibility of supporting HRH forecasting and planning. Thus, at a regional level it does not exist a unique database that contains individual data on employed medical doctors. It means that as a first step we had to identify and integrate multiple data sources in order to define a unique and coherent representation of regional professional demography. For the purposes of this work it would be desirable to associate to each active physician its specialization within the Ministry of Education, University and Research (MIUR) classified ones. For medical doctors it means to identify their specialization title, taking into account that the definition of new specialization schools can lead to identify professionals who currently hold similar jobs, despite having specializations with different names, depending on the year in which they earned the title. SSR employement data come mainly from three sources with restricted access: Economy and Finance Ministry General Accounting Department "Conto annuale", "Elenco nominativo ruoli regionali" and "Rilevazione personale sanitario" of General Directorate of Health and Social Policies, human resources development service, Emilia-Romagna Region (RER-DGSPS).

"Conto annuale" collects annual census surveys compulsory for public institutions as a part of their final balance. As far as health care sector concerns it is possible to identify clinicians, managers without a medical background and non-managerial staff. For each local health authority it is then possible to define professionals managerial position and a detailed information on their qualification and economic compensation. For each qualification (legal-economic position covered by the employee on the basis of national collective bargaining agreement) it is possible to obtain information on age, part-time, career progression and turn-over for the period 2001-2010. The data contained in "Conto annuale" are useful to consider the length of service and to study age related retirement rate probabilities. "Elenco nominativo ruoli regionali" database is mainly used to establishment the components of examination boards for healthcare professionals. It collects information by name and age of the permanent employees of the Health Care and their position. "Rilevazione personale sanitario" is part of a more complex database that is used to monitor personnel and payroll at a regional level. In this database staff information is classified at an aggregated level according to role, responsibilities and specialty with information about the type of employment (temporary / permanent) and the employment area (hospital, support, district, prevention) including University contracts. In addition, as regulated by law since the foundation of National Health Service (SSN), healthcare agencies cooperate with professionals that work under agreement contracts. Today it is possible to classify agreement-contracted professionals as general practitioners, continuity of care doctors, district emergency workforce, pediatricians and health personnel working in university hospitals. Agreement contract data are collected on specific databases and reported in some tables of "Conto Annuale".

For general practitioners (GPs), internal DGSPS-RER "Cedolini" database has been analyzed. In this database GPs, pediatricians (PLS) and their connected substitutes and trainees are recorded. For each medical doctor age, sex and number of related patients are recorded. Sumai and AIOP data are not recorded at a regional level so we collected those pieces of information through ad-hoc surveys.

Once available data sets were analyzed, the supply model was implemented for all 62 specialties active in ER region considering that training and working life can be included in 25-70 years of age and split by gender. We then initialized training and working stock levels with 2011 data. Figure 2.1 shows age and gender distribution

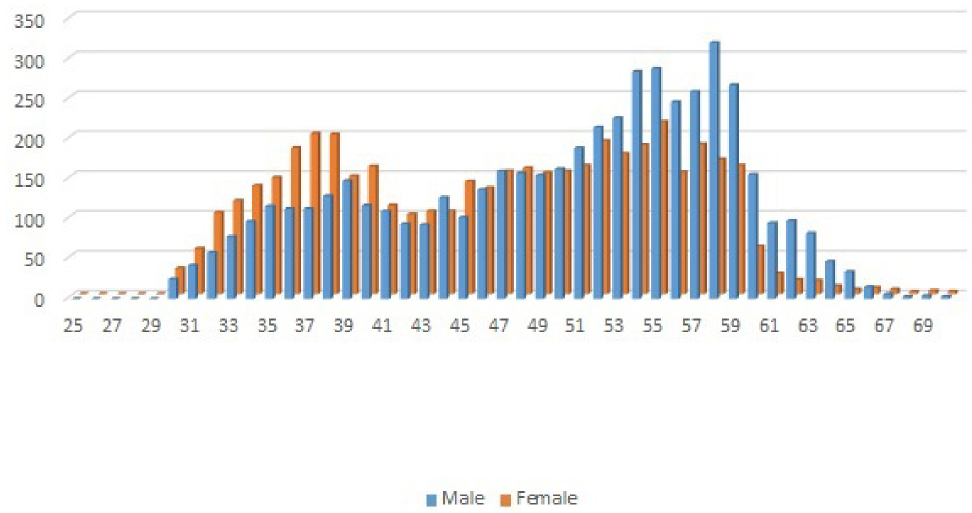


FIGURE 2.1: Specialists working in the Public sector age distribution in 2011

of physicians working in the public sector in 2011. As stated before a fundamental element in SD is the definition of variables and flows that allow to represent a system in a dynamic way. Through the analysis of the supply data bases it has been possible to quantify not only the amount of physicians that were active in 2011 but also the system behavior in the period 2001-2010. Retirement rates have been defined for Public, Private and Accredited sectors, pointing out that physicians behave in a different way depending on their employment sector. Public to Private sectors flows in 45-65 age band have been then modelled. These flows are mainly due to better economic conditions that highly skilled physicians can get shifting to Private Hospitals. Regional healthcare systems can't be considered as self-sufficient sectors. Migration flows due to employment reasons are elements that can't be overlooked. Available data sources allow an outflow analysis for the public sector. We defined migration outflow rates starting from the number of physicians that during 2001-2010 left the public system before retirement age and were not present in the next years in private sector available databases. As a last flow element we modeled the training rate, that is to say, the number of specialized physicians that become available for employment in one of the regional supply sectors. Training flows can be considered as time-delayed flows that come form training school inflows, defined by funded grants in previous years.

2.3.2.2 The demand model

In the literature review we noted that the projection of future workforce demand is a task of great complexity due to the ambiguous definition of demand drivers. However demand should be defined taking into account:

- Population health needs;
- Expected service utilization changes due to organizational and technological improvements;
- Regional employment trends considering both the private and public sectors.

The three scenarios are now discussed in detail.

2.3.2.3 Physician to population ratio scenario

Demographic change is an element of paramount importance for health systems in terms of future service utilization. Migration flows as well as aging population phenomena can strongly affect future service requirements and, consequently, staffing level needs. Based on the analysis of regional population historical trends we must say that, among Italian regions, Emilia-Romagna is the one that had the most rapid and intense transformation of its age structure. As an example elderly population overtook the young one nearly a decade earlier if compared to Italian overall population. In a similar way, migration flows are slowing down the overall aging of the population because foreigners living in Emilia-Romagna are on average younger than the historical population. The Statistical and Geographic Information Service of Emilia-Romagna published in 2011 a set of demographic projections to 203. These forecasts are based on observed population trends recorded from 2001 census up to 2010 annual surveys. Three alternative scenarios were developed: the first one modeling a further expansion of population dynamics, the second one projecting a substantial conservation of observed dynamics and the third one consisting of a slowing down of regional dynamics. By dynamics, we mean birth and death ratios as well as emigration and immigration flows.

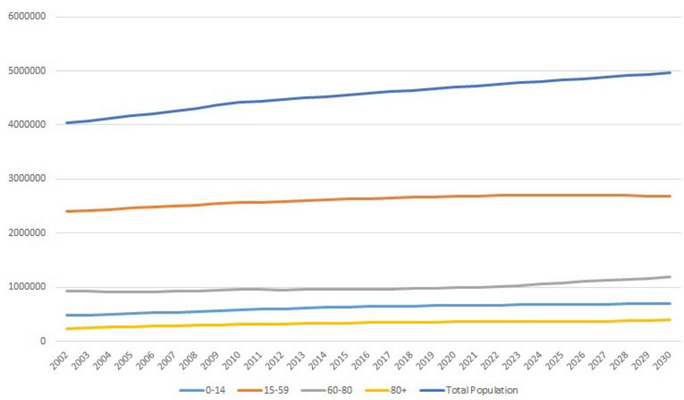


FIGURE 2.2: Emilia-Romagna Population 2002-2030 forecasts

Figure 2.2 depict Emilia-Romagna Population 2002-2030 forecast. An in depth analysis of table 2.1 shows a 12% increase of ER resident population up to 2030 with a consistent increase for elderly (+24%) and young (+18,7%) age bands and a mild increase for central age band (+7%). At 2011 the third scenario seemed the most plausible as the regional population shows the typical characteristics of demographically mature systems with:

- an increase of elderly population with a particular focus on octogenarians;
- an aging and declining trend for working-age population;
- an increase of school-age population.

Age bands	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
0-4	208,076	209,349	212,714	214,083	214,766	215,602	215,937	216,255	216,615	217,117	217,807	218,746	219,952	221,434	223,182	225,205	227,459	229,915	232,546	235,333	238,230
5-9	194,969	198,264	202,224	207,128	212,238	215,684	218,690	220,762	221,859	222,342	223,034	223,220	223,393	223,625	223,991	224,548	225,359	226,434	227,780	229,392	231,249
10-14	181,487	186,302	189,732	194,231	198,476	203,223	206,791	211,149	215,679	220,385	223,488	226,204	228,043	228,957	229,300	229,865	229,930	229,991	230,116	230,393	230,866
15-19	177,878	178,783	179,897	182,640	186,169	190,168	195,893	200,621	204,916	208,968	213,462	216,781	220,853	225,092	229,489	232,328	234,819	236,482	237,243	237,462	237,925
20-24	187,956	190,727	192,814	195,836	197,700	199,207	200,015	201,401	204,000	207,345	211,168	216,652	221,153	225,256	229,107	233,362	236,471	240,301	244,288	248,410	251,025
25-29	231,127	228,304	226,957	225,140	224,990	225,999	228,080	231,408	234,127	235,744	237,060	237,684	238,900	241,230	244,257	247,707	252,762	256,838	260,572	264,057	267,889
30-34	305,354	293,713	288,811	280,818	273,908	269,540	266,719	264,101	262,053	261,545	262,151	263,797	266,659	269,056	270,385	271,465	271,843	272,793	274,750	277,366	280,399
35-39	369,989	366,874	365,177	357,713	348,255	335,048	323,084	312,814	304,919	298,130	293,749	290,872	288,161	285,939	285,203	285,568	286,952	289,549	291,758	292,927	293,861
40-44	372,510	374,798	376,787	380,194	382,281	384,030	383,286	378,258	370,703	361,214	348,109	336,234	326,063	318,206	311,468	307,058	304,136	301,361	299,040	298,175	298,399
45-49	348,910	358,748	363,617	370,842	376,907	378,708	380,855	383,385	386,493	388,374	389,885	388,954	383,761	376,147	366,638	353,600	341,795	331,678	323,851	317,133	312,702
50-54	302,658	311,114	314,523	323,256	334,111	346,969	357,254	366,421	373,292	378,957	380,494	382,386	384,708	387,612	389,367	390,712	389,664	384,387	376,760	367,262	354,317
55-59	267,733	272,633	275,779	283,297	290,456	298,372	306,510	314,257	322,762	333,348	345,896	355,944	364,864	371,475	376,881	378,287	380,028	382,233	385,020	386,727	387,995
60-64	279,104	280,583	275,947	269,299	263,495	263,250	265,236	272,216	279,511	286,450	294,158	302,065	309,691	318,018	328,378	340,665	350,529	359,262	365,716	370,964	372,347
65-69	234,751	233,812	237,892	247,584	257,029	263,496	273,902	265,901	259,737	254,410	254,365	256,451	263,261	270,377	277,149	284,699	292,429	299,948	308,135	318,277	330,313
70-74	239,388	242,001	240,455	236,605	229,485	223,178	212,234	221,446	230,797	239,866	246,077	255,745	248,640	243,250	238,655	238,968	241,270	247,945	254,894	261,516	268,909
75-79	201,741	202,052	198,827	200,942	205,426	207,652	209,881	210,579	207,542	201,817	196,819	187,684	196,576	205,433	213,974	219,830	228,491	222,641	218,347	214,802	215,633
80-84	159,294	161,059	159,373	157,486	157,293	158,772	158,864	157,695	159,834	164,053	166,466	168,859	169,641	167,538	163,484	160,053	153,193	161,450	169,455	177,105	182,339
85+	151,093	156,724	163,911	170,415	175,602	179,842	184,913	188,251	190,437	193,139	196,382	198,906	199,148	200,997	204,931	208,232	211,449	212,031	211,832	211,907	212,078

TABLE 2.1: Population 2010-2030 forecasts

However we decided to use the second scenario projections because regional health system has a strong passive patient mobility (patients coming from abroad or from other regions) that we want to consider as a component of the overall demand.

General evaluation methods on staffing level are based on simple physician to population ratios and, since we would like to test different demand drivers to discuss future imbalances, we define as a first demand driver the physician to population ratio assuming that 2011 staffing levels were appropriate to cope with ER population needs. It is clear that for some specialties such as gynecology and obstetrics, geriatrics, neuropsychiatry and pediatrics, physician to population ratios should be estimated with respect to specific population segments.

2.3.2.4 Service utilization scenario

A predictive model based on epidemiological data can be developed only if, in addition to disease incidence and prevalence information availability, medical care pathways are defined for each disease. By means of care pathways, we define the set of activities and related human resources that are involved during patient's treatment. Emilia-Romagna only collects epidemiological data for the most relevant chronic diseases and very little information is available on care pathway standards. As stated in Persaud and Narine [1999], in case of scarce or not consistent epidemiological data, population need of services, expressed and satisfied through annual inpatient services (emergency and planned) and outpatient ones (diagnostic appointments, treatment and rehabilitation), can be used as a proxy for physician requirement definition. Service utilization data for

(A) Public sector SDO rates in 2002 and 2011 (B) Public sector ASA rates in 2002 and 2011

Figure 2.3: Public sector service utilization rates in 2002 and 2011

the public and private hospitals as well as for outpatient appointments are recorded on Siseps regional database. Inpatient activities are related to hospital discharge (SDO) records while outpatient activities are recorded in the specialist outpatients care (ASA) database.

ASA and SDO records can be related to a specific specialty representing a valuable indicator of specialty-specific demand of care. As a starting point we analyzed how service utilization has changed for inpatient and outpatient activities in the last decade. In the 2002-2011 period a general reduction of public hospitalization can be identified (see figure 2.3). This phenomenon is mild in the 10-40 age band and more consistent in the 40+ one. The peak of service utilization reduction has been recorded for males and females in the 70-74 age band (-22,5% and -21.9%), while a reverse trend has been recorded in the 0-5 age band (+69,4% and +111,1%). Inpatient activities decrease has been balanced by a general increase of outpatient activities demand that has been mainly focused on Medical Specialty disciplines. In particular, we note that male population shows higher consumption rates than women up to 14 years old and starting from 60-64 years. For women, service utilization increase is mainly related to child-bearing age and shows a significant increase in the 40-44 age band. Starting form 2002-2011 data set we extrapolated specific trends for each 'discipline-sex-age' combination up to 2030. The 2002-2011 trend lines are projected until 2021 and then a bounding factor is considered until 2030.

We defined a service utilization driven scenario by applying utilization rates forecasts to 2011-2030 population projections and by assuming that 2011 staffing level was appropriate to cope with service demand in that year.

2.3.2.5 Guidelines-driven scenario

The number of beds assigned to a given specialty is a structural constraint that is often used to estimate the need of health care professionals. This indicator can be considered appropriate when most of the activities involving a specific specialty are related to hospitalization. This is particularly true for surgical area specialties where peri-operative activities are a core element. We have decided to include this driver, because one of the key points of the spending review on health is the reduction or conversion of beds within the public health sector. Beds for discipline can be therefore considered as a constraint that bounds the overall number of specialists that can work in the public sector. The definition of a bed-driven scenario is based on two main elements (see Table 2.2):

- Number of beds available in the public sector for each specialty at a regional level after Decreto 'Balduzzi' criteria to the regional number of public beds available in 2011
- Abruzzo's staffing standards defined by a decree of government-appointed commission.

Medical specialties	Care complexity	Number of medical doctors per bed	Number of Beds
General Surgery	Medium	0,28	1032
Pediatric Surgery	Medium	0,34	63
Plastic Surgery	Medium	0,34	42
Maxillo-facial Surgery	Medium	0,34	59
Gynecology and Obstetrics	Low Medium	0,28	792
Neurosurgery	High	0,46	60
Ophthalmology	Low	0,28	75
Orthopedic and Traumatology	Low	0,28	1118
Otolaryngology	Low	0,28	232
Urology	Low	0,28	429
Cardiac Surgery	High	0,46	10
Thoracic Surgery	Medium High	0,34	79
Vascular Surgery	Medium High	0,34	156
Geriatrics	Low	0,3	293
Thermal Medicine	Low	0,24	2262
Emergency medicine	High	0,3	196
Neurophysiology	Medium High	0	95
Neurology	Medium High	0,3	271
Pediatric neuropsychiatry	Medium	0,3	22
Psychiatry	High	0,42	228
Dermatology and Venereology	Low	0,24	31
Hematology	Medium High	0,3	122
Endocrinology	Low	0,24	20
Gastroenterology	Low	0,24	139
Cardiovascular diseases	Medium High	0,3	478
Respiratory system diseases	Low	0,24	249
Infectious diseases	Low	0,24	214
Nephrology	Medium High	0,3	138
Rheumatology	Low	0,24	23
Nuclear medicine	Low	0,24	18
Radiology	Low	0	2
Radiation therapy	Low	0,3	6
Anesthesiology and Intensive care medicine	Very High	1	373
Physical medicine and rehabilitation	Low	0,24	329

TABLE 2.2: Abruzzo guidelines

The System Dynamics model integrates the three scenarios described in the previous section by importing both regional population forecasts and weighting factors such as 2011-2030 projections on ASA and SDO utilization rates for Scenario 2 and Abruzzo guidelines and regional beds for Scenario 3.

2.3.2.6 Shortage and surplus indicators

Once demand and supply representation were defined and integrated in the SD model (Figure 2.4), we identified two indicators that can be used in order to evaluate Regional Health system forecasts in terms of shortages and surpluses during the planning horizon. We define employment needs as the difference between the projections of the initial stock (reduced over years by retirement rates) and the expected requirement, defined by demand scenarios. We define training needs as the gap between the overall supply projections and the expected demand. By overall supply we mean the sum of all the professionals active at a regional level, either employed or specialized but not yet employed.

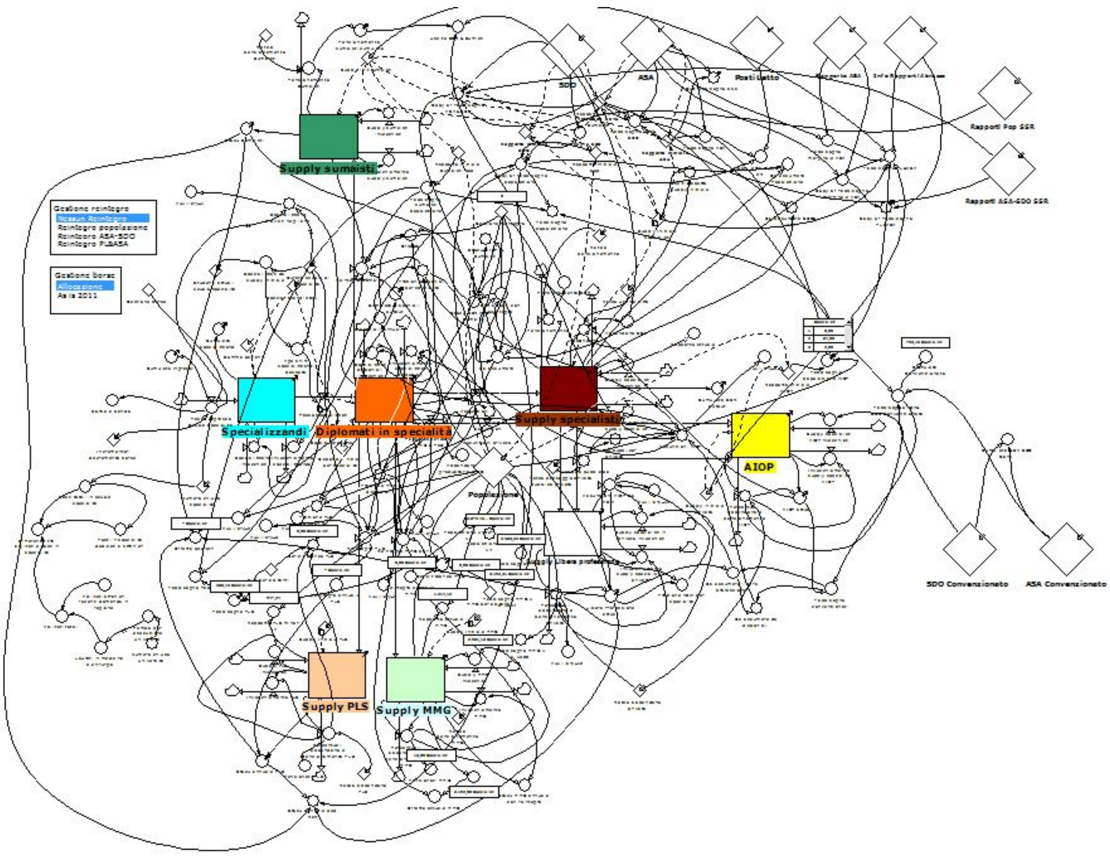


FIGURE 2.4: System Dynamics model

2.3.3 Simulation results

Since HRH planning is a long-term strategic problem we defined a twenty year planning horizon based on 2011 data. We considered in this first phase of the work that the training supported by MIUR will remain constant as observed for the academic year 2012-2013 observing if its funded grants will meet demand forecasts for each of the three demand scenarios (see 2.5.1).

2.3.3.1 Surgical Area

Table 2.3 shows that Physician of General and Specialistic surgery classes, with the exception of Plastic surgery, in 2011 were mainly employed in the public sector and just very few of them were outpatient specialists (SUMAI). Table 2.4 [2] shows that Scenario 1, that is linked to general or specific segments of the population (≤14 years), drives to a general and significant increase, around 12%, for all surgical specialties (detailed forecasts in Subsection 2.5.2).

Medical Specialties	Trained stock		SSR Stock		Sumai Stock		AIOP Stock		Supply Stock	
	2007-2011	2030 MIUR	**2011**	**2030**	**2011**	**2030**	**2011**	**2030**	**2011**	**2030**
Digestive System Surgery	0	0	0	0	0	0	0	0	0	0
General Surgery	22	488	493	134	3	1	84	74	602	696
Pediatric Surgery	2	8	45	22	0	0	2	2	49	32
Plastic Surgery	5	76	26	13	0	0	34	22	65	111
Maxillo-facial Surgery	6	84	30	12	0	0	8	9	44	105
Gynecology and Obstetrics	21	421	450	140	57	18	33	31	561	610
Neurosurgery	3	34	65	19	0	0	28	17	96	70
Ophthalmology	9	155	169	54	96	31	57	44	331	284
Orthopedic and Traumatology	21	467	453	156	14	1	156	148	644	772
Otolaryngology	10	152	125	39	45	11	31	23	211	225
Urology	5	138	161	52	8	3	22	23	196	216
Cardiac Surgery	4	78	27	13	0	0	17	12	48	102
Thoracic Surgery	2	81	43	18	0	0	6	2	51	102
Vascular Surgery	4	135	68	26	1	1	16	14	89	176

TABLE 2.3: Surgical Class Supply: 2011 status quo and 2030 forecasts

Medical Specialties	SSR + Sumai demand			Aiop Demand	Employment gap 2030			Training gap 2030		
	Scenario 1	**Scenario 2**	**Scenario 3**		**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**
Digestive System Surgery	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
General Surgery	556	559	537	94	442	445	423	46	43	65
Pediatric Surgery	50	53	51	2	29	32	30	-21	-24	-22
Plastic Surgery	29	27	29	38	32	30	32	44	46	44
Maxillo-facial Surgery	34	39	33	9	21	27	21	63	57	63
Gynecology and Obstetrics	568	549	536	37	416	397	384	5	24	37
Neurosurgery	73	79	84	31	68	75	80	-34	-41	-46
Ophthalmology	297	317	314	64	232	252	249	-77	-97	-94
Orthopedic and Traumatology	523	544	497	175	393	414	367	74	53	100
Otolaryngology	190	205	191	35	152	166	152	0	-14	0
Urology	189	207	180	25	136	154	127	2	-16	11
Cardiac Surgery	30	31	28	19	25	26	23	53	52	55
Thoracic Surgery	48	60	51	7	34	46	37	47	35	44
Vascular Surgery	77	89	71	18	54	66	48	81	69	87

TABLE 2.4: Surgical Class Demand: 2011 status quo and 2030 forecasts

General Surgeries class. Table 2.4 shows that Scenario 2 (ASA + hospitalizations (SDO)) would result in the 2030 pediatric surgeons increase of 18% due to the expected increase of the residents in the 0-14 age band and to the rising rates of outpatient and

[2] N.D. is an acronym of Not Defined because no training school is active in Emilia-Romagna Region

inpatient activities. Scenario 3, the one that links the amount of inpatient activities to available hospital beds according to Abruzzo guidelines, reduces the increase of the requirement of general surgery specialist if compared to Scenario 2 (+9% vs. +13%). This means that inpatient activities have a significant impact on specialists workload. The implemented national funding at 2011, if maintained during the planning horizon, seems to cover the overall regional demand for General Surgery specialists. Instead Pediatric Surgery specialty will face a lack of workforce because at the moment no Pediatric specialization school is active in the Emilia-Romagna territory.

Specialistic Surgeries class. Table 2.3 shows that among Specialistic surgeries some specialties are characterized by an higher level of SUMAI workforce namely, 30% for ophthalmology and 22% for otolaryngology, while Plastic surgery, Neurosurgery and Othopedic and Traumatology are characterized by an higher level of specialists working in private hospitals. Scenario 2 is the one that produceds a generalized major increase in Specialistic Surgery physicians requirement. This phenomenon, particularly clear for Urology and Maxillo-facial surgery, is due to the significant increase of inpatient and outpatient activities per person rather than to population demographic forecasts. The physicians requirement in Scenario 3 for all Specialistic Surgery specialties but Neurosurgery is generally lower than thai in Scenario 1 and Scenario 2 because Abruzzo guidelines classify those specialties as having a low or medium complexity of care. The implemented national funding at 2011, if maintained during the planning horizon, will not meet estimated physician demand for Ophthalmology and Neurosurgery specialties in each demand scenario. Instead as far as Otolaryngology and Urology concern only for Scenario 2 forecasts show a deficit of trained physicians.

Cardiac thoracic and vascular Surgery class. Cardiac thoracic and vascular Surgery specialties are mainly active in Private Hospitals within the Emilia-Romagna region. Specialist demography is characterized by young physicians, so retirements will mildly affect supply numerical consistency. The implemented national funding at 2011, although small in absolute volume, seems to satisfy all demand scenarios.

2.3.3.2 Medical Area

The medical area has a great importance for the Public sector because it groups the most significant specialties from the numerical point of view such as Pediatrics, Internal medicine, Emergency medicine, Psychiatry and Cardiology. Table 2.6 [3] shows that Scenario 1 forecasts for geriatrics, pediatrics and neuropsychiatry, which are associated to 0-14 and 65+ age bands, a demand rising of 13% on average. A general reduction of

[3] N.D. is an acronym of Not Defined because no training school is active in Emilia-Romagna Region

the 15-65 age bands population in 2012-2030 planning horizon will contain the increasing demand for Psychiatric services. Medical Area specialties are the most stressed by Scenario 2 forecasts because they are characterized by an increasing utilization rate of outpatient appointments and elective surgeries. This effect is remarkable for Nephrology (+46%), Cardiology (+40%), Neuropsychiatry (+37%) and Rheumatology (+35%). Scenario 3 forecasts generally increase less than Scenario 2 ones since these specialties, with the exception of Psychiatry and Emergency medicine, are classified as low or medium from the complexity of care point of view. In general, for Scenarios 2 and 3, the implemented national funding at 2011, if maintained during the planning horizon, will not satisfy the overall specialists requirement (detailed forecasts in Subsection 2.5.3).

Medical Specialties	Trained stock		SSR Stock		Sumai Stock		AIOP Stock		Supply Stock	
	2007-2011	2030 MIUR	2011	2030	2011	2030	2011	2030	2011	2030
Geriatrics	12	238	198	61	20	16	19	16	249	331
Sports Medicine	3	76	5	3	27	6	4	3	39	88
Primary Health care	0	0	0	0	2	2	0	0	2	2
Internal Medicine	27	464	970	253	6	1	21	21	1024	739
Thermal Medicine	0	0	0	0	0	0	0	0	0	0
Oncology	16	261	190	60	6	1	6	6	218	327
Emergency medicine	3	107	590	236	0	0	5	4	598	347
Neurofiisopathology	0	0	0	0	0	0	0	0	0	0
Neurology	12	191	181	54	28	14	24	18	245	278
Pediatric neuropsychiatry	7	137	134	45	32	14	5	6	178	202
Psychiatry	22	395	520	113	7	5	63	66	612	579
Clinical pathology	10	10	1	0	5	4	13	9	29	23
Allergology and Immunology	2	8	0	0	6	3	2	0	10	11
Dermatology and Venereology	12	136	84	22	49	17	17	12	162	186
Hematology	7	143	172	40	0	0	4	4	183	187
Endocrinology	7	125	79	19	34	25	15	8	135	178
Gastroenterology	11	141	129	56	4	1	26	11	170	209
Cardiovascular diseases	27	499	429	132	45	16	79	71	580	718
Respiratory system diseases	13	208	138	43	7	3	18	13	176	267
Infectious diseases	5	90	107	26	0	0	2	2	114	118
Tropical medicine	0	0	0	0	0	0	0	0	0	0
Nephrology	12	241	136	45	3	0	6	6	157	292
Rheumatology	5	87	22	10	11	7	3	1	41	104

TABLE 2.5: Medical Class Supply: 2011 status quo and 2030 forecasts

Medical Specialties	SSR + Sumai demand			Aiop Demand	Employment gap 2030			Training gap 2030		
	Scenario 1	Scenario 2	Scenario 3		Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
Geriatrics	262	288	253	23	192	218	183	46	20	55
Sports Medicine	36	45	36	4	28	38	28	48	38	48
Primary Health care	2	2	2	0	0	N.D.	N.D.	0	N.D.	N.D.
Internal Medicine	1094	1127	1072	24	842	876	821	-378	-412	-357
Thermal Medicine	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Oncology	220	248	222	7	160	188	162	101	73	99
Emergency medicine	661	652	649	6	427	418	415	-320	-311	-308
Neurofiisopathology	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Neurology	234	282	254	27	174	222	194	17	-31	-3
Pediatric neuropsychiatry	195	228	230	6	136	169	171	1	-32	-34
Psychiatry	586	547	541	70	471	432	427	-76	-37	-32
Clinical pathology	7	19	19	15	8	21	21	2	-11	-11
Allergology and Immunology	7	7	8	2	6	7	7	2	1	1
Dermatology and Venereology	149	155	151	19	118	124	120	18	12	16
Hematology	193	230	214	4	153	191	175	-10	-48	-32
Endocrinology	127	153	154	17	91	117	118	34	8	7
Gastroenterology	149	184	171	29	110	146	132	31	-5	9
Cardiovascular diseases	531	686	551	89	400	555	420	99	-56	79
Respiratory system diseases	162	187	172	20	124	149	133	84	59	75
Infectious diseases	120	125	115	2	94	99	89	-4	-9	1
Tropical medicine	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Nephrology	156	205	179	7	112	161	135	129	80	106
Rheumatology	37	45	45	3	23	31	31	64	56	56

TABLE 2.6: Medical Class Demand: 2011 status quo and 2030 forecasts

General Medicine class. Table 2.5 shows that with the exception of the Sports medicine specialty, in 2011 most of General Medicine class physicians were employed in the Public sector. Increased service utilization rates both for inpatient and outpatient activities in conjunction with 0-14 and 65+ age band population growth will result in a demand rise for Sports medicine specialists(+38%), Geriatricians (+31%) and Oncologists (26%), while Internal Medicine and Emergency medicine specialists will have a limited increase (11% and 15%) because their activity is mainly focused on the 15-65 age band population (see Table 2.6). The national funding at 2011-2012, if maintained during the planning horizon, seems unable to cope with Internal Medicine demand for any demand scenario. This increasing gap is due to multiple factors, it is possible to observe in the 2001-2010 period a general decrease of funded grants for this specialty that, in conjunction with a negative turnover (more retirements than hirings) and an increased average age for employed specialists, will lead to a sudden decrease of active physicians. The increasing training gap for the Emergency medicine specialty is due to a rise of service demand and to a significant retirement in the Public sector that can't be balanced by MIUR funded grants.

Neuroscience and Mental disorders class. Table 2.6 shows that, according to Scenario 2 forecasts, Neurologists demand may increase up to 33%. In this case MIUR funded grants will not be sufficient to cover the demand of both Public and Private sectors. Psychiatry seems to be the specialty with the least demand increase because, as stated before, a general reduction of the 15-65 age band population in the 2012-2030 planning horizon will contain the generally increasing demand for Psychiatric services. The national funding at 2011-2012, if maintained during the planning horizon, will not be able to satisfy any of the 3 scenarios because, as observed in the 2001-2010 period, to a positive turn-over of psychiatrists (more hirings that retirements) does not correspond an increase of funded grants. Pediatric neuropsychiatry demand will increase up to 37-38% in Scenario 2 and 3 due to the joint action of the 0-14 age band population increase and outpatient appointment utilization rate. In this case MIUR funded grants will not be able to cope with projected workforce requirements.

2.3.3.3 Diagnostic and Clinical Services Area

As for Medical specialties class, this area contains some of the most significant specialties from the numerical point of view such as anesthesia and intensive care, diagnostic radiology and hygiene and preventive medicine (see Table 2.7). These specialties have been strongly influenced, in recent decades, by technological and organizational innovations like the increasing importance of health professions introduced to replace, in some cases, medical specialties. As a result over the 2002-2011 period a change in productivity has been recorded, resulting in a lower demand of physicians. As far as Scenario 2 and 3 demand drivers concern, it is important to note that for most of Diagnostic and Clinical Services areas no stay-bed is available, anesthesia and intensive care being the only exception. This turns to a Scenario 2 demand that will be mainly influenced by outpatient appointment services, and in some cases even ASA records are not available. For those specialties we decided to keep the 2011 staffing level as the appropriate one. As a first and general analysis of Table 2.8 [4] we can observe that Scenario 1 shows an increase of 12% in 2030, while Scenarios 2 and 3 are not homogeneous. Scenario 2 defines, for those specialties where outpatient activities have been recorded, an increase of 15% on average in 2030, while Scenario 3 is rather similar to Scenario 2 (detailed forecasts in Subsection 2.5.4).

Medical Specialties	Trained stock		SSR Stock		Sumai Stock		AIOP Stock		Supply Stock	
	2007-2011	2030 MIUR	2011	2030	2011	2030	2011	2030	2011	2030
Nuclear medicine	4	65	53	20	0	0	2	2	59	87
Radiology	47	673	659	218	11	2	59	56	776	949
Radiation therapy	7	158	63	26	0	0	5	5	75	190
Hospital pharmacy	6	61	0	0	0	0	0	0	6	61
Health physics	0	7	0	0	0	0	0	0	0	7
Anatomic pathology	4	166	130	32	0	0	0	0	134	198
Clinical biochemistry	10	28	0	0	0	0	0	0	10	28
Microbiology and Virology	12	59	23	11	0	0	0	0	35	70
Clinical pathology	14	123	147	26	6	1	1	1	168	151
Pharmacology	0	0	56	12	0	0	0	0	56	12
Medical genetics	8	17	21	9	0	0	0	0	29	26
Food science	9	87	8	3	2	2	3	3	22	95
Anesthesiology and Intensive care medicine	54	873	935	345	3	0	71	73	1063	1291
Audiology	2	6	7	1	8	0	0	0	17	7
Physical medicine and rehabilitation	22	226	207	54	27	10	28	27	284	317
Toxicology	0	0	0	0	2	2	1	1	3	3
Oral surgery	0	0	0	0	0	0	0	0	0	0
Orthodontics	24	72	0	0	107	35	10	8	141	115
Preventive healthcare	21	307	574	103	16	3	6	5	617	417
Space medicine	0	0	0	0	0	0	0	0	0	0
Occupational medicine	15	226	116	20	5	0	1	1	137	247
Forensic pathology	7	41	80	26	11	4	1	1	99	71
Health statistics	0	0	0	0	0	0	0	0	0	0

Table 2.7: Diagnostic and Clinical Services Class Supply: 2011 status quo and 2030 forecasts

Medical Specialties	SSR + Sumai demand			Aiop Demand	Employment gap 2030			Training gap 2030		
	Scenario 1	Scenario 2	Scenario 3		Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
Nuclear medicine	59	62	62	2	39	42	42	26	23	23
Radiology	751	899	898	66	540	689	688	133	-16	-15
Radiation therapy	71	90	90	6	45	65	64	113	93	94
Hospital pharmacy	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Health physics	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Anatomic pathology	146	150	130	0	113	118	98	53	48	68
Clinical biochemistry	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Microbiology and Virology	26	26	23	0	15	16	12	44	43	47
Clinical pathology	171	154	154	1	145	127	127	-22	-4	-4
Pharmacology	63	64	56	0	51	52	44	-51	-52	-44
Medical genetics	24	27	21	0	14	18	12	3	-1	5
Food science	11	9	13	3	7	4	8	80	83	79
Anesthesiology and Intensive care medicine	1051	1284	1153	80	712	946	814	161	-73	59
Audiology	17	15	15	0	16	14	14	-10	-8	-8
Physical medicine and rehabilitation	262	237	235	31	203	177	175	23	49	51
Toxicology	2	2	3	1	0	0	1	0	0	-1
Oral surgery	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Orthodontics	120	117	117	11	88	85	85	-16	-13	-13
Preventive healthcare	661	596	596	7	558	493	493	-251	-186	-186
Space medicine	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Occupational medicine	136	167	122	1	116	147	102	110	79	124
Forensic pathology	102	92	92	1	73	63	63	-32	-22	-22
Health statistics	0	0	0	0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

Table 2.8: Diagnostic and Clinical Services Class Demand: 2011 status quo and 2030 forecasts

[4] N.D. is an acronym of Not Defined because no training school is active in Emilia-Romagna Region

Therapeutic and diagnostic services class. Pathologists and clinical pathologists are employed in the public sector. Demand scenarios seem to be over satisfied by the national funding policy, and only Clinical pathologists might suffer a 10% shortage from 2024 for Scenario 1.

Medical imaging and Radiation therapy class. Medical imaging and Radiation therapy class specialists are mainly active in the public health sector (see Table 2.7). For each specialty no stay-bed is available and only outpatient appointments will influence Scenario 2 and Scenario 3 forecasts. Table 2.8 shows the increasing importance of diagnostic techniques related to nuclear medicine, diagnostic radiology and radiotherapy specialties is emphasized in Scenario 2 and Scenario 3 forecasts where an increase of 41% for Radiation therapy and of 32% for Radiology can be observed. The national funding policy has invested over the year on these specialties, to the point that even Scenario 2 and 3 would seem satisfied in 2030. As an example Public Radiologist supply, that will face a 67% decrease due to retirements and a 32% demand increase in Scenario 2, will be satisfied just by national funded grants.

Specialistic clinical services class. Anesthesiology and Intensive care medicine is by far the most supported specialty by MIUR having 9,51% of the overall trained specialists during the simulation period. The demand for anesthesiologists increases, during the planning horizon, up to 35% in Scenario 2, while only a 2% increase is forecasted by Scenario 3 due to Intensive Care Unit (ICU) beds constraints defined by Abruzzo's guidelines (see Table 2.8). MIUR training policy would meet future demand according to Scenario 1 and Scenario 3 forecasts, while it must be integrated with additional regional grants in Scenario 2. Physical medicine and rehabilitation, whether linked to ASA and SDO activities, is much more limited compared to Population driven scenario. Scenario 2 and 3 forecasts only a 2% increase of the overall demand while a 12% increase is expected for Scenario 1. Scenario 2 and 3 forecasts are influenced both by demographic factors, such as mild increase of the population within the 15-65 band, and by service utilization rates, namely the decrease of outpatient and inpatient utilization rates. The national funded grants will largely satisfy future requirements.

2.3.4 An optimal allocation model

The construction of a simulation model that defines supply and demand scenarios can be considered as a first fundamental step to support the management of healthcare resources. The identification of gaps between supply and demand allows an initial analysis of training needs through the comparison between the current training policy and regional employment needs under various demand scenarios. Such an analysis is useful to detect that, for some specialties, it may happen that trained physicians will not be sufficient to meet future demand and for some others the current training will be excessive. A comprehensive decision support tool must therefore define allocation rules for specialization training grants. The decision of building a dynamic allocation tool is based on the assumption that a long-term planning should consider the structural impact of decisions by focusing not only on annual perceived shortage, but also considering that for strategic planners it could be wiser to modulate grants funding policies over the entire planning horizon. The allocation of training grants at a regional level can follow two different criteria:

- By assuming that decision power is at a regional level the decision is taken only on additional grants funded by regional health budget because national funded grants (MIUR) are considered as exogenous variables that cannot be modified by regional managers;
- By assuming that regional managers can influence national funding policies, then the decision is taken on all contracts activated within Emilia-Romagna region by considering both MIUR and regional funded.

2.3.4.1 Allocation of regional funded grants

As stated in the previous section Emilia-Romagna Region can fund and allocate a finite number of specialization grants. The number of additional funded positions can be spread over the fifty medical specialties at the regional level. The proposed model considers that the purpose of a good allocation policy is to reduce the overall training gap considered either as scarce or excessive. The final objective is to find the allocation policy that minimizes the overall gap value in the 2011-2024 planning horizon. We consider 2024 as the last year in which a decision that will affect the system can be taken because a 6 year training grant funded in 2024 will affect 2030 training gap.

$$MinZ = \sum_{i \in S} \sum_{t \in T} (p_i / Supply_{i\ 2011}) y_{it}$$

$$\sum_{i \in S} x_{it} \leq Grants_t \qquad \forall t \in T \tag{2.1}$$

$$\sum_{l=2011}^{t} x_{il} + y_{it} = gap_{it} \qquad \forall t \in T, \qquad \forall s \in S \tag{2.2}$$

$$x_{it} \leq BoundGrants_t \qquad \forall t \in T \qquad \forall s \in S \tag{2.3}$$

$$x_{it} \geq 0 \qquad integer \qquad \forall t \in T, \qquad \forall s \in S \tag{2.4}$$

$$y_{it} \geq 0 \qquad integer \qquad \forall t \in T, \qquad \forall s \in S \tag{2.5}$$

There the Mixed-Integer Linear Programming model reads as follows: where S is the set of medical specialties that is composed of 61 classified specialties, T is the planning horizon such that $t \in \{2011, \ldots, 2030\}$. For each medical specialty $i \in S$ we define d_i as the duration of its specialization school, $FabE_{it}$ as its employment need on year t, $Supply_{it}$ as the supply of medical specialists of type i at year t, (already employed in 2011), and $Trained_{it}$ as the number of Medical specialists trained at year t. We then define p_i as the priority coefficient of specialty i, which is computed as a combination of the care complexity classification at specialty $i \in S$ and its numerical relevance for the public sector. Parameter gap_{it} is the training gap of specialty i on year t calculated as $gap_{it} = max(FabE_{i(t+d_i)} - Supply_{i(t+d_i)} - Trained_{i(t+d_i)}, 0)$. As a last element we define $Grants_t$, as the overall number of grants funded by Emilia-Romagna region on year t and $BoundGrants_t$ the maximum number of grants that can be allocated to a single speciality on year t. It is clear that this bound can lead to sub optimal solutions, but, it is unlikely that regional grants can be all allocated to one single specialty because of political reasons.

Decision variables x_{ij} defines how many grants will be allocated on year t to specialty i, while y_{ij} variables measures the training gap of specialty i on year t. This last measure is the one that we want to minimize. Constraint 2.1 ensure that on year t at most $Grants_t$ grants are allocated to medical specialties. Constraint 2.2 ensure that, tacking in consideration previous years allocations, on year t the number of grants assigned to specialty i will not exceed its demand. Constraint 2.3 bounds the maximum number of grants that can be allocated to speciality i on year t

The model is built to capture a rational allocation of training grants and uses the following data:

- Additional number of grants annually financed: 25
- A maximum of 5 grants assigned to a specialty annually: i.e. 20% of regional financed grants;
- Complexity of care indexes (intensive: 2.25, high: 2, medium-high: 1.75, mean: 1.50, medium-low: 1.25, low: 1, predominantly outpatient or service: 0.50).

Medical classes	**Number of grants**		
	Scenario 1	**Scenario 2**	**Scenario 3**
Surgical	74	71	98
Medical	238	277	237
Services	36	0	13
Total	348	348	348

Table 2.9: Class cumulative allocation 2011-2024

Table 2.9 summarizes the cumulative allocation of 25 scholarships per year (23 in 2011 and 25 from 2012) for the 2011-2024 period aggregating schools according to their disciplinary area. It is clear that, for each scenario, additional grants would focus on medical area specialties, while Services area requires a small number of grants due to the high impact of MIUR grants. Table 2.10 shows the number of grants allocated to each specialty and the percentage with respect to the total number of funded positions. Emergency medicine is the specialty with the highest impact (between 22% and 34% depending on demand scenario). Other priority specialties are pediatrics, Internal Medicine (Scenario 1), psychiatry and neuropsychiatry. Psychiatry and Neuropsychiatry receive a large number of grants from the allocation model due to different factors psychiatry workforce will be strongly reduced due to retirements (e.g. only the 21% of SSR specialists will still be active at the 2030), while neuropsychiatry will be affected by a significant increase of the number of population in the 0-6 age band.

As a conclusion it is important to remark that the allocation of 25 additional positions is not sufficient to close training gaps even for some of the specialties that have been strongly affected by allocation policies. Taking as an example emergency medicine, which receives up to 70 grants under Scenario 1 and 3, only 21% of its training need will be covered with the proposed allocation policies; a similar problem can be observed for pediatrics where only 9% of its training need will be covered with the proposed allocation policies.

Medical Specialty	Number of grants			Percentage w.r.t. allocated grants		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
Digestive System Surgery	0	0	0	0.00%	0.00%	0.00%
General Surgery	0	0	0	0.00%	0.00%	0.00%
Pediatric Surgery	0	0	0	0.00%	0.00%	0.00%
Plastic Surgery	0	0	0	0.00%	0.00%	0.00%
Maxillo-facial Surgery	0	0	0	0.00%	0.00%	0.00%
Gynecology and Obstetrics	0	0	0	0.00%	0.00%	0.00%
Neurosurgery	34	43	45	9.80%	12.36%	12.93%
Ophthalmology	39	0	53	11.24%	0.00%	15.23%
Orthopedic and Traumatology	0	0	0	0.00%	0.00%	0.00%
Otolaryngology	0	10	0	0.00%	2.87%	0.00%
Urology	0	18	0	0.00%	5.17%	0.00%
Cardiac Surgery	0	0	0	0.00%	0.00%	0.00%
Thoracic Surgery	0	0	0	0.00%	0.00%	0.00%
Vascular Surgery	0	0	0	0.00%	0.00%	0.00%
Emergency medicine	70	51	70	20.17%	14.66%	20.11%
Pediatrics	40	40	35	11.53%	11.49%	10.06%
Geriatrics	0	0	0	0.00%	0.00%	0.00%
Sports Medicine	0	0	0	0.00%	0.00%	0.00%
Primary Health care	0	0	0	0.00%	0.00%	0.00%
Internal Medicine	45	0	33	12.97%	0.00%	9.48%
Thermal Medicine	0	0	0	0.00%	0.00%	0.00%
Oncology	0	0	0	0.00%	0.00%	0.00%
Neurofisopathology	0	0	0	0.00%	0.00%	0.00%
Neurology	0	36	3	0.00%	10.34%	0.86%
Pediatric neuropsychiatry	0	28	34	0.00%	8.05%	9.77%
Psychiatry	70	37	31	20.17%	10.63%	8.91%
Clinical pathology	0	0	0	0.00%	0.00%	0.00%
Allergology and Immunology	0	0	0	0.00%	0.00%	0.00%
Dermatology and Venereology	0	0	0	0.00%	0.00%	0.00%
Hematology	10	51	31	2.88%	14.66%	8.91%
Endocrinology	0	0	0	0.00%	0.00%	0.00%
Gastroenterology	0	9	0	0.00%	2.59%	0.00%
Cardiovascular diseases	0	16	0	0.00%	4.60%	0.00%
Respiratory system diseases	0	0	0	0.00%	0.00%	0.00%
Infectious diseases	3	9	0	0.86%	2.59%	0.00%
Tropical medicine	0	0	0	0.00%	0.00%	0.00%
Nephrology	0	0	0	0.00%	0.00%	0.00%
Rheumatology	0	0	0	0.00%	0.00%	0.00%
Nuclear medicine	0	0	0	0.00%	0.00%	0.00%
Radiology	0	0	10	0.00%	0.00%	2.87%
Radiation therapy	0	0	0	0.00%	0.00%	0.00%
Hospital pharmacy	0	0	0	0.00%	0.00%	0.00%
Health physics	0	0	0	0.00%	0.00%	0.00%
Anatomic pathology	0	0	0	0.00%	0.00%	0.00%
Clinical biochemistry	0	0	0	0.00%	0.00%	0.00%
Microbiology and Virology	0	0	0	0.00%	0.00%	0.00%
Clinical pathology	21	0	3	6.05%	0.00%	0.86%
Pharmacology	0	0	0	0.00%	0.00%	0.00%
Medical genetics	0	0	0	0.00%	0.00%	0.00%
Food science	0	0	0	0.00%	0.00%	0.00%
Anesthesiology and Intensive care medicine	0	0	0	0.00%	0.00%	0.00%
Audiology	0	0	0	0.00%	0.00%	0.00%
Physical medicine and rehabilitation	0	0	0	0.00%	0.00%	0.00%
Toxicology	0	0	0	0.00%	0.00%	0.00%
Oral surgery	0	0	0	0.00%	0.00%	0.00%
Orthodontics	0	0	0	0.00%	0.00%	0.00%
Preventive healthcare	15	0	0	4.32%	0.00%	0.00%
Space medicine	0	0	0	0.00%	0.00%	0.00%
Occupational medicine	0	0	0	0.00%	0.00%	0.00%
Forensic pathology	0	0	0	0.00%	0.00%	0.00%
Health statistics	0	0	0	0.00%	0.00%	0.00%
Total	348	348	348	100.00%	100.00%	100.00%

TABLE 2.10: Cumulative allocation of 25 grants per year (2012-2024) classified for general area according to the three demand scenarios (1 demographics, 2 ASA + SDO, 3 standard PL and ASA)

2.3.4.2 Allocation of national and regional grants

As previously mentioned, grants funding allocation can be discussed both in terms of support for regional decision makers and as a tool for national negotiation. In latter case it must be assumed that the problem is characterized by a set of constraints that are different from those previously described for the allocation of the additional regional grants. The negotiation is strongly influenced by the number and the size of regional training centers. It is then important to remark that an allocation model whose objective is the training gap reduction could lead to drastic compressions, or temporary suspension of some university departments. It is then possible that, for some specialties, no grant is funded over the 2011-2024 time horizon. It is not a purpose of the study to define a radical reorganization of regional training centers. It appears however interesting to analyze the results of such a hypothetical assignment in order to quantify the deviation between the optimal allocation without structural constraints and repetitive national policies. In this setting the allocation model for national and regional grants differs from the regional one only for the removal of constraint 2.3.

Medical area	Number of grants				Allocation w.r.t. As-is scenario		
	Scenario 1	Scenario 2	Scenario 3	As-is scenario	Scenario 1	Scenario 2	Scenario 3
Surgical	1492	1327	1429	1617	-7.7%	-17.9%	-11.6%
Medical	3343	3815	3400	2743	21.9%	39.1%	24.0%
Services	1726	1564	1849	1998	-13.6%	-21.7%	-7.5%
Total allocated	6561	6706	6678	6358	3.2%	5.5%	5.0%
Total demand	6561	6900	6678	/	/	/	/

Table 2.11: Comparison of as-is and demand-driven scenarios allocation policies

Table 2.11 summarizes the results of the cumulative allocation of national and regional scholarships (495+23 in 2011 and 451+25 from 2012) per year for the 2011-2024 period by aggregating schools according to their disciplinary area. As a first important information we may say that only Scenario 2 will exploit all available scholarships while a proper management of allocation in Scenario 1 and Scenario 3 can save up to 2.16% of funding budget. The complete utilization of scholarship will not be sufficient to satisfy Scenario 2 requirements where an additional 2.89% would be necessary.

An additional information on the results in Table 2.11 is the increase of grants allocated to medical specialties compared to surgery and service ones, meaning that less importance is given by the current planning policies to a set of specialties that will become very important in future years.

A deeper analysis of the national and regional grant allocation in Table 2.12 shows how scarce is the amount of scholarships for emergency medicine, internal medicine and orthopedic and traumatology surgery. Conversely future needs of plastic surgery, sports medicine vascular and thoracic surgeries as well as respiratory system diseases seem overestimated by current policies.

The most interesting outcome of this analysis is that the current budget seems, for two out of he three scenarios, adequate to fulfill future specialist needs. This means that the scary baby-boomers mass retirement can be managed without a dramatic impact on public funds dedicated to the HRH training.

Medical specialties	Number of grants				Allocation with respect to 'as is' scenario		
	Scenario 1	Scenario 2	Scenario 3	As-is	Scenario 1	Scenario 2	Scenario 3
Digestive System Surgery	0	0	0	0	0.00%	0.00%	0.00%
General Surgery	296	276	277	350	-15.43%	-21.14%	-20.86%
Pediatric Surgery	0	0	0	0	0.00%	0.00%	0.00%
Plastic Surgery	17	16	17	56	-69.64%	-71.43%	-69.64%
Maxillo-facial Surgery	0	5	0	57	-100.00%	-91.23%	-100.00%
Gynecology and Obstetrics	331	312	299	295	12.20%	5.76%	1.36%
Neurosurgery	51	57	62	14	6.25%	0.00%	5.08%
Ophthalmology	179	199	196	99	29.71%	101.01%	28.95%
Orthopedic and Traumatology	360	147	334	337	6.82%	-56.38%	-0.89%
Otolaryngology	104	119	105	100	4.00%	8.18%	5.00%
Urology	105	123	96	98	7.14%	6.03%	-2.04%
Cardiac Surgery	6	7	4	56	-89.29%	-87.50%	-92.86%
Thoracic Surgery	14	26	17	57	-75.44%	-54.39%	-70.18%
Vascular Surgery	29	40	22	98	-70.41%	-59.18%	-77.55%
Pediatrics	655	653	622	348	56.70%	63.66%	48.80%
Geriatrics	143	168	133	168	-31.25%	-19.23%	-34.48%
Sports Medicine	12	21	8	56	-78.57%	-62.50%	-85.71%
Primary Health care	0	0	0	0	0.00%	0.00%	0.00%
Internal Medicine	681	715	659	322	111.49%	122.05%	104.66%
Thermal Medicine	0	0	0	0	-100.00%	0.00%	-100.00%
Oncology	80	108	82	168	-52.38%	-35.71%	-51.19%
Emergency medicine	382	373	370	56	582.14%	566.07%	560.71%
Neurofisopathology	0	0	0	0	0.00%	0.00%	0.00%
Neurology	125	173	145	128	-2.34%	5.49%	10.69%
Pediatric neuropsychiatry	100	133	135	98	2.04%	5.56%	2.27%
Psychiatry	385	346	340	280	10.00%	9.15%	9.32%
Clinical psychology	0	0	0	0	0.00%	0.00%	0.00%
Allergology and Immunology	0	0	0	0	0.00%	0.00%	0.00%
Dermatology and Venereology	70	77	72	85	-17.65%	-9.41%	-15.29%
Hematology	121	158	142	100	10.00%	4.64%	8.40%
Endocrinology	60	87	88	86	-30.23%	1.16%	2.33%
Gastroenterology	65	100	87	85	-23.53%	6.38%	2.35%
Cardiovascular diseases	280	435	300	339	-17.40%	22.54%	-11.50%
Respiratory system diseases	69	94	79	140	-50.71%	-32.86%	-43.57%
Infectious diseases	66	71	61	58	8.20%	5.97%	5.17%
Tropical medicine	0	0	0	0	0.00%	0.00%	0.00%
Nephrology	49	98	72	169	-71.01%	-42.01%	-57.40%
Rheumatology	0	5	5	57	-100.00%	-91.23%	-91.23%
Nuclear medicine	21	24	23	42	-50.00%	-42.86%	-45.24%
Radiology	370	519	517	437	-15.33%	18.76%	15.66%
Radiation therapy	9	29	28	114	-92.11%	-74.56%	-75.44%
Hospital pharmacy	0	0	0	30	-100.00%	-100.00%	-100.00%
Health physics	0	0	0	0	0.00%	0.00%	0.00%
Anatomic pathology	76	81	61	124	-38.71%	-34.68%	-50.81%
Clinical biochemistry	0	0	0	0	0.00%	0.00%	0.00%
Microbiology and Virology	0	0	0	28	-100.00%	-100.00%	-100.00%
Clinical pathology	98	0	80	71	6.52%	-100.00%	8.11%
Pharmacology	0	0	0	0	0.00%	0.00%	0.00%
Medical genetics	0	0	0	0	0.00%	0.00%	0.00%
Food science	0	0	0	55	-100.00%	-100.00%	-100.00%
Anesthesiology and Intensive care medicine	496	729	598	591	-16.07%	23.35%	1.18%
Audiology	0	0	0	0	0.00%	0.00%	0.00%
Physical medicine and rehabilitation	125	100	98	141	-11.35%	-29.08%	-30.50%
Preventive healthcare	480	0	408	211	112.39%	-100.00%	93.36%
Occupational medicine	51	82	36	154	-66.88%	-46.75%	-76.62%
Forensic pathology	0	0	0	0	0.00%	0.00%	0.00%
Total	6561	6706	6678	6358			

Table 2.12: Cumulative allocation of regional and national grants per year (2012-2024) classified for general area according to the three demand scenarios (1 demographics, 2 ASA + SDO, 3 standard PL and ASA)

2.4 Conclusion

The Emilia-Romagna region will face huge changes in its HRH supply structure. The demographic mix of physicians employed both in the public and private sectors will turn into a massive retirement. Current training policies can therefore turn to be unsatisfactory to face such a turnover effect. In addition, population demographic trends will push higher stress on specialties that are related to the elderly population.

The proposed simulation-optimization tool gives a comprehensive overview of regional data availability and consequently defines the level of accuracy that can be reached by a quantitative approach to HRH regional planning. The main contribution of this approach is the definition of a systematic representation of Regional Healthcare that must be taken into account while planning training policies. Training decisions will be effective after five or six years and then different employment sectors can compete in order to hire trained physicians. In addition future perceived shortages can be tackled with a long term perspective considering that allocation policies over multiple years can improve the overall reduction of future gaps. Forecasts on supply and demand components as well as allocation weighting factors are based on a series of assumptions that have to be considered while analyzing model outputs for real policy definition. So far there was a lack of information management and of systematic representation of the regional health workforce sector. Through a proper analysis of available data sources we have described, to the best of our possibilities depending on the lack of proper databases, the work-life cycle of Public, Accredited and Private sectors defining the set of dynamics that influence health workforce. We have then identified three possible demand scenarios that can be used in order to evaluate future shortages. It is then clear that some specialties will face in the future a significant lack of workforce that must be solved through a proper re-modulation of both national and regional funding policies. The model suggests that Emilia-Romagna regional additional grants will not be able to cope with future system shortages and only an integrated management of both national and regional funded grants in ER training schools can satisfy future requirements of physicians.

Up to now training policies were based only on Public sector requirements that were mainly based on annual surveys with local health authorities. Those surveys only considered future imbalances based on current staffing level and future retirements. We included Private and Licensed employment areas by considering supply flows that can affect Public staffing level in the future. We also tried to capture future changes in service utilization by analyzing the trend-lines of the last decade and by linking those forecasts with future population evolution. It is clear from this study that the main problem with quantitative HRH modeling is related to available data sources. Full Time Equivalent representation is not possible due to unreliable public data and incomplete private ones. In addition, no information is given on workload stress of outpatient and inpatient activities. Unemployment or underemployment are other important factors that are not modeled at the moment and that can strongly influence future shortages information. It is possible to know how many physicians have been trained in recent years and by a proper record linkage how many of them are currently employed in ER but no information is available on trained and not employed ones. It is not possible to know if those specialists are currently unemployed or if they migrated to another region. Training autarchy is then another strong hypothesis of our model. We have described future emigration flows but we can't trust the amount of specialists that can be absorbed by non ER training schools. Then, it is clear that simulation-optimization model outputs can't be considered as prophetic forecasts because the set of information available at a regional level allowed us to give just an approximate representation of demand drivers. Finally the proposed approach assumes that current staffing levels are adequate to cope with their catchment areas in 2011 both in terms of row population to physician ratio (Scenario 1) and in terms of service utilization ratio (Scenario 2). It is clear then that a first improving step would be to create a unique and easily accessible database on regional HRH. Nevertheless, this is the first quantitative and systematic study and we believe it defines a fundamental step on the methodological evolution of this strategic healthcare management area.

2.5 Appendix

2.5.1 Graphs forecast legend

Stock of physicians working in the Public Sector in 2011 and still active

Stock of physicians working in the Private Hospital Sector in 2011 and still active

Stock of Public self-employed ambulatory specialists (Sumai) working in 2011 and still active

Stock of district pediatricians working in 2011 and still active

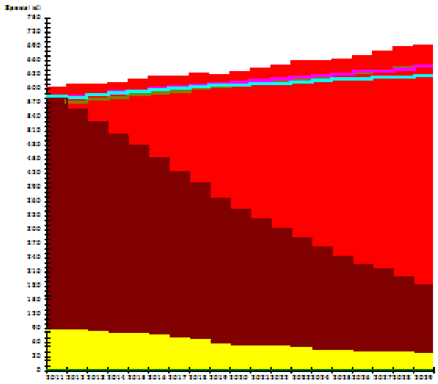
Stock of available physicians (Trained by MIUR + Employed in Public and Private sectors)

Demand forecast of Scenario1

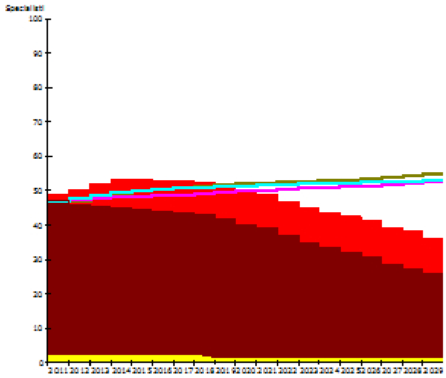
Demand forecast of Scenario2

Demand forecast of Scenario3

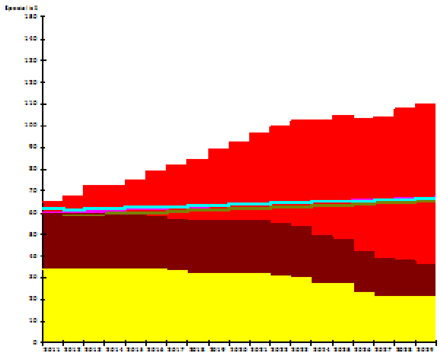
2.5.2 Surgical Area Forecasts



(A) General Surgery

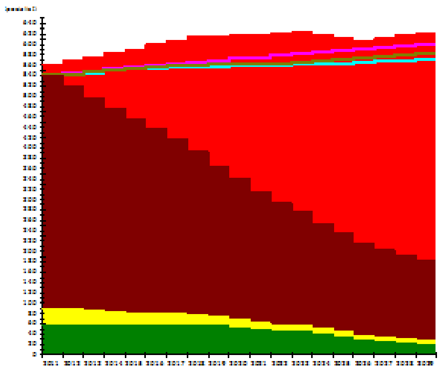


(B) Pediatric Surgery

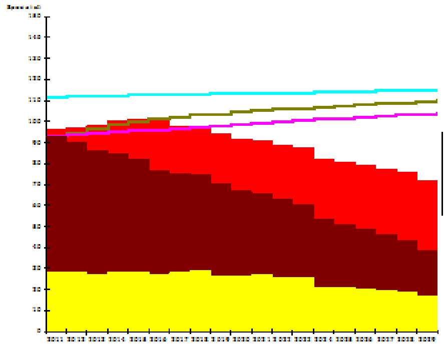


(C) Digestive Surgery

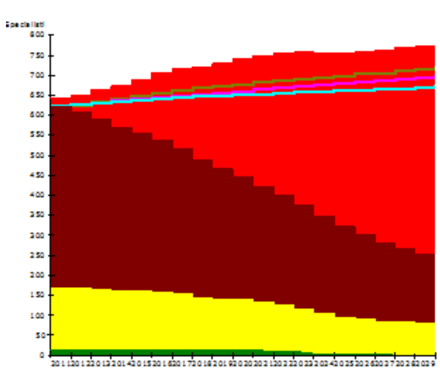
FIGURE 2.5: General Surgeries class



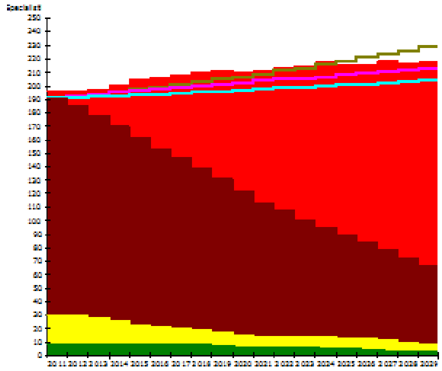
(A) Gynecology and Obstetrics



(B) Neurosurgery

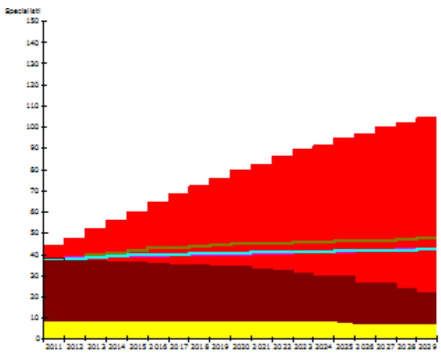


(C) Orthopedic and Traumatology

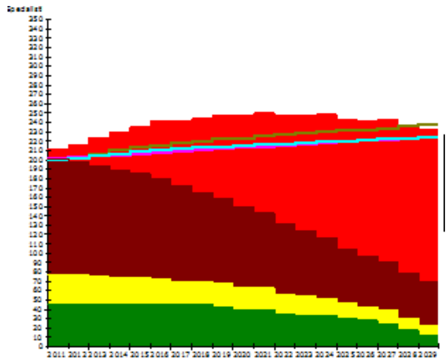


(D) Urology

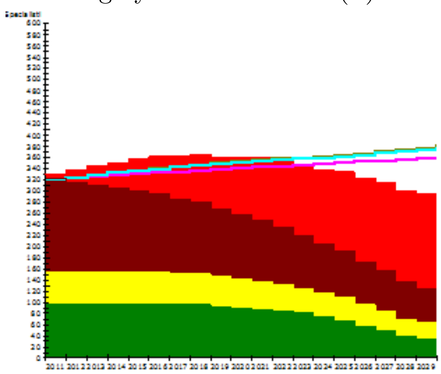
FIGURE 2.6: Specialistic Surgery class



(A) Maxillo-facial Surgery

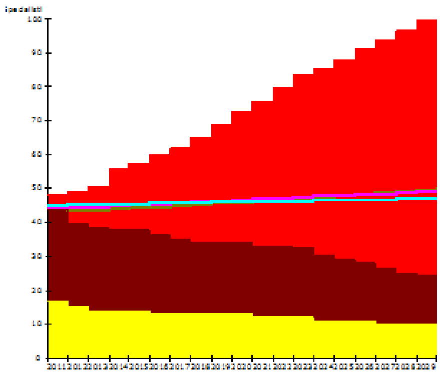


(B) Otolaryngology

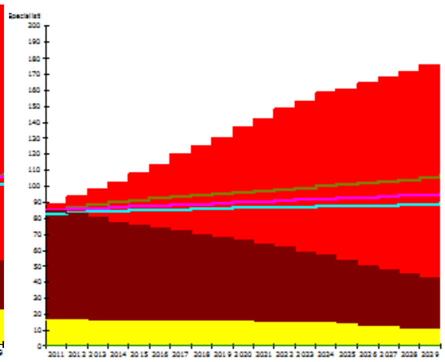


(C) Ophthalmology

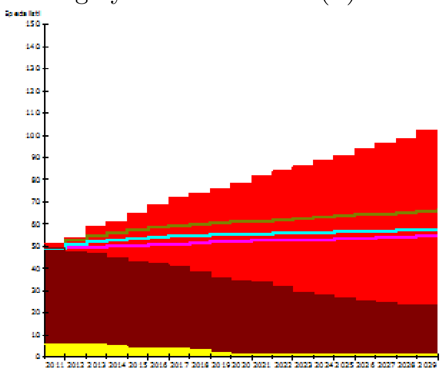
FIGURE 2.7: Head and Neck Surgery class



(A) Cardiac Surgery



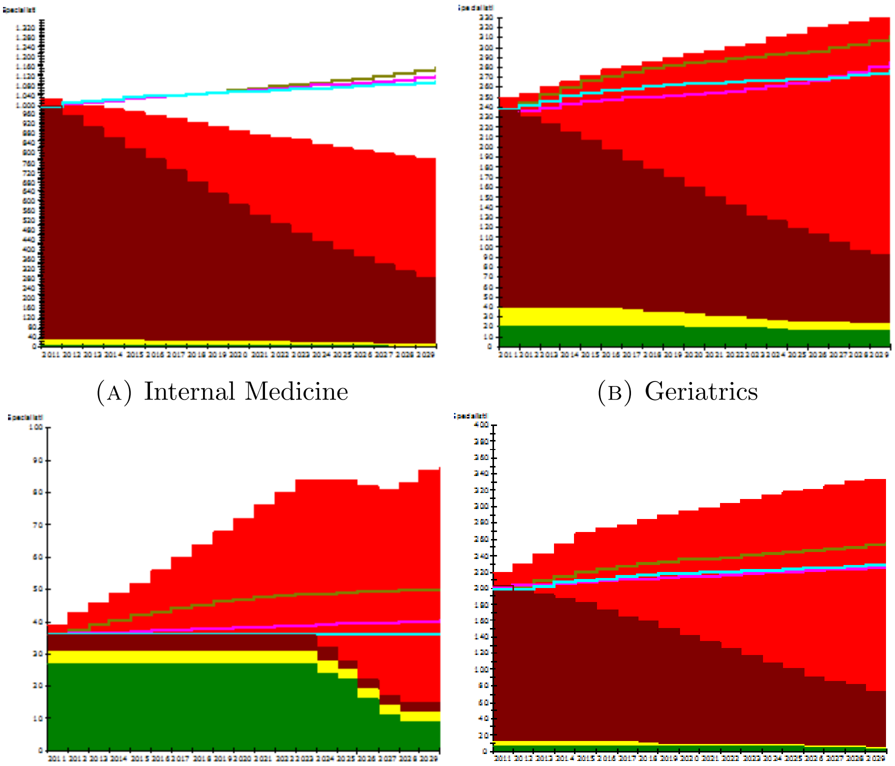
(B) Vascular Surgery



(C) Thoracic Surgery

FIGURE 2.8: Cardiac thoracic and vascular Surgery class

2.5.3 Medical Area Forecasts



(A) Internal Medicine

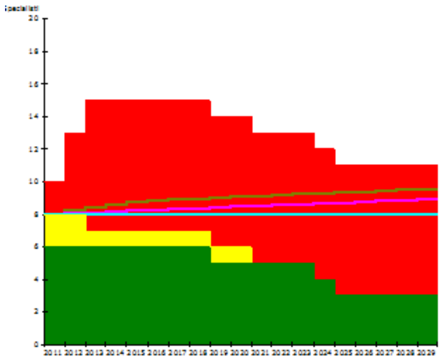
(B) Geriatrics

(C) Sports Medicine

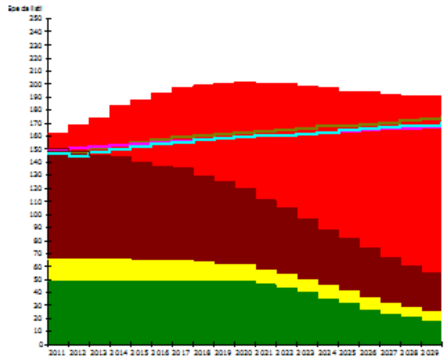
(D) Oncology

FIGURE 2.9: General Medicine class

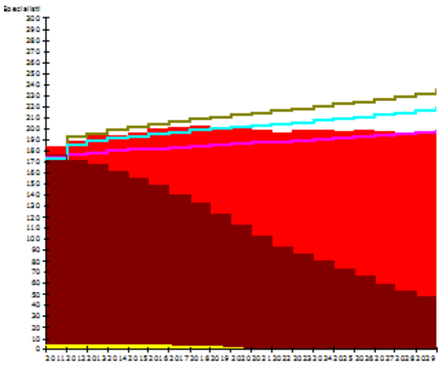
2.5.4 Diagnostic and Clinical Services Area Forecasts



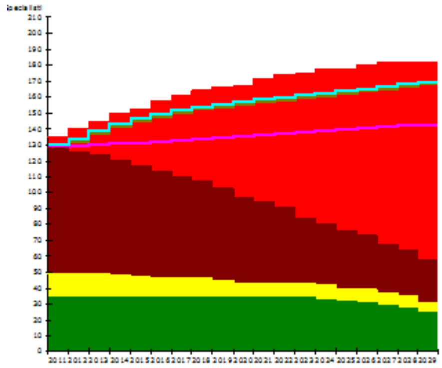
(A) Allergology and Immunology



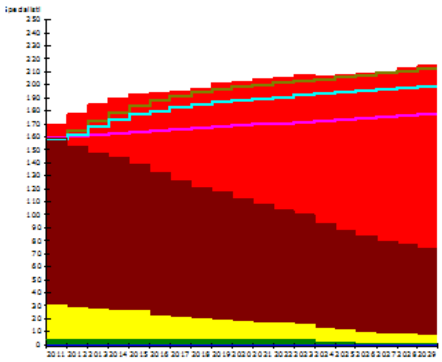
(B) Dermatology and Venereology



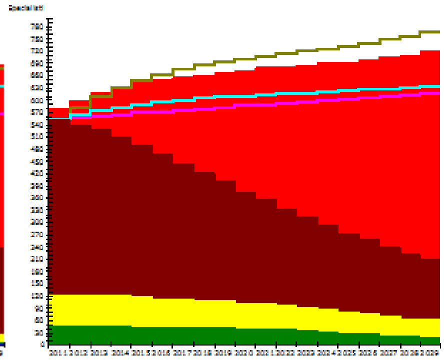
(C) Hematology



(D) Endocrinology

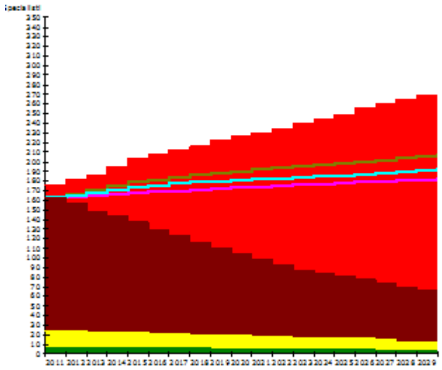


(E) Gastroenterology

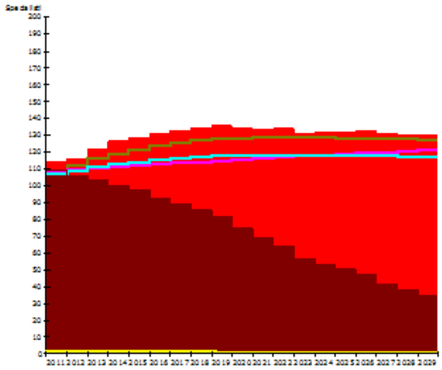


(F) Cardiovascular diseases

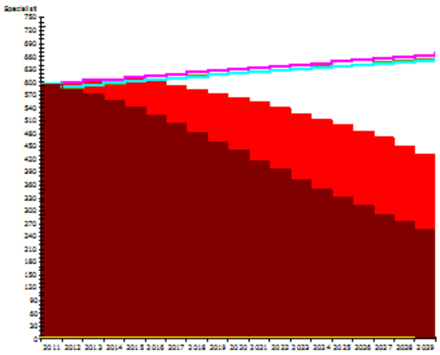
FIGURE 2.10: Specialistic Medical class



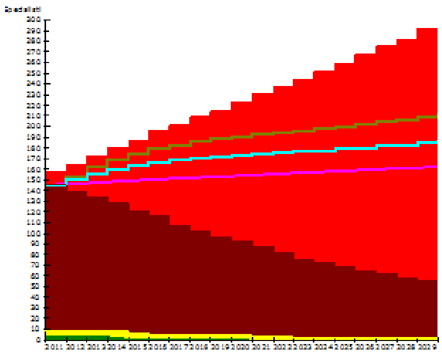
(A) Respiratory diseases



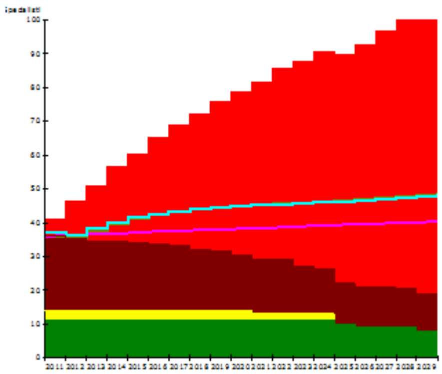
(B) Infectious diseases



(C) Emergency medicine

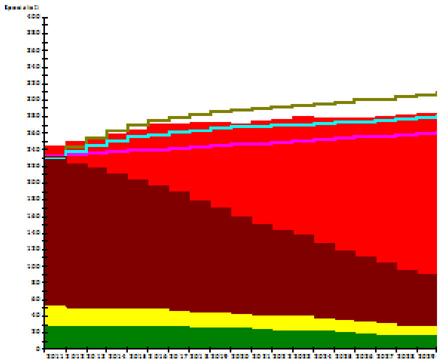


(D) Nephrology

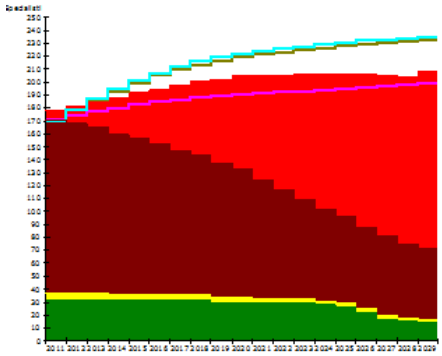


(E) Rheumatology

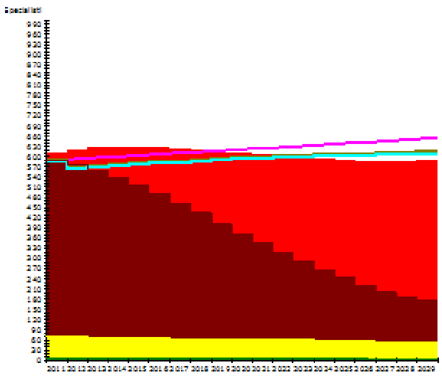
FIGURE 2.11: Specialistic Medical class



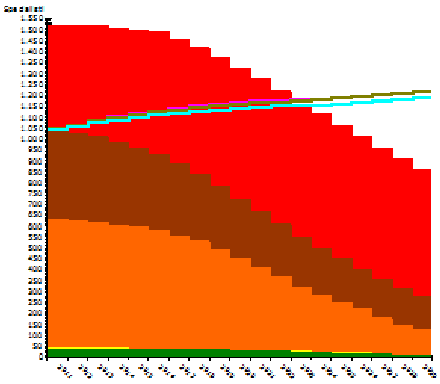
(A) Neurology



(B) Pediatric neuropsychiatry

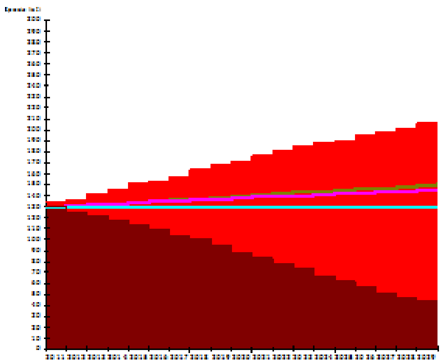


(C) Psychiatry

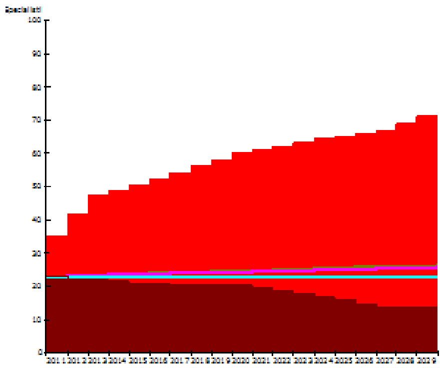


(D) Pediatrics

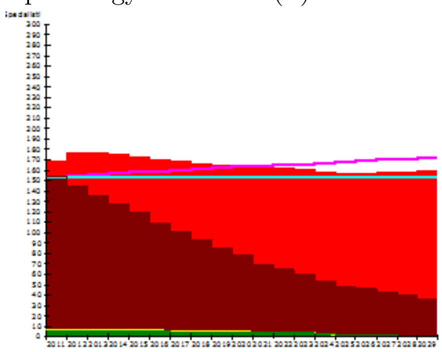
FIGURE 2.12: Neuroscience and Mental disorders class



(A) Anatomic pathology

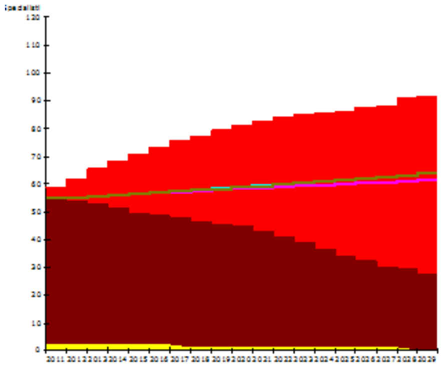


(B) Microbiology and Virology

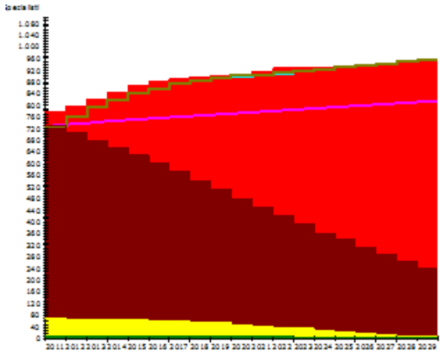


(C) Clinical pathology

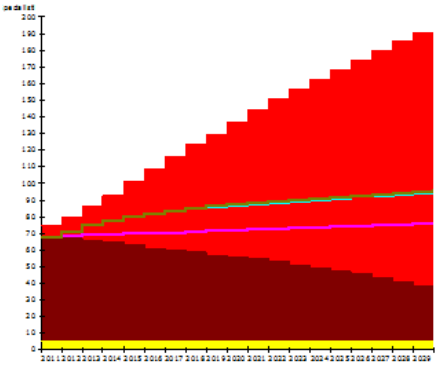
FIGURE 2.13: Therapeutic and diagnostic services class



(A) Nuclear medicine

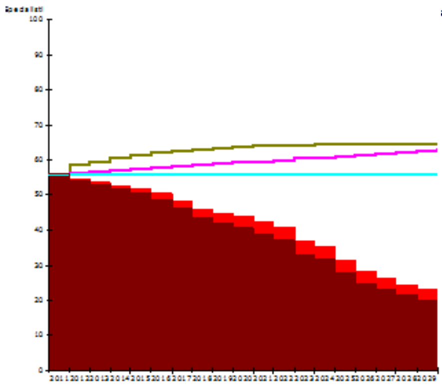


(B) Radiology

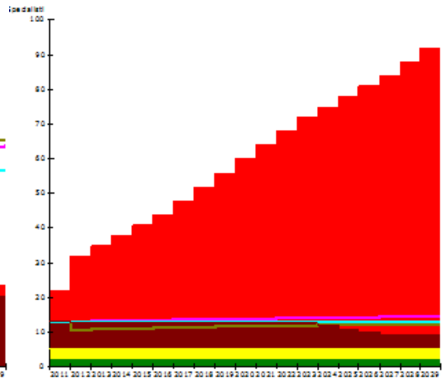


(C) Radiation therapy

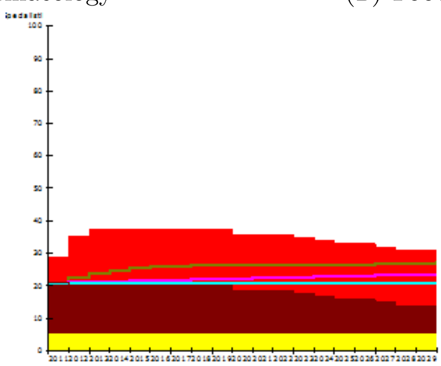
FIGURE 2.14: Medical imaging and Radiation therapy class



(A) Pharmacology

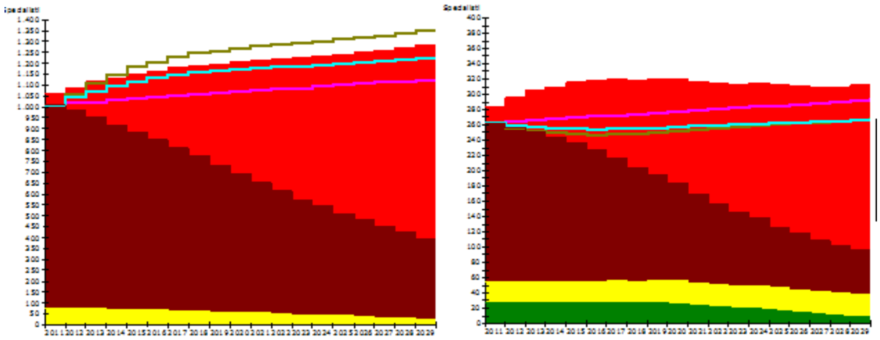


(B) Food science



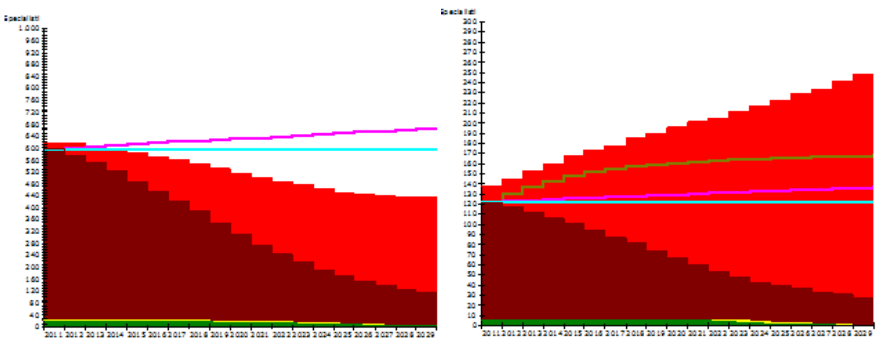
(C) Medical genetics

FIGURE 2.15: Biomedical clinical services class

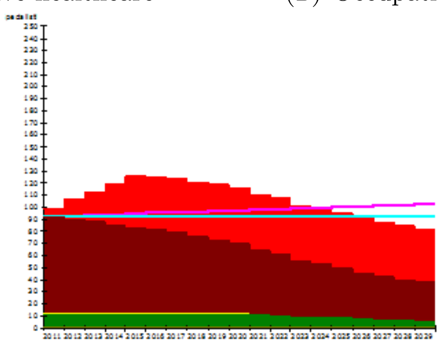


(A) Anesthesiology and Intensive care medicine (B) Physical medicine and rehabilitation

FIGURE 2.16: Specialistic clinical services class



(A) Preventive healthcare (B) Occupational medicine



(C) Forensic pathology

FIGURE 2.17: Public Health class

Bibliography

Physician workforce policy guidelines for the united states, 2000-2020. U.S. Department of Health & Human Services, 2005.

Green paper on the european workforce for health. Commission of the European Communities, 2008.

Models and tools for health workforce planning and projections. World Health Organization, 2010.

Patricia Barber and Beatriz Lopez-Valcarcel. Forecasting the need for medical specialists in spain: application of a system dynamics model. *Human Resources for Health*, 8(1):24, 2010.

S. Birch, G. Kephart, G. Tomblin Murphy, L. O'Brien-Pallas, R. Alder, and MacKenzie A. Human resources planning and the production of health: a needs-based analytical framework. *Canadian Public Policy 2007; XXXIII(Suppl.)*, pages S1–S16, 2007.

Stephen Birch, Linda O'Brien-Pallas, Chris Alksnis, Gail Tomblin Murphy, and Donna Thomson. Beyond demographic change in human resources planning: an extended framework and application to nursing. *Journal of Health Services Research and Policy*, 8(4):225–229, 2003.

Emmanuel Coste and Bui Dang Ha Doan. Forecast of physician workforce in the 22 french regions (1998-2013). *Cah Sociol Demogr Med*, 43(1):5–70, 2003.

BD Doan. Aging of population and medical workforce: a prospective view of health care provision in france in the year 2025. *Cahiers de sociologie et de demographie medicales*, 44(2):243–266, 2004.

S. Duckett. The Australian health workforce: facts and futures. *Aust Health Rev*, 23 (4):60–77, 2000.

P. Gavel. Medical workforce planning in australia: process, methodology and technical issues. *Cah Sociol Demogr Med*, 44(1), 2004.

L. Greenberg and J. M. Cultice. Forecasting the need for physicians in the united states: the health resources and services administration's physician requirements model. *Health Serv Res*, 31(6):723–737, Feb 1997.

Catherine M. Joyce and John J. McNeil. Participation in the workforce by australian medical graduates. *Medical Education*, 40(4):333–339, 2006. ISSN 1365-2923.

Andrea Lodi, Angelina Mazzochetti, Francesca Senese, and Paolo Tubertini. Forecast and fund allocation of medical specialty positions: Emilia-romagna region case study. Technical Report OR 14-5, LabOR DEI - University of Bologna, 2014.

J. G. Nooney and L. M. Lacey. Validating hrsa's nurse supply and demand models: a state-level perspective. *Nurs Econ*, 25(5):270–278, 2007.

Linda O'Brien-Pallas, Andrea Baumann, Gail Donner, Gail Tomblin Murphy, Jacquelyn Lochhaas-Gerlach, and Marcia Luba. Forecasting models for human resources in health care. *Journal of Advanced Nursing*, 33(1):120–129, 2001.

D. Persaud and L. Narine. Cross-national comparison of capitation funding: the american, british and dutch experience. *Health services management research : an official journal of the Association of University Programs in Health Administration / HSMC, AUPHA*, 12(2):121–135, 05 1999.

Dominique Roberfroid, Sabine Stordeur, Cecile Camberlin, Carine Van de Voorde, France Vrijens, and Christian Leonard. Physician workforce supply in belgium. current situation and challenges. *KCE Reports 72C*, 2007.

Dominique Roberfroid, Christian Leonard, and Sabine Stordeur. Physician supply forecast: better than peering in a crystal ball? *Human Resources for Health*, 7(1): 1–13, 2009.

Deborah Schofield. Replacing the projected retiring baby boomer nursing cohort 2001 - 2026. *BMC Health Services Research*, 7(1):87, 2007.

Gail Tomblin Murphy, George Kephart, Lynn Lethbridge, Linda O'Brien-Pallas, and Stephen Birch. Planning for what? challenging the assumptions of health human resources planning. *Health Policy*, 92:225 – 233, 2009.

Chapter 3

Tactical versus operational discrete event simulation: a Breast Screening case study [1]

3.1 Introduction

Since 1996, the Department of Public Health of the Emilia-Romagna region, based on national and international scientific community recommendations, provides a free screening program for early detection of breast cancer. The screening program offers scheduled checks to women, residents or domiciled in the region, falling in those age groups in which the risk of cancer increases the effectiveness of early diagnosis and appropriate treatment, reducing the risk of death. The monitoring is done through periodic checkups, namely mammography tests performed every two years. Local Health Units, which are coordinated and supervised by the Emilia-Romagna Regional Health Authority, are in charge of managing screening programs. The program is characterized by an integrated diagnostic-therapeutic pathway that follows the patient from the screening test up to surgery treatment and follow-up treatments. Until 2010 women in 50-69 age bands were the target population, in 2010 the Regional health Authority extended the breast screening program to the 45-49 and 70-74 age bands. The chapter is organized as follows. In Section 3.2 we report a literature review that analyze how breast screening programs have been studied by means of operational research techniques. In Section 3.3 we describe the Emilia-Romagna breast screening care pathway. Finally, in Section 3.4 we propose the experimental results for two discrete event simulation models a tactical and an operational one.

[1] This chapter is based on Technical Report OR 14-7 (see Lodi et al. [2014]) to appear in ECMS 2014, European conference on modeling and simulation Proceedings book

3.2 Literature review

In the literature Breast screening programs have been studied with simulation techniques considering several aspects. sMichaelson et al. [1999] use biologically based data from the literature on the rates of tumor growth and spread, to calculate the course of breast cancer growth and metastatic state to define the optimal screening interval for early detection of non degenerative breast cancers. In Fryback et al. [2006] the authors focus their attention on epidemiological aspects simulating 25 years of U.S. women population evolution addressing what-if questions about effectiveness of screening and treatment protocols, as well as estimating benefits to women of specific ages and screening histories. Improving health outcomes through effective diagnostic and treatment is certainly the overriding objective of screening programs, nevertheless an in depth evaluation of resources consumption and financial sustainability is very important to guarantee the success of the programs. In Brown and Fintor [1993] and Hunter et al. [2004] simulation studies evaluate the cost effectiveness of breast screening policies testing different scenarios regarding epidemiological trends, population age bands inclusion and possible outcomes and outputs both in terms of quality of care and of financial impact due to involved resources. Similarly the MISCAN (MIcrosimulation SCreening ANalysis) model, which uses Monte Carlo micro-simulation of a large number of life histories according to the epidemiology of the disease in question, has been used to model and test various breast screening issues in Italy (Paci et al. [1995]), Germany(Beemsterboer et al. [1994]) and Australia (Carter et al. [1993]). Surveys on cost effectiveness models regarding Breast screening programs can be found in Brown and Fintor [1993] and in Fone et al. [2003]. In addition to epidemiological and cost-effectiveness analysis some side aspects regarding screening programs have been investigated such as the impact of patients behavior on attendance rates in Brailsford and Schmidt [2003]. In conclusion screening programs can rely on an extensive set of guidelines and benchmarks that are used for performance monitoring. The quality of service can be evaluated according to health goals and organizational objectives. Indicators such as stage at diagnosis (defined as the ability to anticipate the detection of cancer care pathways and activate), quality of care (defined as the reduction of diagnostic errors) and 5-year survival rate after surgery treatment are primary objectives for health managers. Nevertheless screening programs have to be evaluated with an organizational and managerial perspective since, in order to provide services, a set of facilities and associated resources have to be identified. Future volume of activities and their financial sustainability as well as resources availability and waiting times can affect treatment effectiveness and are usually monitored and taken into consideration during planning activities.

3.3 Breast Screening: planning problem

As previously stated, in 2010 Emilia-Romagna Regional Health system decided to extend the breast screening program to 45-49 and 70-74 age bands. Each Local Health Authority, supported by the regional one, had to decide what resources should be resized even if at the time of the planning process it was not clear the impact of the extension of the screening coverage in terms of waiting time and lead time performances. The aim of this work is to study how two different DES software packages could have been effectively applied to support the process from a tactical and operational point of view. The case study is only focused on capacity planning since cost effectiveness analysis as well as screening frequency policies were already defined by strategic planners.

3.3.1 Breast Screening Pathway

Screening program can be generically described as a care pathway. In Naldoni et al. [2012] Emilia-Romagna regional guidelines and benchmarks are reported. It is possible to split the breast screening pathway in three main components: the first contact and appointment management, the first level examination and the second level examinations (see Figure 3.1 for pathway representation).

3.3.1.1 Invitation, appointment reschedule and reminder

Each woman falling in the target age band is invited every two years to undergo a screening test. The woman is invited by the Local Screening Center that communicates day and time of the appointment. If the candidate is unable to attend she can reschedule the appointment, otherwise she goes directly to the diagnostic ambulatory. In case of no shows, the candidate is contacted and a new appointment is planned. Contact activities are managed by the Screening Center, an organizational unit that works as an interface between screening candidates and program activities. This structure is responsible of breast, uterine and colon-rectum screening programs, therefore its operators are shared resources and segment their weekly activity in dedicated time slots for each program. The Screening Center is responsible of monthly supervision of invitations and no shows, appointment re-schedule and monthly reminder letter management.

3.3.1.2 First level examination

The day of the appointment the candidate undergoes a mammography test that is performed by a radiologist technician by using a Breast Computed Radiography Scanner.

After the screening test is performed, the recorded images are sent in a digital way to radiology senology specialists. When available a specialist analyzes the test and gives his/her diagnosis that can be negative, positive or uncertain. A test has to be analyzed at least by two different physicians and if both of them consider it negative, the patient will receive a letter confirming that no evidence of potential cancer was found. If at least one of the diagnosis is uncertain, a third physician analyzes the test and decides if the patient must undergo in-depth examinations. Even though the examinations are carried out at a local level, the analysis can be done in real time anywhere, since the images are remotely available in a digital format. In the ideal case it is possible that on the same day of the examination all diagnostic analysis are performed.

3.3.1.3 Second level examination

In depth examinations are performed for positive or uncertain patients. After a clinical examination the radiologist decides, depending on first level test results, if the patient should undergo detailed mammography, ultrasound or magnetic resonance imaging (MRI) examinations. If non-invasive examinations show the potential presence of a cancer, before proceeding with surgical activities, an invasive test such as cytology or micro-biopsy, is done in order to ascertain the presence of a tumor.

3.3.1.4 Key Performance Indexes

In order to monitor the organizational performances of the screening program two lead times are monitored, the time elapsed from mammography test to the dispatch of the letter with the first level negative result and the time elapsed from the mammography test to the first in-depth appointment. For each KPI two thresholds are monitored, first-level letter dispatch within 15 and 21 days and the first in-depth appointment within 21 and 28 days. Both lead time indicators measure radiology senology specialists and Screening Center performances.

3.3.2 Local Health Authority Data

Tactical versus operational planning has to take into account different levels of detail during the data collection process. Since screening programs are organized at a local level, we focused our case study on a regional local health authority. We collected data regarding volume of activities and resources involved in 2009 in order to build and validate a DES model. In 2009 the target population (women in the 50-69 age band) was equal to 51,462 residents and 24,111 of them were invited to attend the screening test. The 65.77% of invited women attended the test in the planned day, 15.29% called for a reschedule, 27.93% did not attend to the first appointment and

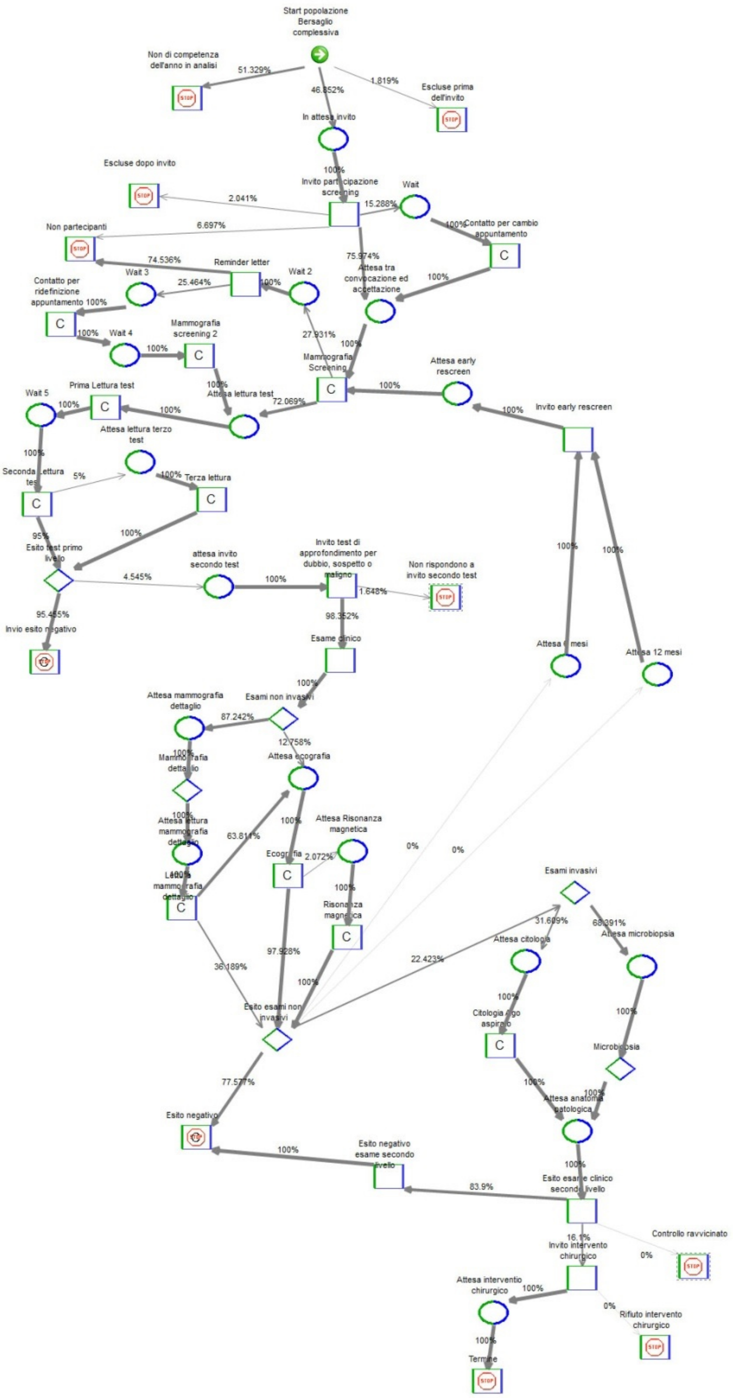


FIGURE 3.1: Simulation model of Breast Screening pathway implemented with Scenario Generator

74.54% of them did not answer to the reminder letter. Two screening center operators managed appointment agenda and need three minutes for each call. Screening tests were organized in eight public clinics (LOC) spread over the local health authority territory.

Resource ID	Number of working days per week	Number of mammograms per day	Working months per year
LOC 1	2	24	10
LOC 2	2	24	10
LOC 3	4	29	10
LOC 4	2	24	10
LOC 5	4	24	10
LOC 6	3	24	10
LOC 7	2	24	10
LOC 8	4	54	11

Table 3.1: First level examination resources available in 2009

Table 3.1 shows the yearly schedule of each clinic. First, analysis were partially outsourced, in 2009 30% of them were managed by an Hospital trust that guaranteed diagnostic results within 5 working days. Seven radiology senology specialists (RSS) working for the Local Health Authority managed the remaining 70% first analysis as well as all second and third ones. Monday, Tuesday and Thursday were dedicated to first level analysis and each mammogram test analysis took on average three minutes (see Table 3.2).

Screening Center manager dedicated on each day one hour and a half to send 140 negative result letters. In-depth examination were managed by Local Health Authority radiology senology specialists by following a fixed schedule, on Wednesday non-invasive examinations were performed invasive ones were planned on Friday.

Specialist ID	Months of activity (included holidays)	Number holidays during activity months (days)
RSS 1	12	20
RSS 2	10	40
RSS 3	3	10
RSS 4	12	52
RSS 5	3	17
RSS 6	4	23
RSS 7	0	365

Table 3.2: Radiology senology specialists available in 2009

Table 3.3 shows the workload of diagnostic sonography tests, magnetic resonance imaging tests, detailed mammogram procedures, micro biopsies and cytology test reading in terms of volume of activities and time required by each activity.

In addition to screening-driven activities, radiologists had to deliver a set of services associated with regular outpatient activities within the public sector (see Table 3.4).

Activity type	Number	Mean time (minutes)
Detailed mammogram procedure	677	15
Magnetic Resonance Imaging	373	20
Diagnostic sonography	373	15
Micro biopsies	119	30
Cytology test reading	55	30

Table 3.3: Second level examination 2009 workload

Activity type	Number	Mean time (minutes)
Computed Radiography	12.379	10
Magnetic Resonance Imaging	762	30
Diagnostic sonography	2.276	20
X-ray computed tomography	863	20

Table 3.4: Non-screening activities in 2009

As a result, in 2009 81.40% of first level negative results were sent within 15 days and 89.20% within 21 days, whereas 78.14% of first in-depth examinations were performed within 21 days and 84.88% within 28 days.

3.4 Simulation results

We implemented an operational model and a tactical one and we validated them on 2009 data. Below we present the results for both models.

3.4.1 Operational level model

We developed an operational model using Simul8 (Concannon et al. 2007), a general purpose DES software, defining radiology senology specialists, public clinics and Screening Center detailed schedules. Then, we tested the program extension impact under different system configurations. The target population in 2010 increased up to 80,289 women in 49-70 age band where 48,165 had to be invited. As a first planning hypothesis we fixed the rates of (i) invited women that attend the test in the planned day, (ii) no-shows after first invitation, (iii) appointment reschedules and (iv) no-shows after reminder letter (see Table 5 for detailed forecasted activity volumes). Then, we tested the performance worsening in case of no radiology senology specialists resizing and by considering that in 2010 the outsourcing contract was expired. As it is clear in Table 2, the real number of active radiology senology specialists in 2009 was less than the theoretical one. We then tested the system behavior under the hypothesis of five radiology senology specialists working full time for the Local Health Authority. That hypothesis holds since one of the seven physicians left in the first days of 2009 and the second one would have been pregnant during 2010. Observing past annual volume of holidays, we identified an average of 49 days off per year per physician. The proposed setting would have led, for 2010, to a 67.01% of first level negative results sent within 15 days and 73.25% within 21 days, while 63.04% of first in depth examination would have been performed within 21 days and 69.43% within 28 days.

As a second planning hypothesis, we considered the impact of increasing the number of radiology senology specialists available until near optimal performances were reached for both KPIs. It is important to say that the result is strongly influenced by the policy implemented for holidays. It is possible to reach a 98.76% of first level negative results sent within 15 days and 99.12% within 21 days with just one additional resource if physicians holidays never overlap. This would not have been the case even with two additional resources if holidays overlap. Two additional resources would have led just to a 76.94% of first level negative results sent within 15 days and 76.94% within 21 days.

Since 2010 data about real performances are available, we decided to test how the resizing proposed by the model considering 2009 population behavior would have been able to cope with real 2010 population behavior (see Table 3.5).

Activity type	2010 Forecast	2010 Real activities
Accepted after first invitation	31.679	28.909
No shows	13453	13.087
Examination after reminder letter	3.426	2.005
First level examination	35.104	30.914
Negative examination	33.509	29.473
Positive examination	1.595	1.43
Not responding to in depth examination	24	11
Responding to in depth examination	1.572	1.419
Detailed mammoraphies (DM)	1.371	1.238
Diagnostic sonographies	201	181
Diagnostic sonographies after DM	875	1.267
MRI tests	22	10
Invasive examinations	319	288

Table 3.5: 2010 forecasted and real volume of activities

The proposed setting would have led to a 97.93% of first level negative results sent within 15 days and 99.19% within 21 days, while 98.77% of first in-depth examinations would have been performed within 21 days and 99.13% within 28 days. These result show the proposed resizing would have been effective on the population behavior for 2010.

In addition to screening activities radiology senology specialists are also involved in general outpatients activities concerning detailed mammography, ultrasound or MRI examinations. In 2010 an increase in the demand of non-screening services was recorded (see Table 3.6) and we tested how that could have impacted on our proposed resource resizing.

Activity type	Number	Mean time (minutes)
Computed Radiography	19.68	10
Magnetic Resonance Imaging	413	30
Diagnostic sonography	8.003	20
X-ray computed tomography	1.536	20

TABLE 3.6: Recorded non-screening activities in 2010

The proposed setting would have led to a 13.75% of first level negative results sent within 15 days and 16.76% within 21 days while 14.48% of first in depth examination were performed within 21 days and 16.94% within 28 days, i.e., a dramatic worsening in performance. To face the increased volume of non-screening activities two additional radiology senology specialists should have been included by the Local Health Authority.

3.4.2 Tactical level model

In the previous section we analyzed how an operational model could have supported Local Health Authority planning. At a regional level one could be tempted to use less detailed information of the system at hands to do a more tactical planning (for example without resource daily schedules). Thus, we tested Scenario Generator (SG), a DES software customized for strategic decision planning, in order to show how it could have been used to test general guidelines regarding breast screening program extensions.

It is important to say that SG software was implemented in order to support long term strategic and tactical planning evaluation by public health managers. Because of this the implementation of new models and clinical pathways had to be very simple in order to ease the utilization to non simulation professionals. SG modeling is then very basic (lacking then of a detailed system modeling), it does not support the definition of daily resource schedules and resource capacity is mainly described by number of Full Time Equivalent (FTE) and minimum, average and maximum number of activities that each FTE can perform in a week. In our case it is then impossible to model the fact that some days of the week are dedicated to first level examinations and some others to in-depth invasive and non-invasive ones. Another modeling constraint is the lack of single entities labeling and management, as a consequence no distinction can be made between 2009 and 2010 screened women. Due to those restrictions we decided to test how SG can be used in order to provide a high level information to regional decision planners. We tested how the system would have behaved in 2010 if the theoretical number of radiology senology specialists working for the regional health authority in 2009 had not been changed. Because of SG restrictions we split physicians capacity in two components, first level examination test and second level non-invasive and invasive examinations (see Table 3.7).

Activity type	Activities per week per FTE
First level examination test	98
Second level examination	4

TABLE 3.7: Volume of activities provided in 2009 by a radiology senology specialists FTE

As a result we measured a stronger impact of program extension in terms of lead time worsening because the proposed setting would have led to a 45% of first level negative results sent within 15 days (see Figure 3.2) and 64% within 21 days.

A resource resizing up to 142 test per 9 radiology senology specialists would have been necessary in order to increase the performances up to 95% of first level negative results sent within 15 days and 97% within 21 days on average.

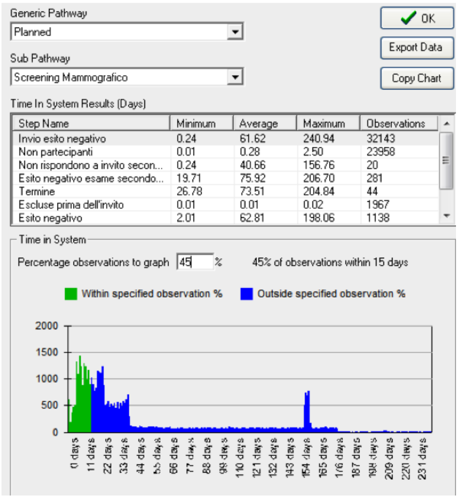


FIGURE 3.2: Scenario Generator 2009-2010 forecasts

3.5 Result interpretation and conclusion

We analyzed the screening program with two DES software packages in order to show how different tools can help stakeholders that operate at different decision levels. We applied Scenario Generator, an ad-hoc DES software for tactical planning for health systems, in order to develop an high level model that can be used to support in a quantitative way the definition of regional guidelines. We implemented a detailed DES model using Simul8, a general purpose software, in order to show to local health authority managers how a more detailed model can be used to understand the reasons of long lead times.

The Scenario Generator software acquired by the Regional Agency for health and Social Care of Emilia-Romagna can be used just to provide long term recommendations and to test guidelines implementation. Such recommendations will build upon available regional and local data and will allow the discussion around new services to be provided by the local trusts. A typical questions that could be answered by SG is the annual number of first level analysis that each radiologist should ensure in order to achieve organizational performance goal. Once defined general guidelines it is up to the Local Health Authority management to define if the number of first level test examinations is sustainable by the number of resources available.

It is clear that Scenario Generator software can not provide a detailed system forecast to Local Health Authority planners. The absence of territorial distinctions for mammography machines as well as the impossibility to describe radiologist weekly activities in a detailed way reveals the unfitness of SG as a tool to support operational planning. In order to better control activities in a weekly or annual time horizon is therefore advisable to use a tool like Simul8. In the proposed application we defined and tested several what-if scenarios and we measured their impact on measured screening lead times. We identified that resource resizing would have been strongly affected more by non-screening activities related to radiology senology specialists than by population rates of attendance to screening programs. We also evaluated how holiday policies could strongly affect the perception of shortages or surpluses in terms of available resources. That has been proven to be an interesting approach by health managers since during the breast screening planning extension it was not clear if and how much the increased number of radiology senology specialists would have been able to meet organizational goals. These results helped the Agenzia to asses the need of Decision Support Systems in general, and of operational planning discrete event simulation in particular.

Bibliography

Petra M. M. Beemsterboer, Harry J. De Koning, Peter G. Warmerdam, Rob Boer, Enno Swart, Marie-Luise Dierks, and Bernt-Peter Robra. Prediction of the effects and costs of breast-cancer screening in germany. *International Journal of Cancer*, 58(5):623–628, 1994. ISSN 1097-0215. doi: 10.1002/ijc.2910580502. URL http://dx.doi.org/10.1002/ijc.2910580502.

Sally Brailsford and Bernd Schmidt. Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. *European Journal of Operational Research*, 150(1):19 – 31, 2003. ISSN 0377-2217. doi: http://dx.doi.org/10.1016/S0377-2217(02)00778-6. URL http://www.sciencedirect.com/science/article/pii/S0377221702007786. O.R. Applied to Health Services.

Martin L. Brown and Lou Fintor. Cost-effectiveness of breast cancer screening: Preliminary results of a systematic review of the literature. *Breast Cancer Research and Treatment*, 25(2):113–118, 1993. ISSN 0167-6806. doi: 10.1007/BF00662136. URL http://dx.doi.org/10.1007/BF00662136.

Rob Carter, Paul Glasziou, Gerrit van Oortmarssen, Harry de Koning, Chris Stevenson, Glenn Salkeld, and Rob Boer. Cost-effectiveness of mammographic screening in australia. *Australian Journal of Public Health*, 17(1):42–50, 1993. ISSN 1753-6405. doi: 10.1111/j.1753-6405.1993.tb00103.x. URL http://dx.doi.org/10.1111/j.1753-6405.1993.tb00103.x.

David Fone, Sandra Hollinghurst, Mark Temple, Alison Round, Nathan Lester, Alison Weightman, Katherine Roberts, Edward Coyle, Gwyn Bevan, and Stephen Palmer. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health*, 25(4):325–335, 2003.

Dennis G. Fryback, Natasha K. Stout, Marjorie A. Rosenberg, Amy Trentham-Dietz, Vipat Kuruchittham, and Patrick L. Remington. Chapter 7: The wisconsin breast cancer epidemiology simulation model. *JNCI Monographs*, 2006(36):37–47, 2006. doi: 10.1093/jncimonographs/lgj007.

Duncan J.W. Hunter, Sean M. Drake, Samuel E.D. Shortt, John L. Dorland, and Ninh Tran. Simulation modeling of change to breast cancer detection age eligibility recommendations in ontario, 2002-2021. *Cancer Detection and Prevention*, 28(6):453 – 460, 2004. ISSN 0361-090X. doi: http://dx.doi.org/10.1016/j.cdp.2004.08.003. URL http://www.sciencedirect.com/science/article/pii/S0361090X04001217.

Andrea Lodi, Roberto Grilli, Francesca Senese, and Paolo Tubertini. Tactical versus operational discrete event simulation: a breast screening case study. Technical Report OR 14-7, LabOR DEI - Università di Bologna, 2014.

James S. Michaelson, Elkan Halpern, and Daniel B. Kopans. Breast cancer: Computer simulation method for estimating optimal intervals for screening. *Radiology*, 212(2): 551–560, 1999. doi: 10.1148/radiology.212.2.r99au49551. PMID: 10429717.

C. Naldoni, A. Finarelli, and R. Mignani. Protocollo diagnostico terapeutico dello screening per la diagnosi precoce dei tumori della mammella della regione emilia-romagna. *Contributi Servizio Sanità Pubblica Regione Emilia-Romagna*, (69), 2012.

E. Paci, R. Boer, M. Zappa, H. J. de Koning, G. J. van Oortmarssen, E. Crocetti, D. Giorgi, M. Rosselli del Turco, and J. D. Habbema. A model-based prediction of the impact on reduction in mortality by a breast cancer screening programme in the city of Florence, Italy. *Eur. J. Cancer*, 31A(3):348–353, 1995.

Chapter 4

Pre-operative activities and Operating Room planning: a local hospital department case study[1]

4.1 Introduction: Operating Theater Planning

The high costs of health care push national and regional health services as well as local authorities to constantly improve management performances in order to obtain more efficient organization of hospitals activities so as to provide patients with the best possible care. Operating Theater (OT) planning is one of the most studied topics for Operations Research applied to health systems management. Surgery departments play a crucial role in hospital organization and surgical activities usually constitute the core of efforts made by hospital managers in order to improve performances. Surgery activities represent a significant cost in the overall hospital budget because they directly involve expensive resources but also because they have an impact on other hospital services and on a multitude of resources indirectly associated. During the last decade the number of research studies that aim at efficiently organizing and planing surgical activities by reducing costs and keeping a good level of care has dramatically grow. A better planning and management of activities and resources directly and indirectly involved in surgical procedures (Operating Theater Planning) can therefore entail a more efficient use of resources, a reduction in waiting times for patients and a better overall performance of the hospital itself. In this chapter we consider the problem of assigning operating room time slots to surgeons in parallel with the definition of the admission planning of elective patients. The chapter is organized as follows.

[1]This chapter is based on Technical Report OR 14-6 (see Lodi and Tubertini [2014])

In Section 4.2 we report a literature review that analyze the various elements that has been considered in (OT) optimization models. In Section 4.3 we propose a Mixed Integer Programming (MIP) formulation based on Emilia-Romagna guidelines. In Section 4.4 we propose a reformulation of the problem tailored on a Local Hospital Orthopedic Department. Finally, Section 4.5 reports the computation results for real wold instances.

4.2 Literature review

Articles on operating theater management are characterized by a wide range of different criteria and approaches. The OT problem, in several articles, has been formulated as an optimization problem and solved with exact or heuristic procedures. Nevertheless a consistent branch of research applies simulation techniques in order to evaluate the effects of specific changes on performance indicators. Operating Theater planning and management can be analyzed over multiple decision making levels and classified by its decision-maker and time horizon characteristics.

4.2.1 Strategic Planning

Strategic planning decisions usually aim at defining the number of operating rooms that will characterize future hospital structures. Decisions concerning the amount of investments for the construction of new operating rooms as well as for technical equipment procurement and the prediction of specialists' need are characteristic elements of strategic high level planning. Furthermore, welfare policies defined by local or regional health management are discriminating factors for this kind of capacity planning activities. Uncertainty about performance impact of structural changes made by the local health system and the complex and uncertain nature of regional and local population needs (socio-demographic composition and related pathologies) usually requires the use of non deterministic decision support system tools.

4.2.2 Tactical planning

Strategic level decisions, despite having a strong impact on operational performance of the system, cannot be considered as the only planning decision that can be managed with quantitative tools. A proper management of available resources can, in some cases, avoid investments for structure upsizing. Once defined the number of available operating rooms it is necessary to define how operating room time slots should be assigned to Operating Units. The above mentioned problem is known as Master Surgical Schedule Problem (MSSP). Master Surgical Schedule is then defined as a cyclic program that determines the daily (or portion of the day) assignment of operating rooms to each surgery team. The MSSP is designed to support the negotiation phase with all Operating Units/Wards involved in order to ensure the most equitable possible solution. Equity and quality of a solution can be defined in various ways:

- Minimum gap between effective use of resources by the Operating Units and benchmarks defined by corporate policies;
- Maximum utilization of operating rooms with leveled bed utilization in order to reduce peaks that can turn into delay for surgeries;
- Minimum gap between required and assigned time for Operating Units.

Designing a weekly MSSP has to take into account restrictions related to surgeons availability, number of teams belonging to different surgical specialties (Testi et al. [2007]), operating rooms capacity and equipment (Chaabane et al. [2006]) and legal constraints imposing conditions on the maximum number of surgeon working hours (Guinet and Chaabane [2003]). As stated before MSSP is defined on a weekly basis and it is usually considered valid for a one year time horizon being cyclically applied rather than modified each week.

4.2.3 Operational Planning

Operating Theater operational planning has to identify, for each Operating Unit, which patients, among the ones in the waiting list, will undergo surgery during the week. This decision-making phase is characterized by short-term (weekly or daily) goals and it is usually called Admission Planning Problem (APP) or Surgical Case Assignment Problem (SCAP) if it includes tactical decisions regarding operating room time slot assignment to surgeons(Fei et al. [2008]). In the literature this problem has been solved with different approaches; as an example in Guinet and Chaabane [2003] and in Conforti et al. [2011] the authors define which subsets of patients have to be operated in each time block assigned to the surgery unity. Adan and Vissers [2002] manages operational assignment according to the partition of patients into different categories, determining how many patients, of each class, have to undergo surgery on each day in order to obtain the best patient's acceptance profile. Riise and Burke [2011] determine which patients will undergo surgery in each day by assigning patients to operating room time slots already dedicated to surgeries related to a specific specialty. Testi and Tanfani [2009] take into account patients priority according to their clinical status and the time spent on the waiting list. Finally, Vissers et al. [2005] evaluate the impact of operating room planning on pre and post surgery resource utilization. Admission Patients Scheduling is a problem of great complexity since it cannot simply be solved with a First-Come-First-Served logic, but it has to take into account clinical patients conditions and Operating Unit staff composition. In order to understand the element that really differentiates scheduling approaches we inspect the cohorts of patients that can be involved and the set of objectives that can influence operational planning decisions.

4.2.4 Patients and resources involved

A first distinction of Operating Theater problems can be driven by the set of patients that are considered during the planning phase. In the literature, two classes of patients are mentioned: elective patients and non-elective patients. We define an elective patient as a person who has to undergo a surgery that can be planned in advance (elective surgery). We define non-elective patients as those who need to urgently undergo surgery due to unexpected events and therefore that cannot be planned in advance. It is possible to split elective patients in two categories: inpatients, namely people that need to be hospitalized for at least one night, and outpatients that enter and leave the hospital on the same day of the surgery. Adan and Vissers [2002] develop a mixed integer programming model for scheduling elective operations, where outpatients are treated as inpatients requiring a hospital stay of only one day, the one in which the surgery takes place. Among non-elective surgeries, however, it is possible to distinguish urgent and emergency ones depending on the timeliness with which they have to be performed according to the patient clinical condition. An emergency indicates that a patient must be treated as soon as possible, while an urgency declares that a patient does not have to be treated immediately but it can wait up to a “short” period of time. Thus, elective surgeries can be planned in advance without any kind of uncertainty, that is the reason why most of the articles that apply optimization to OT planning take in consideration only this component of the problem. It is clear that non-elective surgeries are hardly predictable due to the stochastic behavior of urgencies and emergencies (see Denton et al. [2010] and Adan et al. [2009]).

Most of the research dealing only with elective surgeries planning, propose mathematical models in which a subset of the operating rooms are dedicated to non-elective interventions that as a consequence does not affect elective plans. Alternatively non-elective surgeries can be managed reserving them operating room time slots on each day. Wullink et al. [2007] studied, using a Discrete Event Simulation (DES) model, which is the best solution among the ones described above and concluded that the overall use of operating rooms improves significantly when the OT capacity dedicated to non-elective surgeries is distributed across operating rooms.

A further element of distinction among articles dealing with OT planning, is the variability associalted with various aspects of surgical services. Planning approaches are divided into deterministic and stochastic ones depending on the inclusion or not of variability aspects related to the surgical activity. Uncertainties considered in stochastic approaches are those related to uncertain arrival of non-elective patients or to non deterministic surgery time. The first condition is strictly related to the unpredictability of emergency or urgent case arrivals, while the second is determined by the fact that the real surgery time duration may not coincide with the one assumed during planning. Oostrum et al. [2008] propose a stochastic integer programming model with a set of probabilistic constraints that define that the total duration of all procedures performed in a day, in a certain operating room, must not exceed OT capacity, with a given probability.

4.2.5 Objective Function: performance indicators

Once decision levels have been identified it is import to analyze how combinatorial optimization approaches to Operating Theater management deal with the definition of a good solution of the problem. The final objective of these models is obviously to define the best possible management of operating rooms. In order to define if a given schedule or assignment is more efficient than others, it is necessary to identify performance indicators that have been considered in the literature. It is possible to identify two set of objectives that are taken into account. The first is characterized by those goals that aim at improving the service provided to patients by reducing waiting times and by taking in consideration patient need and priority related to their clinical status. Long waiting lists are one of the most common reported problem in health services. This results in a large quantity of studies that aim at reducing the waiting time of patients in order to increase their degree of satisfaction. This kind of objective can be identified in Tanfani and Testi [2010] and Riise and Burke [2011] papers, where the authors evaluate the overall efficiency of planning as a function of patients waiting time.

The second is focused on the improvement of hospital organizational efficiency. We define hospital organizational efficiency as the branch of research that aims at minimizing the costs associated to surgical activities maximizing the utilization of resources. In this context, Conforti et al. [2011] propose a model that aims at maximizing the number of patients who may be hospitalized. Oostrum et al. [2008] try to minimize the operating rooms' capacity by imposing that a predefined set of surgeries must be performed. This requirement aim at increasing OT productivity avoiding its under utilization by reducing inactivity times between two consecutive surgeries. Resource utilization is one of the most used performance indicators that focus on hospital organizational efficiency. Vissers et al. [2005] develop an integer linear programming model that minimize the deviation between the target utilization of resources (e.g. beds, nursing staff, etc.) and their actual use, keeping predefined objectives on number of surgeries that should be performed during the planning horizon. The decision to perform interventions in overtime may be more convenient in terms of waiting time reduction. On the other hand, keeping an operating room open can be very expensive and, for that reason, the number of planned interventions beyond the regular schedule should be kept under control (see Riise and Burke [2011] and Chaabane et al. [2006]). Resource leveling is another important goal for operating theaters management. For example, the prevention of peaks in resource utilization is taken in consideration in Oostrum et al. [2008].

4.2.6 Constraints and Resources

Among the resources directly involved in surgical procedures operating rooms are, of course, the most important ones. Chaabane et al. [2006] compare two methods of managing operating rooms time capacity. The authors define operating room time capacity as the maximum number of hours that can be worked each day on each operating room. Usually time bounds are between 8 AM and 8 PM. In the first model they manage operating room time as a fixed resource constraint in which, on each day, the sum of surgery hours assigned has to be less than or equal to the time capacity. In the second model, instead, Chaabane et al. [2006] define as a feasible assignment a schedule that might include some overtime with respect to classical operating room capacity. Bound on regular time and overtime operating room capacity are defined in order to forbid an overload of activities. Similar constraints can be found in the three stage approach of Testi et al. [2007]. Chaabane et al. [2006] manage surgical staff availability by requiring that, in each day, interventions are carried out, for each specialty, only if there is at least one surgeon available. In a similar way also Riise and Burke [2011], take into consideration surgeons availability. Riise and Burke [2011], as well as Testi et al. [2007], Testi and Tanfani [2009] and Tanfani and Testi [2010] define non-ubiquity constraints related to surgeons or to surgical staffing units, a non-ubiquity constraint is the maximum number of simultaneous activities that a resource can do. So far only resources directly related to surgery activities have been described. It is clear that there is a set of hospital resources that are indirectly influenced by Operating Theater planning decisions. Ward and Intensive Care Unit (ICU) beds are usually a hard constraint to deal with during OT planning. If no bed is available for after surgery treatments, then the patient cannot be scheduled. Tanfani and Testi [2010] restrict the number of patients that can undergo surgery in a day, according to the number of beds available in that same day both for ward and ICU case. Resources involved in planning activities can also depend on each patient pathology. Conforti et al. [2011] allow patients admission only when all examination can be planned in one week. Such kind of planning is bounded by the capacity of all clinical services. In addition to resource capacity constraints some models consider also political or legal constraints. Chaabane et al. [2006] impose lower and upper bounds on the number of time slots that can be assigned to each specialty. Guinet and Chaabane [2003] set a limit to the maximum number of hours that a surgeon can work in a day.

4.3 Operating Theater management in Emilia Romagna: a theoretical optimization model

Emilia-Romagna Region lacks of optimization-driven decision support systems for Operating Theater management. Most of local hospitals are equipped with data management tools that, at the best of our knowledge, are mainly used to record activities rather than to plan them. Furthermore, local databases sometimes lack of fundamental planning information such surgery duration since the adoption of data management system is in some cases not structured. It must be noted that the lack of data is mainly focused on organization information rather than on epidemiological or clinical ones. The literature review suggested that huge improvements can be made with the adoption of optimization planning tools. We then decided to design an Operating Theater Mixed-Integer Programming model that is tailored on Emilia-Romanga Regional guidelines and known best practices in order to propose this innovative approach to regional health hospitals. At first, we decided to define a theoretical model that takes into account most of regional guidelines and then to evaluate how it could be tailored on specific local case studies.

4.3.1 Decision variables and involved resources

Operating Theater planning in Emilia-Romagna is usually performed on a weekly basis and is defined from two to four weeks in advance. Surgery activities are usually classified as either elective or non-elective. Due to the lack of detailed data regarding non-elective surgeries we focused on elective planning. Planning activities, while having as a primary focus the definition of which patients should undergo surgery, has to take into account the availability of a set of resources that are necessary in order to define a feasible solution. Previous considerations suggested that two main approaches can be developed regarding Operating Theater planning: (i) Surgical Case Assignment or (ii) Master Surgical Scheduling plus Admission Planning. Analyzing Maurizio Bufalini Hospital guidelines we defined SCAP as the most suitable framework. A regional Operating Theater plan must then consider

- Operating rooms: time slots and surgical teams;
- Beds: Short-Term Care, Ordinary and Intensive Care;
- Preoperative outpatient medical examinations.

As far as operating room management concerns a feasible regional planning has to consider: (i) how many teams of each Operating Unit can work in parallel on the same day, (ii) if it is possible in each operating room for each day to perform surgery on patients coming from different Operating Units or characterized by different kind of surgeries and (iii) if the operating room is equipped with the facilities that are necessary in order to perform each class of surgeries.

Patients hospitalization is a topic of great importance since some hospital wards within regional territory have to deal with a quite limited number of beds for inpatients management. Even more, it is important to remind that national guidelines will impose in future years a general reduction in the number of beds available for each ward. So far, in Emilia-Romagna each Operating Unit has a set of reserved beds that are occupied by patients during, before and after surgery hospitalization periods. Pre and post-surgery hospital stays have to be considered since a patient before and after the day of surgery may need a bed for a time at least equal to the expected hospital stay related to his clinical condition. In addition to ward beds, the Intensive Care Unit have a number of beds that are shared between Operating Units in order to treat patients in critical conditions. It is clear that beds, being a limited resource, may constraint surgery scheduling. It may happen that an incorrect distribution of surgeries lead to a bed occupancy peak that prevents the admission of patients and forbids their planned surgery. Within Emilia-Romagna, bed stays can be classified as: (i) ordinary inpatient stay that can be extend beyond 5 days, (ii) short inpatient stay with a maximum length of 5 days and (iii) Day Hospital, which requires hospital admission only for the day of surgery. In general, Day Hospital beds are separated from inpatient ones.

Pre-operative management of outpatient anesthetic examination is an element of great importance while planning surgical interventions. Patients in waiting lists are characterize by an heterogeneous level of critical urgency, associated with their clinical condition. Each patient can then be classified with a degree of potential criticality that, in cases of major significance, can require a pre-operative anesthetic consult in addition to the routine one. This medical examination is associated with the possible, but not certain, need of in-depth medical examinations in order to give to the anesthesiologist a complete clinical picture of the patient so as to properly manage his peri-operative clinical pathway. It is clear that a proper management of schedule case mix is of major importance in order to define an efficient management of resources since it is possible that without a pre-operative anesthetic consult a sub-optimal use of available resources is achieved. This inefficiency is associated with the impossibility to perform some surgical procedures because some of the information necessary in order to preserve patient's health condition is missing. This situation can lead to surgery cancellations that negatively impact both patient care and hospital resource utilization since a time slot may not be used. Usually anesthesiologists define, for each Operating Unit, the maximum number of pre-operative outpatient medical examinations that can be planned and for critical patients for which it was not possible to book a medical examination, hospitalization is move up from one to four days in order to ease an informal and more direct management of anesthetist evaluation.

4.3.2 Planning objectives

Once defined the optimization problem and the set of constraints that will influence the definition of a feasible solution, it is important to identify which are the indicators to be used in order to evaluate the proposed solutions and to identify the most appropriate one. The concept of appropriateness in health care is not always uniquely defined and quantitatively measurable. We define two cohorts of objectives: patient-related quality of service and hospital-driven internal efficiency.

4.3.2.1 Patient-driven indicators

Quality of care and patient centered treatment are considered fundamental driving factors for the Emilia-Romagna public healthcare system and time-to-surgery can be considered as the most important planning factor. We define time-to-surgery as the lead time between patient referral and the day of surgery, where patient referral is defined as the moment in which, after diagnosis, the surgeon decides that the patient needs a surgery treatment. Time-to-surgery can then be classified as an indicator of health system responsiveness to patient needs. It is clear that this kind of information is strongly related with patient clinical status. If we analyze elective surgeries we can then define a classification of patients that is based on clinical condition and evaluation on patient pathology evolution. The clinical status can be defined according to the ASA (American Society of Anesthesiologists) classification:

- ASA Physical Status 1 - A normal healthy patient,
- ASA Physical Status 2 - A patient with mild systemic disease,
- ASA Physical Status 3 - A patient with severe systemic disease,
- ASA Physical Status 4 - A patient with severe systemic disease that is a constant threat to life, and
- ASA Physical Status 5 - A moribund patient who is not expected to survive without the operation.

At a Regional level priority classification is related to pathology evolution risk and to surgical procedure characteristics defining four degrees of priority:

- Grade 1: Small Surgery,
- Grade 2: Medium Surgery,
- Grade 3: Medium-High Surgery, and
- Grade 4: High and Very High Surgery.

Clinical patient classification, pathology evolution risk and surgery complexity are used in order to define four priority classes with related deadlines:

- High priority (A): Time to surgery has to be bounded to 30 days,
- Medium priority (B): Time to surgery has to be bounded to 90 days,
- Low priority (C): Time to surgery has to be bounded to 360 days, and
- Very low priority (D): Time to surgery has no bound and it can exceed 360 days.

The definition of priority can then be analyzed from two different perspectives. First, if two patients enter on the same day in the waiting list, the one with the highest level of priority has to undergo surgery before the other one. This statement is trivial and it can be explained by the fact that the first patient has a deadline that is closer in time with respect to that of the second patient. It is then important to define a function that describes how close is the patient to his deadline. The general idea of this function is that the more we get close to the deadline the more it will cost to postpone the surgery of the patient. By means of postponing we define the fact that the patient will not be inserted in the weekly plan. We assume that the function that models the deadline weight can be considered as constant on a weekly basis but exponential from week to week. That means that objective function weights $\pi_i^{\Delta_i}$ and $\pi_i^{\Delta_i+1}$ related to patient planned or unplanned admission in the week of planning can be defined as follows. Let Δ_i be the number of weeks that the patient spent in the waiting list, and DL_i the number of weeks that the patient can wait before reaching the deadline day according to his priority status. Let then define $Delay_i = \Delta_i - DL_i$ the number of elapsed weeks to deadline. If $Delay_i$ has a negative value this means that the deadline has not yet been reached while if the value of $Delay_i$ is greater than zero it means that the deadline is not respected and the patient has to be operated as soon as possible. The priority level can then be defined as $\pi_i^{\Delta_i} = |Delay_i|^{2Delay_i/|Delay_i|}$.

Figure 4.1 shows that parameter $\pi_i^{\Delta_i}$ is constant during a planning week and varies on a weekly basis exponentially increasing while the patient approaches the deadline week.

So far the priority coefficient distribution that we proposed does not take into account the fact that two patients with different priority levels are considered equal if their $Delay_i$ value is the same. We then decided to weight the $\pi_i^{\Delta_i}$ factor assuming that patients with equal deadline conditions have to be evaluated with respect to their priority level. We then modify $\pi_i^{\Delta_i}$ as $\pi_i^{\Delta_i} = \rho_i|Delay_i|^{2Delay_i/|Delay_i|}$, where ρ_i is the priority coefficient. An example of the impact of the proposed improvement can be seen in Figure 4.2a and 4.2b where Priority A and Priority B patients waiting factor trend line $\pi_i^{\Delta_i}$ is reported .

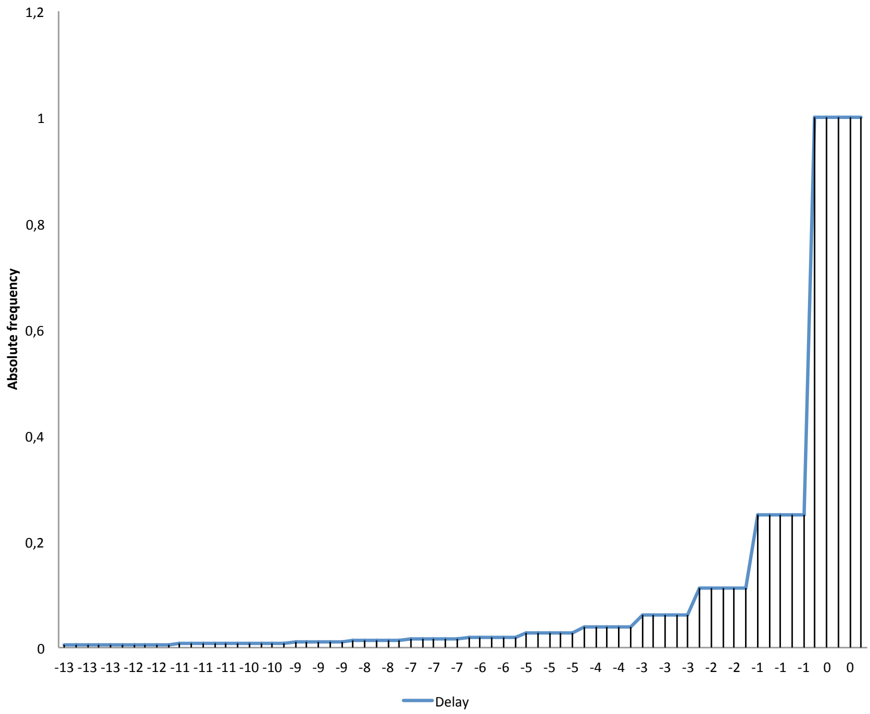
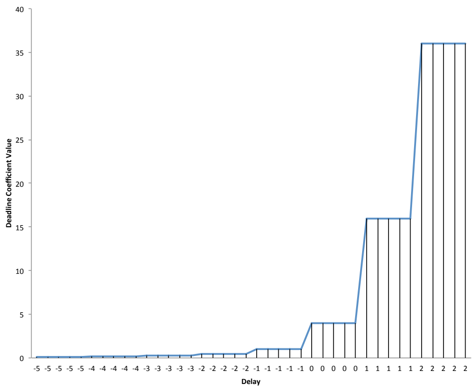
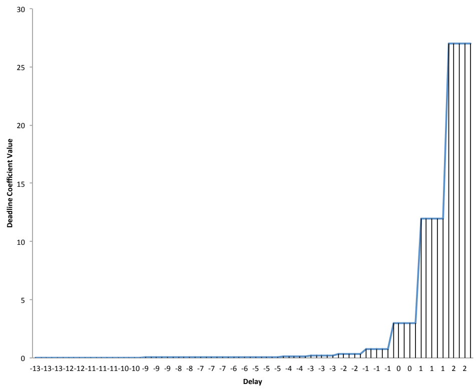


FIGURE 4.1: Example of Priority coefficient for 90 days deadline



(A) Priority coefficient for class A deadline



(B) Priority coefficient for class B deadline

FIGURE 4.2: Weighted trend line of Priority A and Priority B patients waiting factor

4.3.2.2 Hospital driven indicators

Hospital-driven efficiency indicators are related to an efficient resource management. We considered as a first element of organizational evaluation the number of overtime hours. Usually operating rooms are organized in two day sections, the morning and the afternoon one. The afternoon section ends at 8 PM even if it is possible that, as a consequence of very long waiting times, surgeries are planned during overtime hours. We then decided to model this behavior representing overtime hours as an element that negatively impact on the objective function but that can be accepted if a significant improvement in terms of waiting time reduction is achieved. Another important element of organizational efficiency is the planning of pre-operative appointments. As we said in Section 4.3.1 it is possible that if the required anesthetist appointment is not planned the patient must be hospitalized up to four days before the surgery day. This is clearly an inefficient management of pre-operative appointments that has to be reduced as much as possible by considering the number of unplanned appointments as a negative element that has to be weighted in the objective function.

4.3.3 Problem formulation

Let I be the set of patients inserted in waiting lists, K the set of operating rooms and T the set of days in planning time horizon. We consider as planing time horizon a period T equal to one month before the first surgery day and we define the last five working days of period T as the ones where surgery activities will be planned. We define working days as the ones available for assignment because planned surgeries are usually performed from Monday to Friday. Given a set K of operating rooms and a time horizon T, we define the assignment for the last week of the time horizon T of each operating room time slot to a surgery unit. The model we propose considers as a first decision variable the assignment of operating room time slots to Operating Units as follows:

$$y_{wkt} = \begin{cases} 1 & \text{if ward } w \text{ is assigned to operating room } k \text{ on day } t \\ 0 & \text{otherwise;} \end{cases}$$

Each operating room time slot assigned is then filled with a patient that is associated with the ward as follows:

$$x_{ikt} = \begin{cases} 1 & \text{if patient } i \text{ is assigned to operating room } k \text{ on day } t \\ 0 & \text{otherwise} \end{cases}$$

Since not all patients in the waiting list can undergo surgery during the week of planning we define

$$z_i = \begin{cases} 1 & \text{if patient } i \text{ will not undergo surgery on time horizon } T \\ 0 & \text{otherwise} \end{cases}$$

Pre-surgery appointments will be managed by variable

$$\varphi_{it} = \begin{cases} 1 & \text{if patient } i \text{ is examined by an anesthetist on day } t \\ 0 & \text{otherwise;} \end{cases}$$

In order to consider patient hospitalization as a consequence of missing anesthetist examination we define

$$\epsilon_{it} = \begin{cases} 1 & \text{if patient } i \text{ will undergo surgery on day } t\text{, and no anesthetist examination} \\ & \text{was planned} \\ 0 & \text{otherwise;} \end{cases}$$

Variable f_{kt} measures the overtime on operating room k on day t and variable η_{kt} the number of specialties assigned to k on day t exceeding the complexity threshold M (see parameters table). Finally, variable ψ_t measures the number of unused anesthetist appointments.

The complete list of parameters is then reported in Table 4.1

w	Operating Unit index such that $w = 1, 2, \ldots, m$ where m is the number of wards
I_w	Subset of patients that are in the waiting list of ward w
Δ_i	Waiting time of patient i
$\pi_i^{\Delta_i}$	Priority coefficient related to deadline proximity
ρ_i	Priority coefficient of patient i related to his clinical status
P_{WT}	Waiting time weight
F_{kt}	Maximum number of overtime hours planned for operating room k on day t
$Cover_{kt}$	Overtime cost of operating room k on day t
P_{OT}	Objective function overtime weight
P_{SP}	Objective function overlapping assignment
P_B	Objective function unplanned anesthetist appointments weight
P_{AV}	Objective function unused anesthetist appointments weight
M	Maximum number of specialties/Operating Units that can be assigned to the same operating room on the same time slot
$Csur_{kt}$	Cost related to exceeding number of different specialties assigned to the same time slot
$Cstay_w$	Cost of stay bed for a specific ward
p_i	Expected surgery time defined as a deterministic value
L_w	Minimum number of surgery hours to be assigned to ward w
U_w	Maximum number of surgery hours o be assigned to ward w
S_{kt}	Surgery time available on day t for operating room k
μ_i	Clinical condition indicator, equal to 1 if patient i is classified as critical and needs an anesthetist examination before surgery
h_i	Number of days that are suitable between anesthetist examination and surgery time in order to guarantee further in depth examinations if necessary
C_t	Number of anesthetist outpatient appointment slots available on day t
α_i	After surgery hospital stay days based on clinical condition and surgical treatment
I_{IC}	Subset of patients that, after surgery, need an Intensive Care Unit bed
δ_i	Forecasted after surgery stay time for patient i
G_w	Number of beds available for ward w
Q	Number of Intensive Care Unit beds
E_w	Number of surgical teams available for ward w
A_{wt}	Number of patients that undergo surgery in previous weeks but that are still hospitalized in ward w on day t
B_t	Number of patients that undergo surgery in previous weeks but that are still hospitalized in ICU on day t
ϑ_i	Takes value 1 if the patient i has to be hospitalized the day before surgery
N	"big M" value used to link x_{ikt} and y_{wkt} variables

TABLE 4.1: Theoretical Operating Theater model parameters

4.3.3.1 Mathematical formulation

The mathematical programming formulation reads as follows.

$$MinZ = P_{WT}\left(\sum_{i\in I}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}\rho_i\pi_i^{\Delta_i}x_{ikt} + \sum_{i\in I}\rho_i\pi_i^{\Delta_i+1}z_i\right) +$$

$$P_{OT}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}Cover_{kt}f_{kt} + P_{SP}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}Csur_{kt}\eta_{kt}+$$

$$P_B\sum_{w=1}^{m}\sum_{i\in I_w}\sum_{t=|T|-4}^{|T|}Cstay_w\epsilon_{it} + P_{AV}\sum_{t=1}^{|T|-5}\psi_t$$

$$\sum_{i\in I_w}x_{ikt} \le Ny_{wkt} \tag{4.1}$$

$$\forall k\in K;\quad \forall t=|T|-4,\ldots,|T|;\quad \forall w=1,2,\ldots,m$$

$$\sum_{k\in K}\sum_{t=|T|-4}^{|T|}x_{ikt}+z_i=1 \qquad \forall i\in I \tag{4.2}$$

$$\sum_{w=1}^{m}y_{wkt}\ge 1 \qquad \forall k\in K;\quad \forall t=|T|-4,\ldots,|T| \tag{4.3}$$

$$\sum_{w=1}^{m}y_{wkt}\le M+\eta_{kt} \qquad \forall k\in K;\quad \forall t=|T|-4,\ldots,|T| \tag{4.4}$$

$$\sum_{i\in I_w}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}x_{ikt}p_i\ge L_w \qquad \forall w=1,2,\ldots,m \tag{4.5}$$

$$\sum_{i\in I_w}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}x_{ikt}p_i\le U_w \qquad \forall w=1,2,\ldots,m \tag{4.6}$$

$$\sum_{i\in I}x_{ikt}p_i\le S_{kt}+f_{kt} \qquad \forall k\in K;\quad \forall t=|T|-4,\ldots,|T| \tag{4.7}$$

$$\sum_{k\in K}\mu_i x_{ikt}\le \sum_{a=1}^{t-h_i-1}\varphi_{ia}+\epsilon_{it} \qquad \forall i\in I;\quad \forall t=|T|-4,\ldots,|T| \tag{4.8}$$

$$\sum_{i\in I}\varphi_{it}+\psi_t=C_t \qquad \forall t=1,2,\ldots,|T|-5 \tag{4.9}$$

$$\sum_{i\in I_w}\left[\sum_{k\in K}\left(\sum_{\tau=t-\alpha_i}^{t} x_{ik\tau}+\vartheta_i x_{ik(t+1)}\right)+\sum_{\tau=t+1+\vartheta_i}^{t+4}\epsilon_{i\tau}\right]- \tag{4.10}$$

$$-\sum_{i\in I_w\cap I_{IC}}\sum_{k\in K}\sum_{\tau=t-\delta_i+1}^{t} x_{ik\tau}\leq G_w-A_{wt}$$

$$\forall t=|T|-5,\dots,|T|;\quad \forall w=1,2,\dots,m$$

$$\sum_{i\in I_{IC}}\sum_{k\in K}\sum_{\tau=t-\delta_i}^{t} x_{ik\tau}\leq Q-B_t\qquad \forall t=|T|-4,\dots,|T| \tag{4.11}$$

$$\sum_{k\in K} y_{wkt}\leq E_w\qquad \forall t=1,2,\dots,|T|\quad \forall w=1,2,\dots,m \tag{4.12}$$

$$x_{ikt}\in\{0,1\}\qquad \forall i\in I,\quad \forall k\in K,\quad \forall t=|T|-4,\dots,|T| \tag{4.13}$$

$$y_{wkt}\in\{0,1\}\qquad \forall w=1,2,\dots,m,\quad \forall k\in K,\quad \forall t=|T|-4,\dots,|T| \tag{4.14}$$

$$z_i\in\{0,1\}\qquad \forall i\in I \tag{4.15}$$

$$f_{kt}\in\{0,\dots,F_{kt}\}\quad \forall k\in K,\quad \forall t=|T|-4,\dots,|T| \tag{4.16}$$

$$\varphi_{it}\in\{0,1\}\qquad \forall i\in I,\quad \forall t=1,\dots,|T|-5 \tag{4.17}$$

$$\epsilon_{it}\in\{0,1\}\qquad \forall i\in I,\quad \forall t=|T|-4,\dots,|T| \tag{4.18}$$

$$\eta_{kt}\geq 0\quad \forall k\in K,\quad \forall t=|T|-4,\dots,|T| \tag{4.19}$$

$$\psi_t\geq 0\quad integer\quad \forall t=|T|-4,\dots,|T| \tag{4.20}$$

The objective function minimizes the waiting time of patients, weighted by priority indexes based on the patient clinical condition and associated deadline. As secondary objectives it is also minimized the total expected duration of planned interventions in overtime, the excess of surgical specialties associated with an operating room in a day, the number of unplanned anesthetist appointments. Constraints (4.1) define that if a patient is assigned to operating room k on day t it is necessary that this room is assigned on day t to the Operational Unit that is related with patient's specialty treatment. Despite this, the allocation of operating rooms to surgical specialties is also constrained by upper and lower bounds in (4.5) and (4.6). Constraints (4.3) ensure that each operating room will be assigned to at least one surgical specialty in each day of the planning week, while an excessive operating room sharing is bounded by M in (4.4) and penalized in the objective function. The number of working hours for each Operating Unit is bounded by constraints (4.5) and (4.6) within [L_w , U_w]. Constraints (4.7) model overtimes as the amount of time worked over regular time S_{kt}. Anesthetist medical examination planning is managed by constraints (4.8) and

(4.9). More precisely, if a critical patient is selected for surgery his pre-operative examination can be planned at most once due to constraints (4.8). Constraints 4.8 also impose that if a medical examination is planned for patient i, it must be performed at least h_i days before surgery. Parameter h_i is the time required to possibly plan in depth examinations. Constraints (4.9) defines maximum available examinations per day. Constraints (4.10) and (4.11) model beds and intensive care beds capacity constraints. Namely, constraints (4.10) ensure that a patients can undergo surgery only if he can be hospitalized. In order to verify bed availability on day t we consider available beds dedicated to surgical specialties and those already occupied. On day t beds are occupied both by patients in they pre- and post-surgery stay. Constraints (4.11) impose that patients, needing a post-operative ICU stay can undergo surgery only if a bed is available. Finally constraints (4.12) ensure that, for each specialty, at most E_w teams can work simultaneously.

4.4 Operating Theater management for a local hospital orthopedic department

Implementing an OT model requires the collection of a large set of data in order to feed model parameters and coefficients. Tests can be performed and analyzed using randomly generated or real-world data. Once the model based on the regional guidelines has been designed, we had the opportunity of testing it on real world data instances. An orthopedic department of a local health hospital agreed to collaborate with us in terms of model validation and data supply. The model presented and analyzed in section 4.3 has been then revisited in order to fulfill special requirements coming from the collaborating department. Since the revised model is strongly influenced by department characteristics and peculiarities it represents a specific problem that is not usually addressed in the literature.

The proposed problem is focused on department internal dynamics, so it does not require anymore the allocation of time slots to different surgical specialties. In fact, assignment of operating rooms among surgical units is considered in this case as a strategic decision that is taken by hospital managers and that cannot be modified at the operational level. The problem we analyze is then defined as the allocation of operating room time slots, among the ones already assigned to the department, to orthopedic surgeons and the subsequent admission plan for their related patients.

All surgeons working at the department share a common waiting list of patients that have been previously examined within the public health service. In addition each surgeon has a personal list of patients waiting for surgery. The final problem can then be defined as the determination of which operating room time slots will be assigned to each surgeon and which patients, belonging to shared or personal lists, must undergo surgery in the planning week. Orthopedic surgeries can be classified in elective and non-elective ones. Elective surgeries can be distinguished in non-prosthetic and prosthetic. Prosthetic surgeries consists of articulations replacement due to disease or congenital conditions and can be classified in hip replacement, knee replacement, and shoulder replacement. Discussing with department physicians we decided to overlook non-elective surgeries, also known as trauma surgeries, since trauma management is organized reserving on a daily basis separated operating room time slots.

4.4.1 Problem description: comparison with the theoretical model

The Orthopedic department has two reserved operating rooms in each day of the week for a total of 20 time slots available per week, 4 of which are reserved to trauma management. Each day is partitioned into two periods: a morning session of six hours from 8 AM to 2 PM and afternoon one of the same duration, from 2 PM to 8 PM. A feasible schedule dedicates to each specific surgeon one or more day periods during the planning horizon. An operating room, therefore, is occupied by a single surgeon for the duration of the assigned period. In addition, the Orthopedics department managers impose that, in each period, only one class of surgeries can be performed (prosthetic interventions, non-prosthetic or trauma). Finally, it is preferable that prosthetic surgeries are performed during the morning session. As already mentioned, the prosthetic interventions are classified as hip, knee or shoulder surgery. Each type is characterized by a distinct degree of complexity so it is preferable, for organizational reasons, to minimize the number of different type of prothesis surgeries that are performed in the same time slot. No forecast is possible about real surgery time so it is not possible to distinguish the impact of different surgeries on Operating Room Capacity. As a standard description we can consider that two surgeries per time slot can be planned. However, it is possible to plan on overtime an additional surgery just for non-prosthetic class ones. A further modification of the model is related to the request that a surgeon can be supported, for a type of surgery, by a colleague. The model then will have to satisfy these requests by preventing the assignment of an operating room to a surgeon, in case he has to assist a colleague.

Once activities that directly involve Operating Theater have been analyzed we evaluated which could be the set of resources that could indirectly impact on OT planning. Beds availability will not be considered in the model since the number of beds currently dedicated to the orthopedic department appears to be relatively abundant if compared to historical demand and therefore it does not constraint surgical activities. That is not the case for anesthetic appointment management that, according to department managers, is an element of great importance that is at the moment poorly planned. In a similar way autologous blood donation is a preoperative activity that must be planned in relation with surgical activities. Autologous blood donation is strongly promoted by national and regional guidelines since the planning of the blood drawing samples of patients who may have to undergo blood transfusions during surgery is important for several reasons. Autologous blood donation can lead to patient self-sufficiency in terms of blood needs during surgery and can also reduce the volume of requests to Hospital blood bank that has a limited number of blood bags shared by all the operating units in the hospital. Autologous blood donation is characterized by two blood samples the first that has to be organized at least three weeks before the surgery day and the second that must be taken at least one week before surgery.

4.4.2 Problem formulation

Let I be the set of patients inserted in waiting lists, K the set of operating rooms and T the set of days in planning time horizon. We consider as planing time horizon a period T equal to one month before the first surgery day and we define the last five working days of period T as the ones where surgery activities will be planned. We define working days as the ones available for assignment because planned surgeries are usually performed from Monday to Friday. Each working day is subdivided in P time slots. Given a set K of operating rooms and a time horizon T, we define the assignment for the last week of the time horizon T of each operating room time slot to a surgery unit. Let define Q the set of surgery cohorts, namely prosthetic and non-prosthetic surgeries. The model we propose considers as a first decision variable the assignment of operating room time slots to surgeons as follows:

$$g_{wktpq} = \begin{cases} 1 & \text{if time slot } p \text{ on day } t \text{ of operating room } k \text{ is assigned to} \\ & \text{surgeon } w \text{ for class } q \text{ surgeries} \\ 0 & \text{otherwise;} \end{cases}$$

In order to distinguish if more than one type of surgery is performed we define variable:

$$y_{wktpr} = \begin{cases} 1 & \text{if during time slot } p \text{ on day } t \text{ of operating room } k \\ & \text{a surgery of type } r \text{ will be performed by surgeon } w \\ 0 & \text{otherwise;} \end{cases}$$

Each operating unit time slot assigned is then filled with patients that are related to the surgeon waiting list as follows:

$$x_{iktp} = \begin{cases} 1 & \text{if patient } i \text{ is assigned to Operating room } k \text{ on day } t \\ & \text{in time slot } p \\ 0 & \text{otherwise;} \end{cases}$$

Since not all patients in the waiting list can undergo surgery during the week of planning we define:

$$z_i = \begin{cases} 1 & \text{if patient } i \text{ won't undergo surgery on time horizon } T \\ 0 & \text{otherwise;} \end{cases}$$

Pre-surgery anesthetist appointments will be managed by variables

$$\varphi_{it} = \begin{cases} 1 & \text{if patient } i \text{ is examined by an anesthetist on day } t \\ 0 & \text{otherwise;} \end{cases}$$

$$\epsilon_i = \begin{cases} 1 & \text{if patient } i \text{ will undergo surgery} \\ & \text{and no anesthetist examination is planned} \\ 0 & \text{otherwise;} \end{cases}$$

First pre-surgery autologous blood donation appointments will be managed by variable:

$$\sigma_{it} = \begin{cases} 1 & \text{if the first pre-surgery autologous blood donation appointment of} \\ & \text{patient } i \text{ is planned on day } t \\ 0 & \text{otherwise;} \end{cases}$$

Second pre-surgery autologous blood donation appointments will be managed by variable:

$$\lambda_{it} = \begin{cases} 1 & \text{if the second pre-surgery autologous blood donation appointment of} \\ & \text{patient } i \text{ is planned on day } t \\ 0 & \text{otherwise;} \end{cases}$$

$$\zeta_{i} = \begin{cases} 1 & \text{if patient } i\text{, will undergo surgery during the week of planning and} \\ & \text{no autologous blood donation is planned} \\ 0 & \text{otherwise;} \end{cases}$$

Variable f_{ktp} measures the overtime on operating room k on day t for time slot p and variable η_{kt} the number of specialties assigned to k on day t exceeding the complexity threshold M (see parameters table). Finally, variable ψ_t measures the number of unused anesthetist appointments and variable Φ_t the number of unused blood donation appointments.

The complete list of parameters is then reported in Table 4.2

w	surgeon index such that $w = 1, 2, \ldots, m$ where m is the number of surgeons
I_w	Subset of patients that are in the waiting list of surgeon w
I_{ward}	Subset of patients that are in the general waiting list of the Orthopedic Department
$Cohort_q$	Subset of surgery types that characterize $q \in Q$ class of surgeries (e.g. for q=1 three type of surgeries are defined, namely hip, knee and shoulder)
I_r	Subset of patients that need to undergo surgery type $r \in Cohort_q$
Δ_i	Waiting time of patient i
$\pi_i^{\Delta_i}$	Priority coefficient related to deadline proximity
ρ_i	priority coefficient of patient i related to his clinical status
P_{WT}	Waiting time weight
F_{kt}	Maximum number of overtime hours planned for operating room k on day t
$Cover_{kt}$	Overtime cost of operating room k on day t
P_{OT}	Objective function overtime weight
P_{SP}	Objective function overlapping assignment
P_B	Objective function unplanned anesthetist appointments weight
P_{AV}	Objective function unused anesthetist appointments weight
M	Maximum number of prosthetic surgery types that can be assigned to the same operating room on the same time slot
$Csur_{kt}$	Cost related to exceeding number of different specialties assigned to the same time slot
L_w	Minimum number of surgery hours to be assigned to surgeon w
U_w	Maximum number of surgery hours o be assigned to surgeon w
L_{ward}	Minimum number of patients coming from the ward waiting list that should be planned in the week
U_{ward}	Maximum number of patients coming from the ward waiting list that should be planned in the week
S_{kt}	Surgery time available on day t for operating room k
$trauma_{ktp}$	Equal to 1 if on day t operating room k during time slot p is dedicated to trauma surgeries
μ_i	Clinical condition indicator, equal to 1 if patient i is classified as critical and needs an anesthetist examination before surgery
h_i	Number of days that are suitable between anesthetist examination and surgery time in order to guarantee further in depth examinations if necessary
C_t	Number of anesthetist outpatient appointment slots available on day t
ν_i	Clinical condition indicator, equal to 1 if patient i is classified as suitable for autologous blood donation
$Prel_t$	Number of autologous blood donation outpatient appointment slots available on day t
N	"big M" value used to link x_{iktp} and y_{wktpr} variables
D	Set of supporting surgeries define as follows: $d =< w_1, w_2, r_d >$ where r_d is the type of surgery for whichw_1 needs the support of surgeon w_2

TABLE 4.2: Orthopedic Department Operating Theater model parameters

4.4.2.1 Mathematical formulation

The mathematical programming formulation reads as follows.

$$MinZ = P_{WT}\left(\sum_{i\in I}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}\sum_{p=1}^{P}\rho_i\pi_i^{\Delta_i}x_{iktp} + \sum_{i\in I}\rho_i\pi_i^{\Delta_i+1}z_i\right)+$$

$$+P_{OT}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}Cover_{kt}\sum_{p=1}^{P}f_{ktp} + P_{SP}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}Csur_{kt}\eta_{kt}+$$

$$+P_V\sum_{i\in I}\epsilon_i + P_{AV}\sum_{t=1}^{|T|-5}\psi_t + P_{BS}\sum_{i\in I}\zeta_i + P_B\sum_{t=1}^{|T|-5}\Phi_t$$

$$\sum_{i\in I_w\cap I_r} x_{iktp} \leq Ny_{wktpr} \tag{4.21}$$

$$\forall k\in K;\quad \forall t=|T|-4,\ldots,|T|;\quad \forall w=1,2,\ldots,m;\quad \forall p\in P;\quad \forall r\in Cohort_q;\quad \forall q\in Q$$

$$\sum_{i\in I_{Ward}\cap I_r} x_{iktp} \leq N\sum_{w=1}^{m} y_{wktpr} \tag{4.22}$$

$$\forall k\in K;\quad \forall t=|T|-4,\ldots,|T|;\quad \forall p\in P;\quad \forall r\in Cohort_q;\quad \forall q\in Q$$

$$\sum_{i\in Ward}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}\sum_{p\in P} x_{iktp} \geq L_{ward} \tag{4.23}$$

$$\forall w=1,2,\ldots,m$$

$$\sum_{i\in Ward}\sum_{k\in K}\sum_{t=|T|-4}^{|T|}\sum_{p\in P} x_{iktp} \leq U_{ward} \tag{4.24}$$

$$\forall w=1,2,\ldots,m$$

$$\sum_{k\in K}\sum_{t=|T|-4}^{|T|}\sum_{p\in P} x_{iktp} + z_i = 1 \qquad \forall i\in I \tag{4.25}$$

$$\sum_{r \in Cohort_q} y_{wktpr} \leq S_{ktp} g_{wktpq} \tag{4.26}$$

$$\forall w = 1, 2, \dots, m; \quad \forall k \in K; \quad \forall t = |T| - 4, \dots, |T|; \quad \forall p \in P; \quad \forall q \in Q$$

$$\sum_{w=1}^{m} \sum_{q \in Q} g_{wktpq} = 1 - trauma_{ktp} \tag{4.27}$$

$$\forall k \in K; \quad \forall t = |T| - 4, \dots, |T|; \quad \forall p \in P$$

$$\sum_{w=1}^{m} \sum_{r \in Cohort_1} y_{wkt1r} \leq M + \eta_{kt} \tag{4.28}$$

$$\forall k \in K; \quad \forall t = |T| - 4, \dots, |T|$$

$$\sum_{w=1}^{m} \sum_{r \in Cohort_1} y_{wkt2r} = 0 \tag{4.29}$$

$$\forall k \in K; \quad \forall t = |T| - 4, \dots, |T|$$

$$\sum_{k \in K} \sum_{t=|T|-4}^{|T|} \sum_{p \in P} \sum_{q \in Q} g_{wktpq} \geq L_w \tag{4.30}$$

$$\forall w = 1, 2, \dots, m$$

$$\sum_{k \in K} \sum_{t=|T|-4}^{|T|} \sum_{p \in P} \sum_{q \in Q} g_{wktpq} \leq U_w \tag{4.31}$$

$$\forall w = 1, 2, \dots, m$$

$$\sum_{i \in I_r} x_{iktp} \leq S_{ktp} \tag{4.32}$$

$$\forall k \in K; \quad \forall t = |T| - 4, \dots, |T|; \quad \forall p \in P; \quad \forall r \in Cohort_1$$

$$\sum_{i \in I_r} x_{iktp} \leq S_{ktp} + f_{ktp} \tag{4.33}$$

$$\forall k \in K; \quad \forall t = |T| - 4, \dots, |T|; \quad \forall p \in P; \quad \forall r \in Cohort_2$$

$$\sum_{t=1}^{|T|-5} \varphi_{it} + \epsilon_i = \mu_i (1 - z_i) \tag{4.34}$$

$$\forall i \in I$$

$$\sum_{k\in K}\sum_{p\in P}\mu_i x_{iktp} \leq \sum_{a=1}^{t-h_i-1}\varphi_{ia} + \epsilon_i \tag{4.35}$$

$$\forall i \in I; \quad \forall t = |T|-4,\dots,|T|$$

$$\sum_{i\in I}\varphi_{it} + \psi_t = C_t \tag{4.36}$$

$$\forall t = 1,2,\dots,|T|-5$$

$$\sum_{t=1}^{|T|-5}\sigma_{it} + \zeta_i = \nu_i(1-z_i) \tag{4.37}$$

$$\forall i \in I$$

$$\sum_{k\in K}\sum_{p\in P}\nu_i x_{iktp} \leq \sum_{a=1}^{t-19}\sigma_{ia} + \zeta_i \tag{4.38}$$

$$\forall i \in I; \quad \forall t = |T|-4,\dots,|T|$$

$$\sum_{t=1}^{|T|-5}\lambda_{it} + \zeta_i = \nu_i(1-z_i) \tag{4.39}$$

$$\forall i \in I$$

$$\sum_{k\in K}\sum_{p\in P}\nu_i x_{iktp} - \zeta_i \leq \sum_{a=t-14}^{t-6}\lambda_{ia} \tag{4.40}$$

$$\forall i \in I; \quad \forall t = |T|-4,\dots,|T|$$

$$\sum_{i\in I}(\sigma_{it} + \lambda_{it}) + \Phi_t = Prel_t \tag{4.41}$$

$$\forall t = 1,2,\dots,|T|-5$$

$$\sum_{k\in K}\sum_{q\in Q} g_{wktpq} \leq E_w \tag{4.42}$$

$$\forall w = 1,2,\dots,m; \quad \forall t = |T|-4,\dots,|T|; \quad \forall p \in P$$

$$\sum_{k\in K}(y_{w_1 ktpr_d} + \sum_{q\in Q}\sum_{r\in Cohort_q} y_{w_2 ktpr}) \leq 1 \tag{4.43}$$

$$\forall d \in D : d =< w_1, w_2, r_d >; \quad \forall t = |T|-4,\dots,|T|; \quad \forall p \in P$$

$$x_{iktp} \in \{0,1\} \tag{4.44}$$

$$\forall i \in I, \quad \forall t = |T|-4,\dots,|T|, \quad \forall p \in P, \forall k \in K$$

$$y_{wktpr} \in \{0,1\} \tag{4.45}$$

$$\forall w = 1,2,\dots,m, \quad \forall t = |T|-4,\dots,|T|, \quad \forall p \in P, \forall k \in K, \quad \forall r \in Cohort_q, \quad \forall q \in Q$$

$$g_{wktpq} \in \{0,1\} \tag{4.46}$$

$$\forall w = 1,2,\dots,m, \quad \forall t = |T|-4,\dots,|T|, \quad \forall p \in P, \forall k \in K, \quad \forall q \in Q$$

$$z_i \in \{0,1\} \quad \forall i \in I \tag{4.47}$$

$$\varphi_{it} \in \{0,1\} \quad \forall i \in I, \quad \forall t = 1,\dots,|T|-5 \tag{4.48}$$

$$\psi_t \in \{0,\dots,C_t\} \quad \forall t = 1,\dots,|T|-5 \tag{4.49}$$

$$\epsilon_i \in \{0,1\} \quad \forall i \in I \tag{4.50}$$

$$\zeta_i \in \{0,1\} \quad \forall i \in I \tag{4.51}$$

$$\sigma_{it} \in \{0,1\} \quad \forall i \in I, \quad \forall t = 1,\dots,|T|-5 \tag{4.52}$$

$$\lambda_{it} \in \{0,1\} \quad \forall i \in I, \quad \forall t = 1,\dots,|T|-5 \tag{4.53}$$

$$\Phi_t \in \{0,\dots,Prel_t\} \quad \forall t = 1,\dots,|T|-5 \tag{4.54}$$

$$\eta_{kt} \geq 0 \quad \forall t = |T|-4,\dots,|T|, \forall k \in K \tag{4.55}$$

$$f_{ktp} \in \{0,\dots,F_{ktp}\} \quad \forall t = |T|-4,\dots,|T|, \quad \forall p \in P, \forall k \in K \tag{4.56}$$

The objective function minimizes the waiting time of patients, weighted by priority-based clinical condition coefficients, and the penalty associated with exceeding deadline associated with each patient. In addition, the objective function takes into account the number of non-prosthetic interventions planned in overtime and the number of different types of prosthetic interventions, scheduled during the same period in one operating room. Finally, the objective function minimizes the number of unplanned medical anesthetist appointments and blood samples, for patients defined as critical or suitable for autologous blood donation. Constraint (4.21) are used to link the binary variable y_{wktpr} to the value assumed by x_{iktp} by ensuring that, if a patient surgery is planned for a specific time slot, then the operating room k is assigned to his related surgeon w on day t, slot p for the type of surgeries related to patient's pathology r. Similarly, constraints (4.22) ensure that, if the patient belongs to the ward waiting list, he can be assigned to every time slot assigned to every surgeon if it fits with the type of surgeries that will be performed in the time slot. The maximum and minimum number of surgeries that it is possible to run during each week is defined by constraints (4.23) and (4.24) for each surgeon depending on Department internal organization. Constraints (4.25) ensure, for each patient, that he can be either planned in a specific day, operating room and time slot or his surgery is postponed. Constraints (4.26) activate variable g_{wktpq} as a function of the value assumed by y_{wktpr}. In particular, with the constraints (4.27), we make sure that every operating room, if not assigned to trauma management, must be dedicated to surgeries belonging to the same class of programmable operations. Planning constraints related to prosthetic surgeries characterize constraints (4.28) and (4.29), specifically, in (4.28) the maximum number of different types of prosthetic surgeries that can be scheduled during the same time slot is defined. Constraints (4.29), state that prothesis surgeries can be planned only during morning time slots.

The number of time slots assigned to each surgeon is bounded by constraints (4.30) and (4.31). Constraints (4.32) impose that the maximum number of prothesis surgeries that can be planned in each period is bounded by S_{ktp}, while constraints (4.33) define the possibility of overtime for non-prosthetic surgeries.

Anesthetist appointment management for planned patients that have been classified as critical is managed by constraints (4.34), (4.35) and (4.36). Specifically, constraints (4.34) states that if a patient is selected for surgery and it has been classified as critical, his anesthetist appointment can be planned at most in one day of the time horizon. If an anesthetist consult is planned for the patient, constraints (4.35) impose that it must be attended at least h_i days before the surgery day. Constraints (4.36) manage the planning of medical examinations taking into account the number of appointment slots still available.

Similarly, the planning of autologous blood donation is handled in the constraints (4.37), (4.38), (4.39), (4.40) and (4.41). Namely, constraints (4.37) and (4.39) ensure that, if a patient is selected to undergo surgery and it has been defined as suitable for autologous blood donation, both first and second sampling can be planned in at most one day of the planning horizon. Constraints (4.38) ensure that, if the first blood sample is taken, it is planned at least twenty days before the surgery date. As a consequence of first sample planning, constraints (4.41) ensure that the second blood sample must be executed in a range that goes from fifteen days to five days prior to the surgery date. Constraint (4.41) manage the planning of blood sampling activities taking into account the number of appointment slots still available. In constraints (4.42) we require that each surgeon is not assigned to more than one operating room in the same period. Since it is possible that, for certain types of intervention r_d, some surgeries must be managed together by two surgeons, constraints (4.42) require that a surgeon w_2 is not assigned to any operating room in the period in which he has to support another surgeon w_1.

4.5 A tool to support the planning activities: results interpretation

The practical model presented in Section 4.4 has been implemented on IBM ILOG CPLEX Optimization Studio 12.5, and tested with realistic instances provided by a Local Hospital Health department using perturbed data for privacy reasons.

In order to build a pilot study, the data were provided in an Excel file containing, for each patient, his pseudo ID, priority class, the date of the admission in the waiting list and the type of surgery he has to undergo. Since age information was available just for a portion of the patients enrolled in the waiting list we assumed that the observed age distribution is valid for all patients. In addition to patient's basic data, information regarding his critical status, and the subsequent need to schedule an anesthetist appointment, as well as his suitability to autologous blood donation were defined. In order to implement the model by using real information as input we implemented an Access Data Base, through which it is also possible to obtain a clear representation of output optimal schedules. It is known that a proposed plan not always turn to be a feasible one since patients availability is not known in advance. After the definition of the subset of patients that could undergo surgery during the week of planning, a phone call to each of them has to be done in order to verify their availability. Since elective patients are not in life-threatening condition they may decide to postpone the surgery. It is then necessary to run again the optimization in order to define a new potential subset of patients that will replace not available ones. The DB interface was created in order to simplify the insertion of new patients on the list, the change of patients' day availabilities and the analysis of proposed schedules so as to facilitate the adoption by medical managers. The DB interface may also be used as a tool for scenario analysis developing a multi-week plan. As a consequence routines have been implemented in order to remove from waiting lists patients that have been planned in previous weeks.

4.5.1 Real instances and results interpretation

The initial waiting list is composed by 1014 patients, 861 of them are waiting for a non-prosthetic surgery while 153 need a prosthetic one. The mean waiting time of prosthetic patients is 270 days. A detailed distribution of waiting times is shown in figure 4.3.

Non-prosthetic patients have a mean waiting time of 428 days, significantly higher than the one of the patients in the prosthetic waiting list. Figure 4.4 shows that in some cases the waiting time can be larger than two years.

Waiting time distribution has to be evaluated according to the priority classification of each patient. The waiting list is composed by 57 priority A patients and 11 of them are

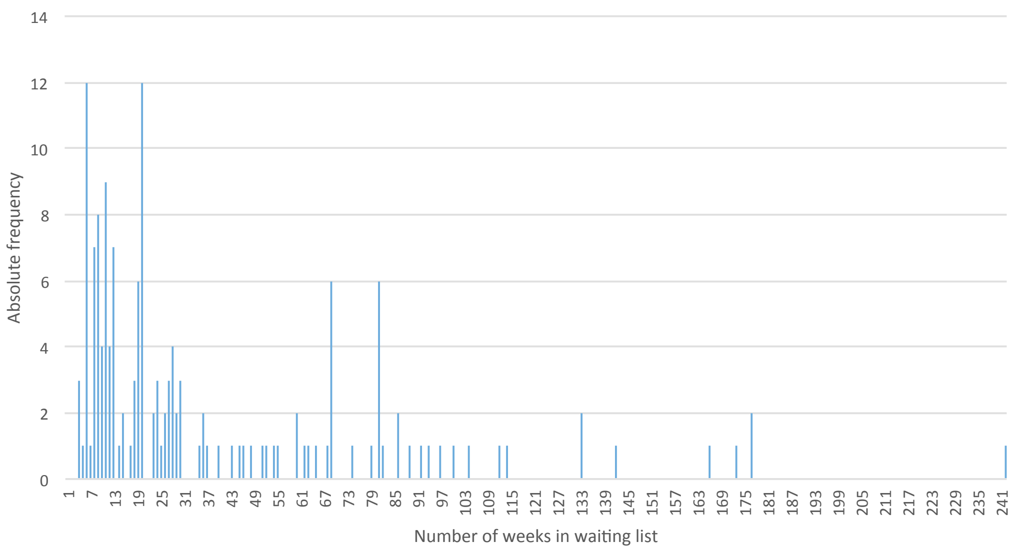


FIGURE 4.3: Non-prosthetic surgery waiting time absolute frequencies in weeks

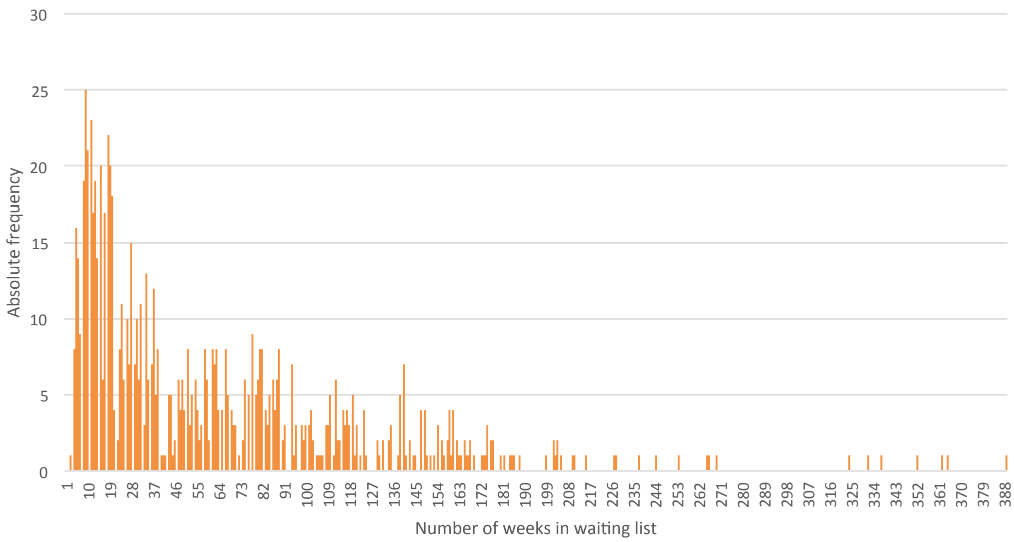


FIGURE 4.4: Prosthetic surgery waiting time absolute frequencies in weeks

waiting for a prothesic surgery. Figure 4.5 shows the waiting time absolute frequency of priority A patients comparing prosthetic and non-prosthetic ones. It is important to underline that only a few component of both prosthetic and non-prosthetic patients is currently respecting the regional deadline.

As far as priority B class concerns there are 208 patients in the waiting list and 17.79% of them are waiting for a prosthetic surgery. Figure 4.6 shows the waiting time absolute frequency for both cohorts; in this case the deadline has not been passed by 37.84% of prosthetic patients and 45.03% of prosthetic ones.

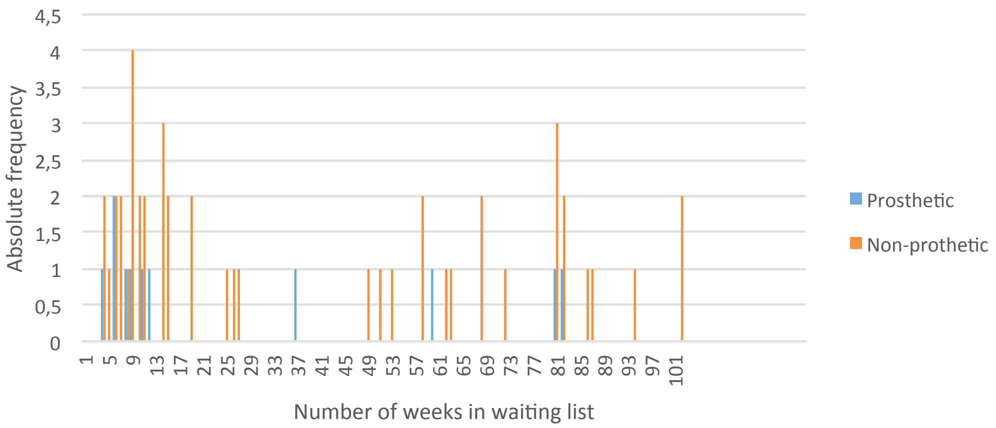


FIGURE 4.5: A priority patients waiting time absolute frequencies in weeks

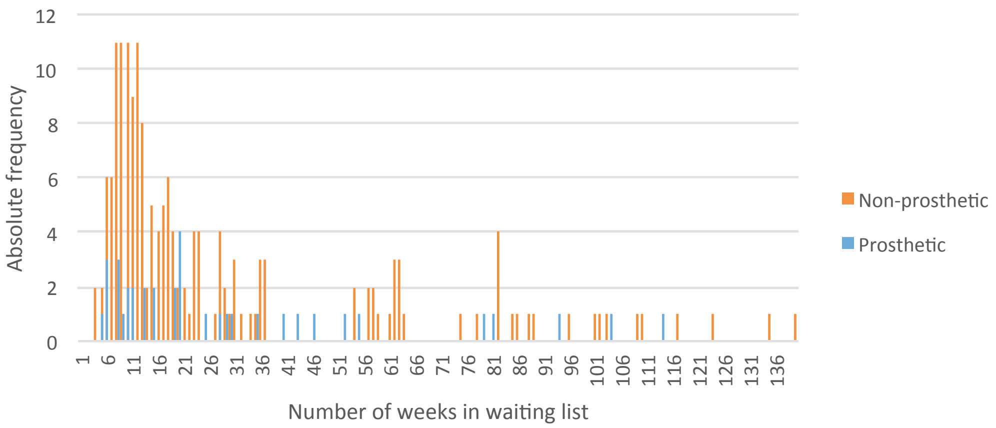


FIGURE 4.6: B priority patients waiting time absolute frequencies in weeks

The number of patients still waiting increases with the decreasing of the priority level since there are 272 priority C patients and 475 priority D ones still waiting for a surgery treatment. Figures 4.7 and 4.8 shows the waiting time absolute frequency respectively for class C and D patients. For both cohorts it is interesting to observe that there is an increased percentage of patients that is not already missing the deadline. Respectively 73.13% and 66.67% for non-prosthetic and prosthetic patients of class C and 64.58% and 88.33% for non-prosthetic and prosthetic patients of class D.

According to the data provided, 20 time slots per week may be assigned to surgeons and usually some of them are reserved to trauma management (see Table 4.3). The department has fifteen surgeons that may be active each week.

By consulting historical data on Operating Room weekly assignments it was possible to define the minimum number L_w of periods assigned to each of the fifteen surgeons working at the Department of Orthopedics. Since no information is available for the

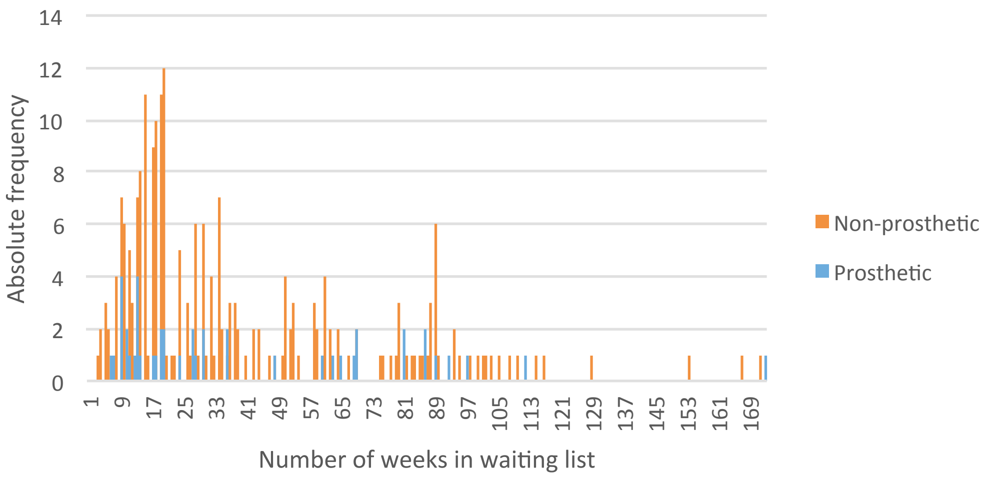


FIGURE 4.7: C priority patients waiting time absolute frequencies in weeks

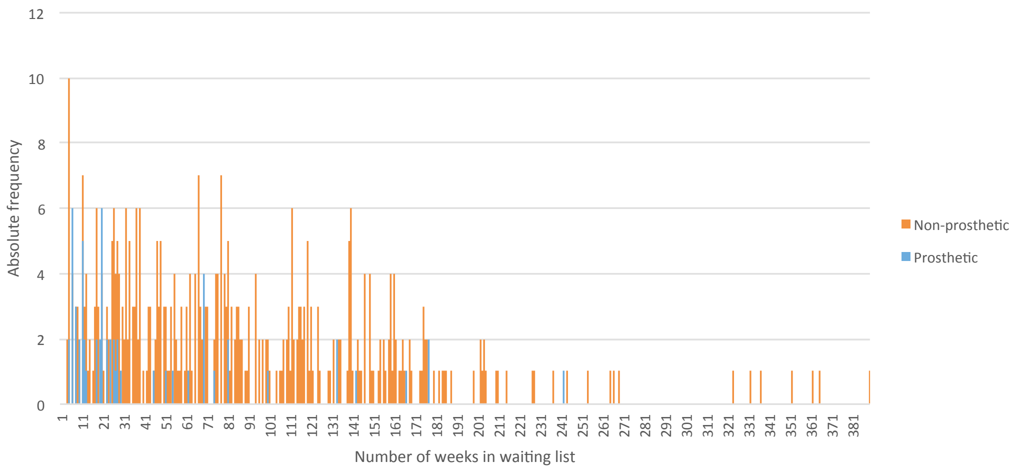


FIGURE 4.8: D priority patients waiting time absolute frequencies in weeks

Day	Time slot	Operating Room ID
Monday	8 AM - 2 PM	1
Tuesday	2 PM - 8 PM	1
Wednesday	8 AM - 2 PM	1
Friday	2 PM - 8 PM	1

TABLE 4.3: Time slots and operating rooms reserved for trauma management

maximum number of assignments for each surgeon U_w we defined as a valid bound the number of time slots that can be assigned to a single surgeon without violating non-ubiquity constraints (4.42).

As already mentioned in Section 4.4, in each period (morning or afternoon sessions), only surgeries related to the same class can be performed. Only by testing the model with the real-world instances we realized that the optimal solution of the model consisted only of non-prosthetic surgeries. This is due to the low percentage of patients with very high waiting times that need to undergo a prosthetic surgery. Since the department wants to ensure a minimum number of time slots per-week dedicated to perform prosthetic surgeries due to organizational goals we added the following lower bound on the number of time slots that have to be weekly assigned to prosthetic activities.

$$\sum_{w=1}^{m} \sum_{k \in K} \sum_{t=|T|-4}^{|T|} \sum_{p \in P} g_{wktp1} \geq E_{prosthetic} \tag{4.57}$$

A slight strengthening of the model can be obtained by putting an upper bound on the number of time slots assigned to each surgeon for prosthetic surgeries in relation with the number of patients in need of prosthetic in the surgeon's waiting list. More precisely,

$$\sum_{w=1}^{m} \sum_{t=|T|-4}^{|T|} \sum_{p \in P} S_k g_{wktp1} \leq |P_w| \quad \forall w = 1, 2, \dots, m \tag{4.58}$$

Note that in some cases $|P_w| = 0$ (where $P_w := \{I_w \cap I_r : r \in Class_1\}$), thus the effect of constraints (4.58) is to fix variables to 0 simplifying the model.

As an output the optimization tool proposes two set of informations that are respectively related to how operating room time slots are assigned to surgeons and which patients have been assigned to operating room time slots.

At first we run the MIP model with an initial configuration of equal weights for all objective function components and analyze the optimized schedule. Figure 4.9 shows the output information regarding operating room time slots assignment to surgeons. Each time slot that is not dedicated to trauma management is assigned to a surgeon and the class of surgeries that will be perfumed is selected.

Figure 4.10 shows for each time slot which is the subset of patients that have been selected for surgery. The first proposed approach shows that with the initial weights configurations all possible overtime slots would be used in order to reduce overall patients waiting time that in most cases are already missing the priority-driven deadlines. It is clear that the impact of overtime is, in this case, too low if compared to patients long waiting time.

We then decided to tune the weighting coefficients in order to identify the configuration that admits overtime just in case that deadlines are missing. This coefficient is

Date	Time slot	Surgery class	Operating room id	Waiting list
28/05/2012	08:00-14:00	Prothesic	2	8
28/05/2012	14:00-20:00	Not Prothesic	1	1
28/05/2012	14:00-20:00	Not Prothesic	2	4
29/05/2012	08:00-14:00	Prothesic	1	1
29/05/2012	08:00-14:00	Not Prothesic	2	2
29/05/2012	14:00-20:00	Not Prothesic	2	13
30/05/2012	08:00-14:00	Not Prothesic	2	6
30/05/2012	14:00-20:00	Not Prothesic	1	3
30/05/2012	14:00-20:00	Not Prothesic	2	1
31/05/2012	08:00-14:00	Prothesic	1	4
31/05/2012	08:00-14:00	Not Prothesic	2	13
31/05/2012	14:00-20:00	Not Prothesic	1	11
31/05/2012	14:00-20:00	Not Prothesic	2	4
01/06/2012	08:00-14:00	Not Prothesic	1	10
01/06/2012	08:00-14:00	Prothesic	2	9
01/06/2012	14:00-20:00	Not Prothesic	2	1

FIGURE 4.9: Operating room assignment sample with equal weights

Date	Time slot id	Operating room id	Patient id	Waiting list id	Surgery type	Priority	Anesthetist ;	First blood	Second blood
28/05/2012	1	2	1348	8	Knee	B	25/04/2012	23/04/2012	14/05/2012
28/05/2012	1	2	1349	8	Hip	B	02/05/2012	23/04/2012	14/05/2012
28/05/2012	2	2	562	4	Not prothesic	A			
28/05/2012	2	2	574	4	Not prothesic	B			
28/05/2012	2	2	565	4	Not prothesic	A			
28/05/2012	2	1	282	1	Not prothesic	D			
28/05/2012	2	1	178	1	Not prothesic	B			
28/05/2012	2	1	182	1	Not prothesic	B			
29/05/2012	1	1	179	1	Shoulder	B	02/05/2012	23/04/2012	14/05/2012
29/05/2012	1	2	708	2	Not prothesic	B			
29/05/2012	1	2	846	2	Not prothesic	D			
29/05/2012	1	1	187	1	Shoulder	B	25/04/2012	23/04/2012	21/05/2012
29/05/2012	1	2	709	2	Not prothesic	B			
29/05/2012	2	2	690	13	Not prothesic	D			
29/05/2012	2	2	689	13	Not prothesic	D			
29/05/2012	2	2	688	13	Not prothesic	D			
30/05/2012	1	2	1279	6	Not prothesic	D			
30/05/2012	1	2	1280	6	Not prothesic	D			
30/05/2012	1	2	1232	6	Not prothesic	B			
30/05/2012	2	1	2	3	Not prothesic	A			
30/05/2012	2	1	4	3	Not prothesic	A			
30/05/2012	2	2	165	1	Not prothesic	A			
30/05/2012	2	2	177	1	Not prothesic	B			
30/05/2012	2	2	176	1	Not prothesic	B			
30/05/2012	2	1	1	3	Not prothesic	A			
31/05/2012	1	1	566	4	Hip	A	16/05/2012	30/04/2012	14/05/2012
31/05/2012	1	2	683	13	Not prothesic	C			
31/05/2012	1	2	682	13	Not prothesic	C			
31/05/2012	1	2	687	13	Not prothesic	D			
31/05/2012	1	1	578	4	Hip	B	09/05/2012	23/04/2012	14/05/2012
31/05/2012	2	1	1184	11	Not prothesic	B			
31/05/2012	2	2	558	4	Not prothesic	A			
31/05/2012	2	2	559	4	Not prothesic	A			
31/05/2012	2	1	1182	11	Not prothesic	A			
31/05/2012	2	1	1183	11	Not prothesic	A			
31/05/2012	2	2	557	4	Not prothesic	A			
01/06/2012	1	1	1202	10	Not prothesic	B			
01/06/2012	1	1	1219	10	Not prothesic	D			
01/06/2012	1	2	1340	9	Hip	D	02/05/2012	23/04/2012	21/05/2012
01/06/2012	1	2	1313	9	Knee	B	02/05/2012	23/04/2012	21/05/2012
01/06/2012	1	1	1218	10	Not prothesic	D			
01/06/2012	2	2	283	1	Not prothesic	D			
01/06/2012	2	2	181	1	Not prothesic	B			
01/06/2012	2	2	175	1	Not prothesic	B			

FIGURE 4.10: Operating room planning sample with equal weights

strongly influenced by the characteristics of the waiting lists. In theory overtime may be admissible if and only if there is a patient that, if not planned in overtime, will not undergo surgery before his deadline. This means that, observing priority coefficient trend line in Figure 4.2a, the definition of the ratio between overtime weight P_{OT} and waiting time weight P_{WT} should be equal to four in order to admit overtime only in case a patients is missing his deadline. Even if this approach could be reasonable from a theoretical point of view it is clear that the long waiting times that characterize the patients in the waiting list will lead to the utilization of all possible overtime (see Table 4.4).

Number of patients	Number of surgeries performed in overtime	Number of OR time slots	CPU time	Objective function value	Gap %
1014	12	16	190.60	4,262,171.45	0.00
970	12	16	265.14	1,991,193.91	0.00
928	12	16	140.00	1,314,310.51	0.00
887	12	16	561.67	906,458.50	0.00

Table 4.4: Real word instances results with overtime allowed only for patients missing their deadline

Observing Figure (4.2) it is then possible to affirm that an overtime discourage can be achieved only by setting the overtime weighting factor P_{OT} up to three order of magnitude higher than the waiting time one (see Figure 4.11 and Figure 4.12 for an example of resulting planning).

Date	Time slot id	Operating room id	Patient id	Waiting list	Surgery type	Priority	Anesthetist	First blood	Second blood
28/05/2012	1	2	1348	8	Knee	B	16/05/2012	23/04/2012	14/05/2012
28/05/2012	1	2	1349	8	Hip	B	09/05/2012	23/04/2012	14/05/2012
28/05/2012	2	1	557	4	Not prothesic	A			
28/05/2012	2	2	687	13	Not prothesic	D			
28/05/2012	2	1	558	4	Not prothesic	A			
28/05/2012	2	1	559	4	Not prothesic	A			
28/05/2012	2	2	688	13	Not prothesic	D			
29/05/2012	1	2	1340	9	Hip	D	02/05/2012	23/04/2012	21/05/2012
29/05/2012	1	1	682	13	Not prothesic	C			
29/05/2012	1	1	689	13	Not prothesic	D			
29/05/2012	1	2	1313	9	Knee	B	09/05/2012	30/04/2012	14/05/2012
29/05/2012	2	2	846	2	Not prothesic	D			
29/05/2012	2	2	709	2	Not prothesic	B			
29/05/2012	2	2	708	2	Not prothesic	B			
30/05/2012	1	2	179	1	Shoulder	B	09/05/2012	23/04/2012	14/05/2012
30/05/2012	1	2	187	1	Shoulder	B	16/05/2012	30/04/2012	14/05/2012
30/05/2012	2	1	282	1	Not prothesic	D			
30/05/2012	2	2	1218	10	Not prothesic	D			
30/05/2012	2	1	178	1	Not prothesic	B			
30/05/2012	2	2	1202	10	Not prothesic	B			
31/05/2012	1	1	563	4	Knee	A	02/05/2012	23/04/2012	14/05/2012
31/05/2012	1	2	177	1	Not prothesic	B			
31/05/2012	1	2	176	1	Not prothesic	B			
31/05/2012	1	1	578	4	Hip	B	02/05/2012	23/04/2012	14/05/2012
31/05/2012	2	2	1280	6	Not prothesic	D			
31/05/2012	2	2	1279	6	Not prothesic	D			
31/05/2012	2	2	1232	6	Not prothesic	B			
31/05/2012	2	1	1	3	Not prothesic	A			
31/05/2012	2	1	2	3	Not prothesic	A			
31/05/2012	2	1	5	3	Not prothesic	A			
01/06/2012	1	1	1182	11	Not prothesic	A			
01/06/2012	1	1	1184	11	Not prothesic	B			
01/06/2012	1	2	565	4	Not prothesic	A			
01/06/2012	1	2	562	4	Not prothesic	A			
01/06/2012	1	2	574	4	Not prothesic	B			
01/06/2012	1	1	1183	11	Not prothesic	A			
01/06/2012	2	2	175	1	Not prothesic	B			
01/06/2012	2	2	165	1	Not prothesic	A			

Figure 4.11: Operating room planning sample

Date	Time slot	Surgery class	Operating room id	Waiting list
28/05/2012	08:00-14:00	Prothesic	2	8
28/05/2012	14:00-20:00	Not Prothesic	1	4
28/05/2012	14:00-20:00	Not Prothesic	2	13
29/05/2012	08:00-14:00	Not Prothesic	1	13
29/05/2012	08:00-14:00	Prothesic	2	9
29/05/2012	14:00-20:00	Not Prothesic	2	2
30/05/2012	08:00-14:00	Prothesic	2	1
30/05/2012	14:00-20:00	Not Prothesic	1	1
30/05/2012	14:00-20:00	Not Prothesic	2	10
31/05/2012	08:00-14:00	Prothesic	1	4
31/05/2012	08:00-14:00	Not Prothesic	2	1
31/05/2012	14:00-20:00	Not Prothesic	1	3
31/05/2012	14:00-20:00	Not Prothesic	2	6
01/06/2012	08:00-14:00	Not Prothesic	1	11
01/06/2012	08:00-14:00	Not Prothesic	2	4
01/06/2012	14:00-20:00	Not Prothesic	2	1

FIGURE 4.12: Operating room assignment sample

The analysis of model outputs is also interesting in terms of number of time slots dedicated to prosthetic surgeries. It is clear that waiting time imbalance between prosthetic and non-prosthetic patients in waiting lists showed in figures 4.3 and 4.4 leads to a number of time slots assigned to prosthetic surgeries equal to the lower bound that we imposed. If we observe the outputs not just for the initial waiting list but over a one month planning horizon by applying a rolling planning approach that clears the planned patients from the waiting list every time a new planning week is evaluated, we see that after the third week of planning the number of time slots assigned to prosthetic surgeries increases.

The proposed planning approach is at the moment not fully comparable with the handmade solutions. This is because the current approach is totally informal. Specifically, each surgeon defines his own subset of patients after time slot assignment. The weekly plan is not made one month in advance but in most cases two weeks in advance. Anesthetist appointments as well as autologous blood donations are at the moment poorly planned. As far as anesthetist appointment concerns, there is no fixed number of appointments assigned to the orthopedic department, so, every time a critical patient is planned, the surgeon staff tries to arrange an appointment and, if not possible, plans an early hospitalization for the patient. The short advance in Operating Theater planning in some cases forbids an effective planning of autologous blood donation. Patients selection follows a qualitative and informal approach in which the waiting list is analyzed and patients with longer waiting times are selected after some considerations based on their priority level. Every time a potential patient is identified, a phone call is made in order to verify his availability on the proposed days. It is worth mentioning that the handmade solution takes approximately one and a half to two hours to manually define the surgical schedules while the proposed approach takes on average five minutes to solve the problem and to propose solutions that can be iteratively adjusted to accommodate specific needs.

To analyze the impact of the various components of the model and in particular of its complex and heterogeneous objective function, we decided to examine how the relaxation or the strengthening of some constraints would affect the optimal proposed solution.

We tested four alternative configurations:

1. Initial configuration;
2. Overtime disabled;
3. Political bound on minimum number of time slots assigned to surgeons disabled;
4. Initial configuration with one additional operating room.

Table 4.5 shows the results of the comparison of the four configurations. The tests have been made on four instances that have been extracted from the initial waiting list running the standard configured model over a two months planning horizon.

Instance ID	Number of patients	Number of OR	Overtime enabled	Bound enabled	CPU time (sec)	Objective function value	Variation w.r.t. base line configuration
0.1	1014	2	1	1	135.83	4,261,089.06	/
0.1	1014	2	0	1	73.84	4,265,938.06	0.114%
0.1	1014	2	1	0	105.26	4,258,092.06	-0.070%
0.1	1014	3	1	1	2454.86	4,247,993.06	-0.307%
0.2	970	2	1	1	155.77	2,001,231.38	/
0.2	970	2	0	1	64.16	2,004,402.39	0.158%
0.2	970	2	1	0	52.95	1,996,938.39	-0.215%
0.2	970	3	1	1	646.36	1,991,053.39	-0.509%
0.3	928	2	1	1	84.47	1,322,456.19	/
0.3	928	2	0	1	57.39	1,324,779.19	0.176%
0.3	928	2	1	0	91.71	1,319,353.19	-0.235%
0.3	928	3	1	1	1606.92	1,314,820.94	-0.577%
0.4	887	2	1	1	81.15	918,003.97	/
0.4	887	2	0	1	63.13	919,619.97	0.176%
0.4	887	2	1	0	70.65	915,434.97	-0.280%
0.4	887	3	1	1	202.35	911,884.97	-0.667%

Table 4.5: Computational results under different configurations

Initially, we evaluated the impact of overtime prohibition by comparing configuration (1) and (2). Analyzing the various components of the objective function it is possible to observe that, due to the high volume of patients missing their deadline, the waiting time component is, by far, the most relevant one. It is then clear that a reduction in terms of overtime is less attractive than a reduction in terms of patients waiting time. As a consequence in all the four test instances that we used it is possible to measure a worsening in terms of objective function value that is caused by the fact that 12 patients less are planned in each tested instance. Since the number of potential assignments is reduced it is also possible to observe that the computational time to find an optimal solution is 35.99% lower for configuration (2) if compared to the initial one.

If we then compare (1) and (3) configurations it is possible to observe a slight improvement in terms of objective function solution value just by defining a different subset of time slot assignment to surgeons. That is to say that the actual subdivision of operating room time slots to surgeons is not correlated with the real condition of patients waiting in different lists. That is even more important because the improvement in terms of solution quality is reached by planning exactly the same number of patients, exploiting all possible regular and overtime time slots. We can then affirm that the political bounds that is actually implemented in the department leads to suboptimal solutions.

As a last comparison we took the results obtained by configuration (4) evaluating the impact of an increase in the number of operating rooms assigned to the orthopedic department. In this situation it is possible to see that all additional time slots related to the new operating room would be used, thus increasing the number of patients undergoing surgery in each planning week up to 30 more than the initial configuration. This would result in an improvement in terms of final objective function value related to waiting time reduction. This result is obtained at the price of higher computing times that can, in some cases, get 10 times higher. As a final remark it is important to point out that the increased number of operating rooms would lead, due to the situation of the waiting lists, to an increase in terms of overtimes. As stated for (1)-(2) comparison, the overtime component increase is, for the tested instances, of minor importance if compared with the waiting time one.

If we focus our attention only on Table 4.5, the differences between the three proposed policies seem to be very little in terms of overall waiting time reduction. That is not the case if we analyze the impact of the four proposed policies over a wider planning horizon. In order to do that, we defined as a starting point for all the four configurations the initial waiting list and run a one month rolling planning for each of them.

Table 4.6 shows that, by increasing the time horizon, the differences between the proposed policies clearly emerge. The comparison between configuration (1) and (2) shows that the overtime disabling would lead to a performance worsening of 12.88% that is mainly due to the decrease of 46 patients among the planned ones with a resulting increase in terms of overall length of stay in the waiting list.

Instance ID	Number of patients	Number of OR	Overtime enabled	Bound enabled	CPU time (sec)	Objective function value	Variation w.r.t. base line configuration
0.1	1014	2	1	1	135.83	4,261,089.06	
0.2	970	2	1	1	155.77	2,001,231.38	
0.3	928	2	1	1	84.47	1,322,456.19	
0.4	887	2	1	1	81.15	918,003.97	
						8,502,780.60	/
2.1	1014	2	0	1	73.84	4,265,938.06	
2.2	982	2	0	1	66.77	2,293,036.92	
2.3	950	2	0	1	65.55	1,715,845.56	
2.4	918	2	0	1	77.47	1,323,454.50	
						9,598,275.04	12.88%
1.1	1014	2	1	0	105.26	4,258,092.06	
1.2	970	2	1	0	107.56	1,963,624.92	
1.3	926	2	1	0	80.4	1,089,167.58	
1.4	883	2	1	0	72.92	717,039.59	
						8,027,924.14	-5.58%
3.1	1014	3	1	1	2454.86	4,247,993.06	
3.2	940	3	1	1	468.38	1,289,106.92	
3.3	869	3	1	1	421.83	600,397.58	
3.4	798	3	1	1	479.98	284,630.58	
						6,422,128.14	-24.47%

TABLE 4.6: Computational results under different configurations over a 4 week planning horizon

The impact of the assignment bounds can be measured evaluating configuration (3). Even if the overall number of patients that undergo surgery in the planning period is very similar (843 versus 839) the overall waiting time reduction would lead to a 5.58% decrease in terms of objective function value. That phenomenon strengthens the perception that political bounds lead to heavily suboptimal solutions because they forbid a planning cycle that follows overall patients needs. If we then evaluate the assignment of an additional operating room to the orthopedic department we can observe that an improvement in terms of objective function value equal to 24.47% could be reached as a consequence of waiting time reduction. The additional operating room would allow, over a one month planning horizon, to plan 112 additional patients, thus reducing the number of them still missing their deadline.

4.6 Conclusions

Healthcare system management (in terms of waiting list reduction), health care appropriateness and resource utilization efficiency is a subject of great importance. The increasing need of virtuous programming practices for health services and the subsequent development of quantitative tools to support decision-making processes had grown mainly for the set of activities that are considered most critical from the organizational point of view. Operating rooms planning is, among those studied by operations management scientists, the one with the highest systemic impact due to the number and the economic importance of involved resources. Operating rooms are shared and limited resources, therefore they have a strong impact on hospitals both in terms of quality of service provided to the patients and of economic and organizational performance. In the presented chapter we analyzed the Operating Theater planning problem from two perspectives. First, we designed a Mixed Integer Programming model following benchmark guidelines that are defined by excellent Regional hospitals. As a result we designed a model that is an hybridization of literature SCAP models with the inclusion of regional-driven side constraints. The definition of the planning objectives comes from the high level of importance that is given in Emilia-Romagna (regional) guidelines to patient care. In particular a strong importance is given to the respect of priority-driven guidelines. Computational experiments suggested that, even if the data we used came from a big regional hospital trust, the problem can be easily solved using a commercial MIP solver. The integration of not yet implemented or poorly managed regional guidelines into the optimization tool emphasizes the importance of the case study as a driving factor to ease the planning of complex activities. Another interesting factor that could be analyzed in the future is the impact of national and regional policies that aim at reducing the number of beds available for each department and, in some cases, to consider shared beds among departments. Even if the model in section 4.3 considers bed capacity, in Section 4.4 we have not considered those resources because at the time of data mining they did not influence the final solution. We are aware that a reduction in the number of available beds is now under evaluation and it is plausible that in the future department OT planning will have to consider those resources as active constraints that can reduce the overall number of performed surgeries if they are not properly managed.

Bibliography

I. J. B. F. Adan and J. M. H. Vissers. Patient mix optimisation in hospital admission planning: a case study. *International Journal of Operations & Production Management*, 22(4):445–461, 2002. doi: 10.1108/01443570210420430.

Ivo Adan, Jos Bekkers, Nico Dellaert, Jan Vissers, and Xiaoting Yu. Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning. *Health Care Management Science*, 12(2):129–141, 2009. ISSN 1386-9620.

S. Chaabane, N. Meskens, A. Guinet, and M. Laurent. Comparison of two methods of operating theatre planning: Application in belgian hospital. In *Service Systems and Service Management, 2006 International Conference on*, volume 1, pages 386–392, 2006.

Domenico Conforti, Francesca Guerriero, Rosita Guido, MarcoMatucci Cerinic, and MariaLetizia Conforti. An optimal decision making model for supporting week hospital management. *Health Care Management Science*, 14(1):74–88, 2011.

Brian T. Denton, Andrew J. Miller, Hari J. Balasubramanian, and Todd R. Huschka. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, 58(4-part-1):802–816, 2010. doi: 10.1287/opre.1090.0791.

H. Fei, C. Chu, N. Meskens, and A. Artiba. Solving surgical cases assignment problem by a branch-and-price approach. *International Journal of Production Economics*, 112 (1):96 – 108, 2008. Special Section on Recent Developments in the Design, Control, Planning and Scheduling of Productive Systems.

A. Guinet and S. Chaabane. Operating theatre planning. *International Journal of Production Economics*, 85(1):69–81, 2003.

Andrea Lodi and Paolo Tubertini. Pre-operative activities and operating room planning: a local hospital department case study. Technical Report OR 14-6, LabOR DEI - Università di Bologna, 2014.

Jeroen M. Oostrum, M. Houdenhoven, J.L. Hurink, E.W. Hans, G. Wullink, and G. Kazemier. A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, 30(2):355–374, 2008.

Atle Riise and Edmund K. Burke. Local search for the surgery admission planning problem. *Journal of Heuristics*, 17(4):389–414, 2011.

Elena Tanfani and Angela Testi. A pre-assignment heuristic algorithm for the master surgical schedule problem (mssp). *Annals of Operations Research*, 178(1):105–119, 2010.

Angela Testi and Elena Tanfani. Tactical and operational decisions for operating room planning: Efficiency and welfare implications. *Health Care Management Science*, 12 (4):363–373, 2009.

Angela Testi, Elena Tanfani, and Giancarlo Torre. A three-phase approach for operating theatre schedules. *Health Care Management Science*, 10(2):163–172, 2007.

J. M. H. Vissers, I. J. B. F. Adan, and J. A. Bekkers. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*, 16(3):281–304, 2005.

G. Wullink, M. Van Houdenhoven, E. W. Hans, J. M. van Oostrum, M. van der Lans, and G. Kazemier. Closing emergency operating rooms improves efficiency. *J Med Syst*, 31(6):543–546, Dec 2007.

Chapter 5

Kidney Exchange Problem: a simulation-optimization approach [1]

5.1 Introduction

Renal disease is a growing problem, affecting thousands of patients. Although a patient can incur in hemodialysis treatment this is undesirable as the patients life quality is severely hampered and the treatment has high costs. To solve this problem, the patients traditionally enter in the local deceased donor list, where they hope to get a deceased donor kidney for transplant. However organ demand largely overcomes organ supply, so new alternatives have been developed. More recently some countries developed policies where they allowed a patient to receive an organ from a living donor that volunteered to donate a kidney. The donor is assumed to be emotionally or blood related with the patient and must be compatible in order for the transplant to be performed. This compatibility is tested both at blood and tissue-type level. When the pair is shown to be incompatible there are several alternatives, depending on the country:

- the patient may start a desensitization treatment to progressively overcome his donors HLA incompatibility. This treatment typically takes a long time and incurs in high costs;
- the donor may hand over his kidney to a patient in the deceased donor list in exchange for an higher position for his associated patient in the same list, however the quality of a living donor organ is assumed to be much superior to an organ provided by a deceased donor [Zenios, 2002];

[1] This chapter is based on Technical Report OR 14-8 (see Constantino et al. [2014])

- both patient and donor may enter a kidney exchange program. In these programs the donor will cede his organ to another patient in similar conditions in exchange for a living donor kidney for his related patient.

This last option seems superior as the patient may hope to find an organ of similar quality to the one that is given. The corresponding matching of patients is known as the Kidney Exchange Problem (KEP).

Although the initial kidney exchange programs were composed exclusively of incompatible pairs they evolved continuously and nowadays may include: (i) patients that have multiple incompatible donors; (ii) donors without an associated patient that are willing to donate a kidney for no return, usually called altruistic donors; (iii) patients that have a compatible donor but enter the exchange program hopping to find a higher quality organ. A natural way of representing a KEP is through a directed graph. For each patient/donor pair we create a vertex. If the donor of a given vertex is compatible with the patient of another one we draw an arc from the first to the second to signal the compatibility. In figure 5.1we show the example of a pool with three patient/donor pairs. The donor of pair 1 is compatible with the patients of both pair 2 and pair 3. The donor of pair 2 is compatible with the patient of pair 3 and the donor of pair 3 is compatible with the patient of pair 1. Vertex 4 represents an altruistic donor that is compatible with the patient of pair 2. The clearing of the problem may be achieved both by using cycles (sequence of patient/donor pairs) or chains (sequence that starts with an altruistic donor and afterwards includes patient/donor pairs). Note that in a cycle, the donor of the last pair cedes its organ to the patient of the first pair, while in chains and depending on the adopted policies, the donor of the last pair may give his kidney to the deceased donor list or he may be used as an altruistic donor to start another chain in the future. The second option is know as Never Ending Altruistic Donor chain (NEAD) as the chain may be prolonged indefinitely. Cycles are also characterized by the number of pairs included: if a cycle has two elements we say it represents a two-way exchange. More generally if a cycle contains k is deemed a k-way exchange. In the presented graph we can find a new transplant for each patient with: a cycle $c_1 = \{1, 3\}$ and a chain $h_1 = \{4, 2\}$; a single cycle $c_2 = \{1, 2, 3\}$; a single chain $h_2 = \{4, 2, 3, 1\}$.

While the KEP is simple to state, its management is extremely hard since optimization, logistic or even ethical questions may rise during each step of the process. In this work we describe a simulation framework for the KEP. The developed tool is extremely flexible, allowing to simulate different pool management policies and contains features not found in other simulators. The proposed solution is also efficient, being able to solve problems of realistic sizes. In the next sections of this work we review the work related with the KEP and describe our proposed simulator for the problem. Finally we conclude presenting computational results, conclusions and future work.

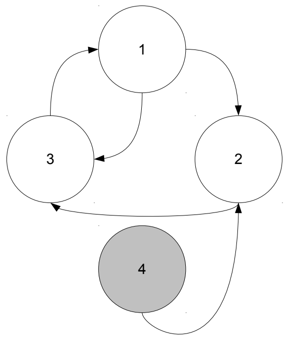


FIGURE 5.1: KEP pool example

5.2 Literature review

First theoretical references to KEP are found in [Rapaport, 1986] and the first transplantations following this paradigm took pace in 1991 in South Korea. Due to its importance, the number of works related to this topic is growing rapidly since 2000. There are two common approaches: static (offline) or dynamic (online). In the static variant we have a fixed pool of patient/donor pairs that we want to clear according to some criteria. In the dynamic case the pool evolves over time: pairs enter the pool seeking for a new kidney and leave after getting one, or if there is another occurrence such as sickness, pregnancy or even patient death. In the following subsection we describe the most relevant (to our knowledge) work in both the static and dynamic KEP and we provide a more detailed description of simulator features available in the literature.

5.2.1 Static KEP

When dealing with a static pool of pairs the solution methods vary. When only 2-way exchanges are considered the problem is solvable in polynomial time with the Edmonds algorithm [Edmonds, 1965]. Similarly if we consider k-way exchanges, where k is equal to the number of pairs, the problem is solvable in polynomial time as an assignment problem. However studies show that if only two-way exchanges are considered an important number transplants is missed [Roth et al., 2007]. The authors also show that by including 3-way exchanges they capture most of the possible pairings and that the advantage of including larger valued exchanges is marginal. Unfortunately, when $k \geq 3$ the clearing problem is NP-Hard [Abraham et al., 2007] and as the number of patients in the pool increases the problem quickly becomes intractable. To solve the KEP when exchanges greater than two are considered the great majority of works use Mixed Integer Programming based either on the cycle or edge formulations also described in the same article. Still in the same article the authors describe a branch and price algorithm developed to tackle very large pools. More recently Constantino et al. [2013] proposed and analysed the performance of alternative compact formulations. Recent focus on probabilistic KEP studies the impact of patient withdrawal and cross match failure (not considered in the deterministic model). In Pedroso [2013] the author proposes a method for computing the maximum expectation for the length of the maximum set of vertex-disjoint cycles in a digraph where vertices and/or arcs have a known probability of failure. The algorithm relies on prepared database of possible configurations to speed up the process. Computational results are presented for both deterministic and probabilistic scenarios.

5.2.2 Dynamic KEP

Considering dynamic KEP some early articles consider the allocation of cadaverous kidneys and its interaction with a KEP.

Zenios et al. [2000] study the allocation of cadaverous kidneys to potential transplant recipients. The authors develop a linear differential equation model and propose a dynamic index policy based on the approximate analysis of the resulting optimal control problem. A tri-criteria objective is considered, maximizing the quality-adjusted life expectancy of transplant candidates and minimizing the likelihood of transplantation of the various types of patients and the difference in mean waiting times across patient types.

The same objective is considered in Zenios [2002] but applied to the case where the donor has the option of making a direct or indirect exchange. They consider a double ended queuing model for an exchange system with two types of donor candidate pairs. An optimal dynamic exchange policy is obtained by invoking a Brownian approximation.

In Segev et al. [2005] the authors conduct a simulation study to evaluate waiting times of different patients. The patients are characterized by factors such as blood type, ethnicity and geographical distance and the authors suggests if a patient should enter a kidney paired donation program or alternatively choose a desensitization treatment.

In Awasthi and Sandholm [2009] a trajectory-based online stochastic optimization algorithm is proposed. At each step the algorithm analyses future scenarios before committing to a decision. Afterwards the algorithm invokes an offline algorithm for matching the maximum number of pairs.

In Ünver [2010] the authors propose efficient dynamic matching mechanisms for two-way and multi-way exchanges. The model is based on blood type compatibility and minimizes the average waiting cost.

In Beccuti et al. [2011] the authors study the effect of considering different time intervals between matches in the number of transplants performed and waiting times. The model is based on patient's blood compatibility and 2-way exchanges are allowed.

In Li et al. [2011] the authors propose a probabilistic graph model. To each edge in the compatibility graph they associate a weight based on an utility function and a probability of failure. After running the clearing algorithm a fall-back mode is also considered, to compensate for patient withdrawal. Computational results are presented based on a developed microsimulation platform. This work is further extended in Chen et al. [2011] with the inclusion of altruistic donors and the development of a graphical user interface simulation platform for managing KEP programs.

In Dickerson et al. [2012a] the authors introduce the concept of potentials. The main idea is to penalize donors that can be matched with under demanded patients and save them for the future. These weights are calculated based in blood compatibility with a parameter tuning package. Computational results compare the total number of transplants obtained in the weighted, unweighed and full information model.

In Dickerson et al. [2012b] the same authors study the influence of the size of altruistic donor chains in the total number of transplants. In Dickerson et al. [2013] the authors study theoretical and practical implications of the probabilistic graph model and chains in the performance of the algorithm of Abraham et al. [2007] and propose new bounding schemes to improve its scaling properties. Based on their previous experience they also propose a bimodal distribution to model the failure probabilities of matching.

In Ashlagi et al. [2013] the authors study the effect of different waiting periods between match runs, cycle size and chain inclusion in the total number of transplants. They present a greedy algorithm that prioritises matching of high PRA patients and study its behaviour for the online, offline and multiple waiting scheme variants of the problem.

5.2.3 Feature comparison

Although the previous subsection describes work done in the dynamic variant of the KEP it is hard to compare the features and capabilities of the various simulators for the problem. For a more throughout comparison we provide table 5.1 where the main features are summarized in the columns.

In the first column we have the author information.

In the second column we describe the pool management system. More precisely: if matchings are conducted dynamically (d) or if a static algorithm is invoked periodically (s); information on how the compatibility graph is generated: more specifically if the model considers blood compatibility, tissue compatibility or both (respectively B, T and BT); and the maximum cycle size allowed.

In the third column we describe particular simulator features that are not common across all implementations: w if we are considering a weighted version of the problem; eu if an expected utility function is used to express weights and probabilities between the donors/patients; fb if the simulator includes a fall-back mechanism to minimize the impact of drop outs; ch^n if the simulator considers altruistic donor chains, n representing the maximum chain size allowed, when specified.

In the fourth column we describe the objective function considered in each article.

article	pool	extra	objective
Segev et al. [2005]	s, BT, 2	w	maximize weighted number of transplants
Awasthi and Sandholm [2009]	d, BT, 3		maximize number of transplants
Ünver [2010]	d, B, n		minimize discounted surplus
Beccuti et al. [2011]	s, B, 2		maximize number of transplants
Li et al. [2011]	s, BT, 3	eu, fb	maximize expected utility
Chen et al. [2011]	s, BT, 3	eu, fb, ch^3	maximize expected utility
Dickerson et al. [2012a]	s, BT, 3	w, ch	maximize weighted number of transplants
Dickerson et al. [2012b]	s, BT, 3	ch^5	maximize number of transplants
Dickerson et al. [2013]	s, BT, 3	eu, ch	maximize expected utility
Ashlagi et al. [2013]	s/d, T, 3		maximize number of transplants

Table 5.1: Comparison of simulator features

5.3 Simulator features

One of the greatest challenges of a Kidney Exchange Program is the choice of policies that should be implemented in order to ensure an effective and fair management of all included patients. Policy evaluation and validation through appropriate Decision Support System (DSS) lead us to identify the need of an integrated tool that can be used in order to evaluate, through simulation, how different policies can impact outcomes of K.E. Programs.

In order to understand the implications of different policies of patient allocation and pool management we developed a simulation tool for a realistic model of KEP pools (see Figure 5.2). As stated in the previous section different techniques have been used in order to model the dynamics of an exchange pool. We decided to implement the simulation component via a Discrete Event Model. Discrete Event Simulation (DES) modeling gives the possibility of representing the key elements of KEP in a simple and straightforward way through the definition of entities and classes of events that can occur during the dynamic evolution of the program.

The main goal of the implemented simulator is to take into consideration the widest typology of actors that can be included into an exchange pool as well as the different policies that can be used in order to manage the matching process. To achieve this goal several national programs were studied (i.e. Portugal. UK, The Nederlands) and their main characteristics were captured. The application is then able to test different configurations regarding matching frequency and matching and pool characteristics.

The simulator was developed in a modular way and has the following components:

- a configuration module where the user selects the parameters for running the simulation and defines the distributions of the components;
- a data generation module responsible for generating data according to the specified distributions;
- a pool management module that will control the evolution of the simulation and manage the succession of events that occur;
- an optimization module that calculates the matching of pairs in the pool for a given time.

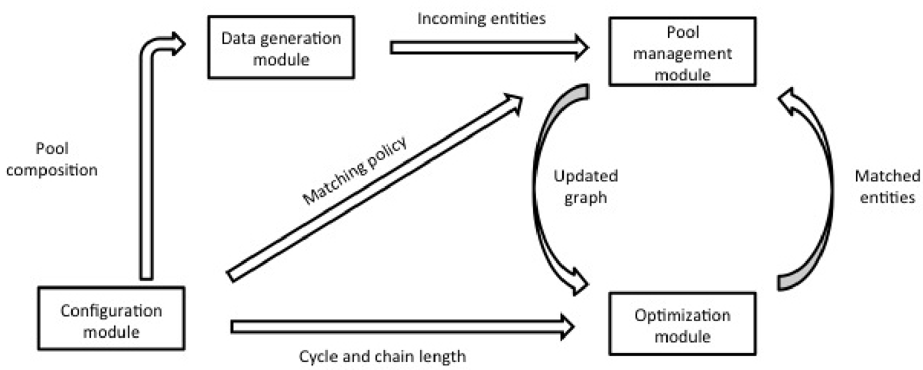


FIGURE 5.2: Simulation-optimization tool components interaction

In the following subsections we show a description of the modules.

5.3.1 Configuration module

The configuration module enables the user to set the characteristics of the simulator in terms of pool composition and matching policy, namely:

- Inclusion of incompatible pair, incompatible pair and altruistic donor;
- HLA or PRA representation of patient and donors tissue type
- After matching crossmatch incompatibility evaluation
- Chain and cycle maximum length definition
- Matching run frequency
- Patient sickness temporary dropout

- Donor sickness temporary dropout
- Patient death permanent dropout
- Donor death permanent dropout
- Definition of weighted or maximum transplants matching rule

the following subsections we describe in detail all the features that can be enabled.

5.3.1.1 Pool Characteristics

We define as pool characteristics the typology of actors that will be considered as being part of the Kidney Exchange program. Most of the papers related to KEP model just incompatible pairs, but it is known that nowadays KEP programs have evolved from this first definition of exchange pool. In the UK, as well as in the USA, exchange pools are composed not only by incompatible pairs but also by compatible ones and by altruistic donors. The introduction of this new kind of actors into exchange pools expands the number of possible alternative matchings and lead to a more complex definition and representation of actors and of graph building rules. It is necessary to evaluate the improvements that a KE program can experience if those additional actors are integrated in the exchange pool.

The configuration model of the simulation that we propose allows the user to select if the pool is composed by incompatible patient/donor pairs, patients with multiple donors, compatible pairs and altruistic donors. In the event of considering altruistic donors the user is also able to determine what happens to the donor at the end of an altruistic donor chain. More precisely if this donor is discarded or if he will be used as an altruistic donor to start a future chain. In order to characterize the three classes of actors that have been mentioned, a set of common attributes can be enabled during the configuration phase so as to properly describe entity behavior within kidney exchange pool. Patients and donors can be described through their blood type, tissue type and age.

Blood type representation. Blood type can be considered has the first element of compatibility evaluation. Blood type is determined by the presence of blood-type proteins called A and B. As a consequence, it is possible to classify blood types in four components, 0, A, B, and AB. A donor can donate a kidney to a recipient who has all the blood-type proteins that the donor possesses, thus:

- 0 blood-type donors are blood-type compatible with all recipients;
- A blood-type donors are blood-type compatible with A and AB blood-type recipients;

- B blood-type donors are blood-type compatible with B and AB blood-type recipients;
- AB blood-type donors are blood-type compatible with AB blood-type recipients.

Compatibility evaluation can not be reduced to blood type test since it is possible that a donor-recipient couple, even if blood type compatible, turns to be tissue type incompatible. In this case donor-recipient pairs can be defined as incompatible if the recipient has developed antibodies to at least one of the antigens that characterize the donor tissue type. Antibody generation is a natural process that help to protect the body against the invasion by foreign antigens. Antibodies can also be created by the body to attack tissue, particularly the tissue of another human. Blood transfusions, transplants, and pregnancy can trigger the production of specific antigen antibodies. Tissue type compatibility can be then managed following two approaches: (i) detailed Human Leukocite Antigen (HLA) proteins description, (ii) Panel Reactive Antibody (PRA) probability classification (see Procurement and Network). The idea of the presented policy evaluation simulation-optimization approach is to implement a Decision Support System tool that is as flexible as possible in order to be easily adapted to national available datasets. As a consequence we implemented the two representations of tissue type incompatibility.

HLA representation. As we stated above, tissue type incompatibility is related to human leukocyte antigen (HLA) proteins that characterizes each human being DNA. HLA compatibility is then performed testing the reaction between potential donor and potential recipient tissues. Not all HLA information is relevant to evaluate tissue type compatibility. We defined the subset of HLA information that should be incorporated in the configuration module following the ones that US Organ Procurement and Transplantation Network (OPTN) currently record and use to evaluate potential tissue incompatibilities (see table 5.2).

PRA representation. A panel reactive antibody test is a blood test that specifically looks for tissue type incompatibilities measuring the level of PRAs that are present in the recipient blood type; that is the reason why PRA levels are usually recorded and monitored for all patients waiting for a kidney transplants. A patient, after the PRA blood test, can then be classified as being high, medium or low with respect to tissue sensitization. If antibody levels are high a patient is defined as highly sensitized and a transplant with a donor organ can be difficult since the chances that the body will reject the kidney increase. Panel reactive antibody levels are based on percent measurements. The percentages of PRAs in the blood play a role in determining the likelihood of finding a matching donor. A low level sensitization means that the recipient has a probability $p \in [0, 0.09]$ of being tissue type incompatible to a generic

A										
	1	2	3	9	10	11	19	23	24	25
	26	28	29	30	31	32	33	34	36	43
	66	68	69	74	80	203	210	2403	6601	6602
B										
	5	7	8	12	13	14	15	16	17	18
	21	22	27	35	37	38	39	40	41	42
	44	45	46	47	48	49	50	51	52	53
	54	55	56	57	58	59	60	61	62	63
	64	65	67	70	71	72	73	75	76	77
	78	81	82	703	804	1304	2708	3901	3902	3905
	4005	5102	5103	7801	8201					
BW										
	4	6								
DR										
	1	2	3	4	5	6	7	8	9	10
	11	12	13	14	15	16	17	18	103	1403
	1404									
DR51/52/53										
	51	52	53							
DQ										
	1	2	3	4	5	6	7	8	9	

TABLE 5.2: Recipient unacceptable antigens

donor, while for a medium level one the probability is $p \in]0.09, 0.8]$. Highly sensitized patients instead are tissue type incompatible with most of their potential donors since the incompatibility probability cover almost all possible transplants ($p \in]0.8, 1]$). As for HLA, pregnancy, previous transplants and blood transfusions may increase panel reactive antibody levels, even if these elements do not always result in a PRA increase. PRA classification can then strongly influence patients waiting times before a transplant is performed since it is more difficult to find a matching donor kidney and it is also more likely to experience organ rejection. It is known that there are two techniques, immunoadsorption and plasmapheresis that can successfully reduce panel reactive antibody; since those treatments are difficult to model in terms of simulation we decided that the first release of the simulator will not take them into account.

Crossmatch. PRA evaluation is a useful indicator of potential tissue type incompatibility even though it does not give a certain indication on specific donor-patient incompatibility. In order to have an in-depth evaluation of tissue type incompatibility a cross match test has to be performed. Cross matching is a very sensitive and final test that compares the tissue of the potential donor kidney with the one of the theoretically matched recipient. These tests enable physicians to specifically define how the recipient may respond to the transplanted kidney, evaluating if he will reject it. As a final result these compatibility tests define if there is a positive or negative crossmatch. A positive crossmatch means that the patient will reject the kidney and then the transplant should not be carried out. A negative crossmatch instead means that the transplant, if performed, should be safe.

PRA information and cross match distribution relationship have been studied in Glorie [2012]. The author, given PRA_j the PRA level of patient j, applies a probit regression model to estimate the probability of a positive crossmatch after a negative virtual one as follows: $Pr[T_{i,j} = 1 : PRA_j = \Phi(-1.5007 + 0.0170 * PRA_j)$ where $T_{i,j} = 1$ means that the crossmatch between donor i and patient $j$is positive. Figure 5.3, form Glorie [2012], shows fitted probabilities approach trend. This crossmatch evaluation technique can be enabled by the user.

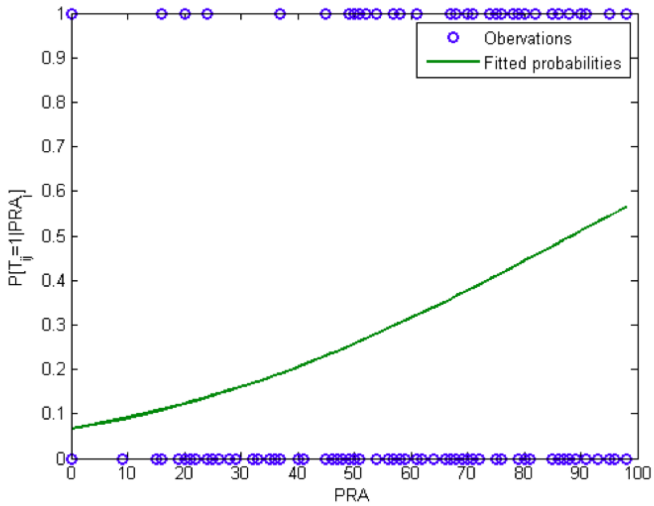


FIGURE 5.3: Fitted probabilities

5.3.1.2 Simulation and matching policy

Once the pool configuration has been defined the user can decide which is the planning policy that he would like to test, which is the behavior of the pool in terms of dropouts and define the planning horizon on which he would like to evaluate the effectiveness of the proposed policy. The user is then able to define the matching rule, the number of runs that will determine the studied running time of the experiment and enable warm up period if the user prefers not to start from an empty pool before relevant statistics start being gathered. The simulator allows two matching rule, a frequency driven and an arrival driven one. Frequency-driven option enable the user to define fixed times between two consecutive matching runs. The frequency of matching can be selected by the user so as to allow a comparison of KEP effectiveness with respect to matching frequency. Arrival-driven option triggers a matching every time a new entity enters the pool. The configuration module can then be considered as a tuning factor for both the simulation and optimization components.

Patient and donors additional behavior characteristics. As far as the simulation component concerns the user can decide if the crossmatch test, that has been described in subsection 5.3.1.1, is considered during the experimental analysis. If crossmatch testing is enabled a test is performed after each matching for the subset of entities that are candidate for transplantation. In case a positive crossmatch is found both patients and donors involved in the exchange rejoin the pool.

It is also possible to include probabilistic information to characterize the uncertainties related with the KEP. As we are modeling the dynamic evolution of a matching pool it is clear that between two consecutive matchings it could happen that patients or donors can exit the pool, temporarily or permanently. Uncertainties can be related to patient or donors length of stay in the pool due to age-related death or sickness causes. It can happen that a recipient got sick during the waiting period and can't perform a transplantation. This results in a temporary drop out of the entire entity (recipient and related donor/donors) from the pool. In a similar way donors can get sick and be temporary unavailable for transplantation. Donors sickness does not automatically result in entity dropping out from the pool. If a donor of a recipient with multiple incompatible donors gets sick only all compatible transplantations from the donor that got sick to compatible recipients will not be taken in consideration. A recipient, in case of donor sickness, becomes unavailable only if all its related donors got sick. Sickness-related unavailability is modeled as an age-related probability of being unavailable for one matching run. Patient and donors can also drop out due to permanent unavailabilities as a consequence of serious sickness or death. This kind of circumstance can lead to entity withdrawal in case of recipients involvement or if all donors related to the recipient are permanently unavailable. In case serious sickness or death occur for a donor of a multiple donor incompatible entity then only the donor becomes permanently unavailable while the remaining available donors and the related recipient remains in the pool.

Matching policy definition. The user can decide if the optimization component has to characterize the degree of compatibility between pairs or not. If weighted compatibility is selected an utility function is used to feed the solver. If no utility function is selected the optimal solution will be considered in terms of maximum number of matched pairs. In addition to weights selection, the user is also able to set up the matching parameters defining the maximum cycle and maximum chain size allowed. Given k_1 and k_2, respectively maximum length of cycle and chain, the optimization component could be theoretically set up with any value.

Besides selecting which features will be considered in the simulation the user is also prompted to define the distributions of the selected features through the Data generation module.

5.3.2 Data generation module

After configuration is complete the simulator steps into the data generation module. At this point all information that will be used during the simulation is generated and stored.

5.3.2.1 Patient and donor generation

The system starts by generating pairs of the selected types as requested in the specified distributions to match the requested ABO, PRA and age parameters.

Patient characteristics generation procedure. At first a random number uniformly distributed in $[0,1]$ is generated and used to define if the patient will be 0, A, B or AB blood type. Blood type labeling is based on incoming patients historically recorded distributions. The utilization of this information prevents a potential mis-representation of blood type assignment that can happen if the probabilities are based on the characteristic of patients in the pool in a given time stamp. In the latter case the blood type percentages would not represent the real distribution of patient blood types because they would be affected by under demanded pairs (see Sönmez and Ünver [2010]). Once the blood type is defined, tissue type generation is executed. If HLA feature is enabled the patient antigen and antibody profile is randomly generated following the blood type approach. If PRA profiling is enabled, first Low, Medium or Highly sensitization levels are randomly selected following national or local available pool statistics. After the PRA class is defined, a specific PRA sensitization level is generated for each patient. If the patient has been labeled as lowly sensitized, the PRA level will be randomly selected in $]0,0.09]$, if he is medium sensitized the random sampling is defined with $p \in]0.09,0.8]$. Otherwise the probability is defined within $]0.8,1]$ interval. Finally the age is randomly assigned using historical data. Patient generation is run only for compatible or incompatible entity generation and different input sources can be used for the two classes of entities. The number of related donors is the last generated information. The module allows the generation of multiple donors just for incompatible pairs. Once the patient characteristics have been defined the donor characteristics generation procedure is called. The procedure is run a number of times equal to the number of donors that should be related to the patient.

Donor characteristics generation procedure. As for the patient procedure at first donor blood type generation is randomly defined. In case the donor generation is related with a compatible patient a blood type compatibility check is performed and if positive the entity is confirmed. In case of blood type compatibility during incompatible donor generation we confirm that the couple is tissue type incompatible after a crossmatch test. Age distribution is then randomly assigned using historical data based on incoming donors age distribution. Different input sources can be used for compatible, incompatible and altruistic donor generation.

Simulation behavior generation procedure. In order to model pool dynamic composition, incompatible, compatible and altruistic entities are generated and join the pool with different frequencies. We describe entities arrival behaviour using a Poisson distribution with mean value equal to λ_I for incompatible entities, λ_C for compatible ones and λ_A for altruistic donors. That procedure is used in order to assign to each generated pair the time in which it enters the pool. Finally, if dropout simulation is enabled, the module randomly defines if and when the patient or the donor will dropout in a temporary or definitive way. The dropout probability is defined on historical data based on age-based survival rates.

Compatibility graph generation procedure. The generation of an incoming incompatible or compatible pair, as well as of an altruistic donor, entail an update in terms of potential transplantations. The definition of new potential transplantations is described tacking advantage of a graph representation of the kidney exchange pool (see Algorithm 1). Graph theory is a natural framework to describe the Kidney Exchange Problem since the pool can be modelled as a directed graph $G = (V, A)$ were V is composed of vertices representing compatible and incompatible pairs, vertices representing altruistic donors and vertices representing the expansion of patients with multiple donors. More precisely, let $\mathcal{P}$ be the set of all patients in the pool and let $\mathcal{D}(p)$ be the set of all donors related with patient p, for each patient donor combination $(p, d), \forall p \in \mathcal{P}, \forall d \in \mathcal{D}(p)$ we consider a different vertex in the graph. This expansion is essential to model kidney exchange programs where the weight of the arc between two elements depends on both donors, e.g., age difference between donors. Conversely A contains the arcs representing the compatibility between vertices. Each incoming entity can then be represented as an additional vertex of the graph, or more than one in case of multiple incompatible donors, and an additional set of arcs that depend on potential donor-recipient compatibilities. A crossmatch graph is also generated to represent the pairs that are found to be incompatible only after the referred test is performed. This information will be progressively discovered as the simulation advances. Although values of the generated data may not be used during the current run of the simulation by generating all data in advance the user is able to save the pool information and test/compare afterwards with the results of using different matching policies.

Data: T theoretical compatibility arcs, C crossmatched arcs : $C \subset T$, $\mathcal{P}$ set of patients in the pool
Result: T, C, $\mathcal{P}$
p: incoming patient index;
B_p: blood type of incoming patient;
Pra_p: PRA level of incoming patient;
ty_p: type of the incoming entity;
(incompatible pair = 1, compatible pair = 2, altruistic donor = 3);
D_p: number of donors of incoming patient;
$B_{p,d}$: blood type of incoming donor $d \in \mathcal{D}(p)$;
l_d label related to node of (p, d) pair;

```
foreach i ∈ P do
    if ty_p = 1 or ty_p = 2 then
        foreach k ∈ D(i) do cycle on each donor of the patient in the pool
            if B_p compatible with B_{i,k} then
                pr ← random number in [0, 1];
                if pr ≤ Pra_p then
                    for d ∈ D(p) do cycle on each donor of the incoming patient
                        T = T ∪ (l_k, l_d);
                        pr' ← random number in [0, 1];
                        if pr' ≤ Φ(−1.5007 + 0.0170 ∗ Pra_p) then
                            C = C ∪ (l_k, l_d);
                        end
                    end
                end
            end
        end
    end
    for d ∈ D(p) do
        if B_i compatible with B_{p,d} then
            pr ← random number in [0, 1];
            if pr ≤ Pra_i then
                for k ∈ D(i) do
                    T = T ∪ (l_d, l_k);
                    pr' ← random number in [0, 1];
                    if pr' ≤ Φ(−1.5007 + 0.0170 ∗ Pra_i) then
                        C = C ∪ (l_d, l_k);
                    end
                end
            end
        end
    end
end
P = P ∪ p
```

Algorithm 1: Compatibility and crossmatch arc update procedure

5.3.3 Pool management module

With all configuration and data information available the pool evolution and management is triggered. The pool management can be described by two main procedures: a before matching procedure that describes the evolution of the pool, and an after matching one that cleans the pool from the matched patients and donors. The two procedures are separated by the triggering of the Optimization module.

Before matching procedure. At each time, the engine checks if there are new pairs to include in the pool, and similarly if any of the current pairs exceeded the maximum allowed time or is temporarily unavailable updating, if necessary, the $\mathcal{P}_t$ set of active patients and related donors. Afterwards the pool management module builds a compatibility graph based on the characteristics of the pairs that currently compose the pool (see Algorithm 2).

Data: C, T, $\mathcal{P}_t$
Result: $G = (V, A)$
$\bar{\mathcal{P}} = \emptyset$ (set of patients already inserted in the pool);
$V = \emptyset$;
$A = \emptyset$;
foreach $p \in \mathcal{P}_t$ **do**
  B_p: blood type of patient p;
  Pra_p: PRA level of patient p;
  ty_p: type of entity p;
    (incompatible pair = 1, compatible pair = 2, altruistic donor = 3);
  D_p: number of donors of patient p;
  $B_{p,d}$: blood type of donor $d \in \mathcal{D}(p)$ of patient p;
  l_d label related to node of (p, d) pair;
  foreach $i \in \bar{\mathcal{P}}$ **do**
    if $ty_p = 1$ *or* $ty_p = 2$ **then**
      foreach $k \in \mathcal{D}(i)$ **do** cycle on each donor of the patient in the pool
        if $(\mathrm{l}_k, \mathrm{l}_d) \in T$ **then**
          $A = A \cup (l_k, l_d)$;
        end
      end
    end
    for $d \in \mathcal{D}(p)$ **do**
      $V = V \cup l_d$;
      if $(\mathrm{l}_d, \mathrm{l}_k) \in T$ **then**
        $A = A \cup (l_d, l_k)$;
      end
    end
  end
  $\bar{\mathcal{P}} = \bar{\mathcal{P}} \cup p$;
end

Algorithm 2: Graph generation procedure

As an example let us take the matching pool graph of Figure 5.4. The pool is composed by two incompatible entities a compatible one and an altruistic donor. Incompatible entity A has two donors (2,3) linked with recipient 1, incompatible entity B instead has three donors (5,6,7) linked with recipient 4. Incompatible entities can't perform an exchange because none of the two donors of entity A is compatible with recipient 4. Instead donors 6 and 7 of entity B are compatible with recipient 1. Vertex 9 represents a compatible pair where the donor is compatible with recipients 1 and 4 and the recipient is compatible only with donor 3. Finally vertex 10 represents an altruistic donor that is compatible only with recipient 4 (see Figure 5.4).

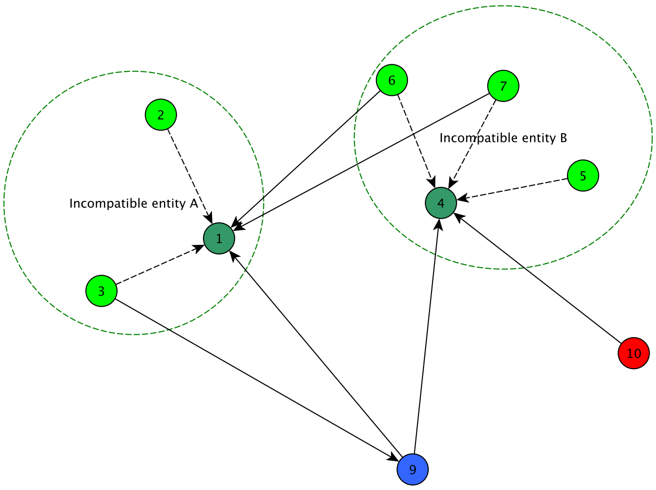


FIGURE 5.4: Matching pool graph

After matching procedure. After matching module checks which pairs were selected for matching and, if required, applies the crossmatch test. In case a matching involves an arc (l_d, l_k) that is in the crossmatch set C the cycle or chain is considered as not performed and the theoretical matching set is updated as follows $T = T \setminus (l_d, l_k)$. Otherwise the matching is accepted and the set $\mathcal{P}_t$ of active patients in the pool is updated. Afterwards the pool is updated and relevant statistics are stored.

The module then proceeds to the next time stamp. The process is repeated until the required number of runs is fulfilled.

5.3.4 Optimization module

The optimization module gets the compatibility graph of the pairs and the pair description reads its configuration parameters and calls a Mixed Integer Programming (MIP) solver to evaluate the possible matchings in the pool. The MIP is based on the cycle formulation Abraham et al. [2007] with the extensions proposed in Constantino et al. [2013]. The model considers the inclusion of both incompatible and compatible pairs, altruistic donors and patients with multiple donors.

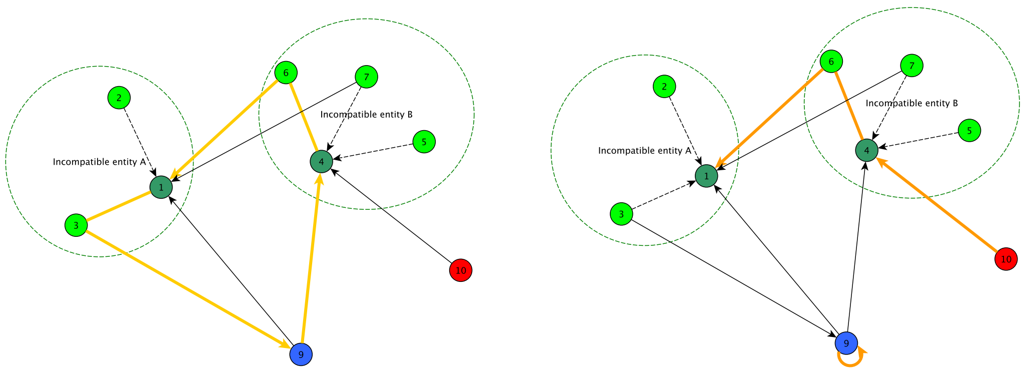
Let k denote the maximum cycle length allowed and k' denote the maximum chain length allowed. Let $\mathcal{C}(k, k')$ be the set of all cycles and chains in G of size less or equal than the referred. We define a variable z_c for each element $c \in \mathcal{C}(k, k')$ such that:

$$z_c = \begin{cases} 1 & \text{if element } c \text{ is selected for the exchange,} \\ 0 & \text{otherwise.} \end{cases}$$

Taking $V(c) \subseteq V$ as the set of vertices in element c and $w_c = \sum_{(i,j)\in c} w_{ij}$, the MIP model is given by:

$$\begin{aligned} &\text{Maximize} && \sum_{c\in\mathcal{C}(k,k')} w_c z_c && && (5.1a) \\ &\text{Subject to:} && \sum_{k\in\mathcal{D}(i)} \sum_{c:i\in c} z_c \leq 1 && \forall i \in \mathcal{P} && (5.1b) \\ & && z_c \in \{0,1\} && \forall c \in \mathcal{C}(k,k'). && (5.1c) \end{aligned}$$

Equation (5.1a) maximizes the weighted number of transplants. Constraints (5.1b) ensure that a vertex is in at most a selected cycle/chain even if the vertex is associated with a multiple donor. Note that the model includes unitary cycles representing compatible pairs. Their weight corresponds to the weight of the self-arc. Note also that when considering NEAD chains, the multiple altruistic donors resulting from a patient with multiple donors at the end of an altruistic donor chain in a previous time iteration are still related. So equation (5.1b) guarantees that only one of them may be chosen for transplant. As a consequence, in the previous example a potential matching that maximizes the number of exchanges could be the cycle [6-1, 3-9, 9-4] or the altruistic chain [10-4, 6-1] with the compatible pair having it's own transplantation surgery (see Figure 5.5).



(A) 3-cycle matching (B) 3-chain matching

FIGURE 5.5: Matching examples

5.4 Simulation results

The simulation-optimization tool has been tested under multiple scenario configurations in order to evaluate the impact of different policies regarding pool management and matching management. Since nor Italian neither Portuguese data were available, in order to fully feed the simulator characteristics, we decided to collect data from datasources of different nations. As far as blood type configuration concerns we could access to UK kidney exchange program data on incoming patients and donors. The UK data have been used to describe patients and donors blood type distribution, number of multiple donors per patient and inter arrival times of incompatible, compatible and altruistic entities. Table 5.3 shows the blood type distribution for both patients and donors. Tissue type incompatibility has ben modeled enabling the PRA configuration. As we stated in subsection 5.3.1.1 patients can be defined as Lowly, Medium or Highly sensitized. We use Dutch national data presented in Glorie et al. [2013] where 48% of patients involved in KEP programs are classified as low sensitized, 35% as Medium sensitized, and 17% of them as Highly sensitized.

Blood Type	Patients	Donors
0	57.509%	31.119 %
A	24.908%	50.349%
B	15.385%	15.385%
AB	2.198%	3.147%

TABLE 5.3: Blood type distribution

Since we were not able to get detailed information regarding UK patients age and death related probability, we used United States Renal Data System data (see Table 5.4 and 5.5).

Age Band	Donors	Patients
0-17	0%	6%
18-34	19%	26%
35-49	38%	31%
50-59	24%	20%
60-69	16%	14%
70-90	3%	3%

TABLE 5.4: Age distribution

The optimization component can potentially manage k-way cycles and k'-length chain simultaneous transplantations, where k (k') (donor-recipients) couples are involved in the exchange. It is known that due to logistical issues, a k-way cycle can be performed simultaneously only if 2k operating room are available. Since we decided to test the model for a realistic case, at most 3-way cycles and 3-chain exchanges will be tested.

Age Band	Patient death Probability
0-4	8.2
5-9	0.9
10-14	0.7
15-19	1.2
20-29	2.8
30-39	4.1
40-49	5.9
50-59	9.7
60-64	13.5
65-69	17.4
70-74	22.1
75-79	28.8
80-84	37.5
85+	49.3

Table 5.5: Death probabilities for patients with renal diseases

Matching frequency	Cycle length	Chain length	Compatible pairs enabled	Theoretical number of matched pairs before crossmatch	Total number of matched pairs after crossmatch
30	2	0	0	610	388
30	2	0	1	738	487
30	2	2	0	815	625
30	2	2	1	921	704
30	2	3	0	853	672
30	2	3	1	1003	786
30	3	0	0	854	452
30	3	0	1	1012	544
30	3	3	0	938	662
30	3	3	1	1124	771
60	2	0	0	627	397
60	2	0	1	747	493
60	2	2	0	806	608
60	2	2	1	941	716
60	2	3	0	895	688
60	2	3	1	1004	758
60	3	0	0	853	441
60	3	0	1	986	543
60	3	3	0	948	639
60	3	3	1	1098	731
90	2	0	0	616	399
90	2	0	1	750	499
90	2	2	0	806	613
90	2	2	1	984	737
90	2	3	0	901	667
90	2	3	1	1036	762
90	3	0	0	855	451
90	3	0	1	1024	545
90	3	3	0	936	612
90	3	3	1	1109	713

Table 5.6: Simulation runs under different policy and pool management configurations

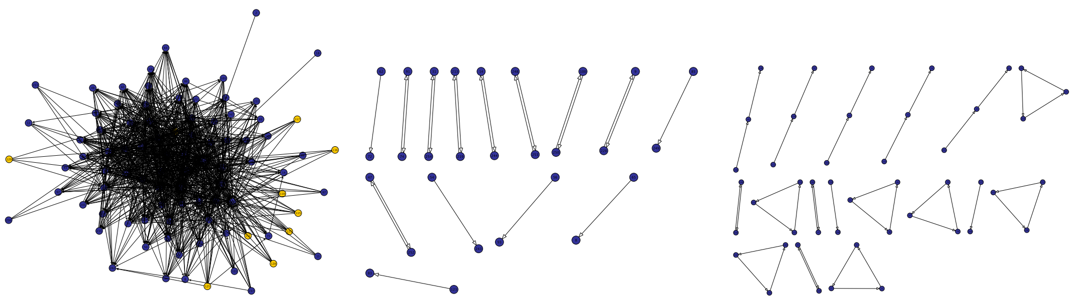
We now present preliminary results simulating dynamic kidney exchange under the model described above. Table 5.6 describes the simulation results comparing the impact of potential matches and real ones after crossmatch test is performed. The first column defines the matching frequency as the number of days between two consecutive matching runs. We tested performances of KEP comparing monthly, bimonthly and quarterly matching frequencies. The second column describes the cycle length (we tested two or three way cycles). The third column describes the chain length. If zero it means that altruistic donors are not considered, otherwise two or three length chains are considered. The fourth column describes if compatible pairs are included in the pool. The second to last column considers potential matching; that is to say the number of matches that at each matching run could have been performed if no after matching crossmatch evaluation was taken into consideration during the simulation phase. The last column takes into consideration the real number of exchanges after the crossmatch test is evaluated as described in subsection 5.3.3. We report extensive simulation results for twenty replication runs. Patient inter-arrival time is equal to 2.6 days while both compatible pairs and altruistic donors have an average inter-arrival time of 23.947 days. Let at first fix the matching frequency and compare the results under different pool configurations.

5.4.1 Monthly matching frequency

The basic kidney exchange program considers only incompatible pairs with cycle length two. Under this configuration the simulator lead to 388 transplantation over a five year planning horizon after 20 replication runs. If we consider a modification in the matching policy allowing 3-way exchanges we can observe that the number of matched entities rises up to 16,66%. The effectiveness of 3-cycle exchanges compared to 2-way ones is well known and has been already observed in Saidman et al. [2006] as well as in Abraham et al. [2007]. If we then evaluate a change in the pool configuration enabling compatibility pairs we can observe that one or two additional entities per matching run would lead to an increase of 24.84% in case of two cycle setting with respect to incompatible 2-cycle configuration, while for the 3-cycle length case the increase would be equal to 39.90%. If we compare the relative improvement of compatible inclusion fixing the number of cycles equal to 3-way we can observe a relative increase in terms of transplants equal to 20.09%. The reduced impact of the compatible pair inclusion for the 3-cycle case, if compared to the 2-cycle one, can be considered as a consequence of after matching crossmatch testing. Three way exchanges are in some way more vulnerable to crossmatch because if only one arc has a positive crossmatch none of the three matched entities will undergo surgery. Till now we are considering an arc configuration with weights all equal to one. As a consequence, the optimization component looks for the solution with the highest number of transplants; it is clear that a more robust evaluation of cycles could lead to better final solutions (see Pedroso [2013]).

If we then evaluate the impact of cycle length increase under a pool configuration that contains both compatible and incompatible pairs we can observe that the increase in terms of matched entities would be equal to 11.75%. Let then evaluate the configuration of incompatible and altruistic donor inclusion in the pool with no compatible pairs joining the exchange. Table 5.6 shows that in case of 2-cycle/2-chain configuration the increase in terms of overall number of exchanges is equal to 61.08% that can rise up to 70.77% if we increase both cycle and chain length. The increased level of exchanges is at first sight unexpected since the inter-arrival time of compatible and altruistic donors is equal. The gap in terms of compatible versus altruistic donors inclusion can be explained as the result of three factors. At first we must say that, if we configure the optimization module for a maximization in terms of number of transplants, all compatible pairs will remain in the pool for at most one matching run. It is possible to assert this since or the compatible entity will be contained in a 2/3 way cycle or a self transplant is executed. A solution that does not take into consideration the self-transplant if no other exchange is possible is considered a suboptimal one by the optimization module since at least one more transplant is possible. Altruistic donors instead does not leave the pool unless a chain is started. The second factor is related to NEAD chains modeling that consider the donor of the last patient in the chain as an altruistic one for the next matching run. The last factor is related to the different impact of positive crossmatch on chains and cycles. Chains are less affected by this phenomenon than cycles since, unless the positive crossmatch is between the altruistic donor and the first patient of the chain, part of the proposed transplants can be preserved.

Lastly we evaluate the results with both compatible pairs and altruistic donors enabled. For the 2-cycle, 2-chain configuration the increase in terms of number of final transplantations is equal to 81.64% and a switch to 3-cycle, 3-chain configuration would lead to an additional 6.02% increase (example in Figure 5.6).



(A) Graph at third monthly matching (B) Matched entities under 2-cycle 2-chain policy (C) Matched entities under 3-cycle 3-chain policy

FIGURE 5.6: Comparison of matched entities with incompatible pairs with multiple donors and altruistic donors enabled

5.4.2 Bimonthly matching frequency

As we did for 30 days matching frequency it is possible to see that if only incompatible pairs are considered, a 3-cycle exchange policy can lead to a 10.90% increase of final transplantations if compared to the 2-cycle one. If compatibility pairs are enabled the increase in terms of final transplantations would be equal to 24.16% in case of 2-cycle setting and to 36.68% in case of 3-cycle one. We can then assert that doubling the time between two consecutive matching does not affect the behavior of the pool. If altruistic donors are included in the pool with no compatible pairs joining, the 2-cycle 2-chain configuration lead to an increase in terms of overall number of exchanges equal to 53.07% for the 2-cycle, 2-chain configuration and to 60.73% for the 3-cycle, 3-chain one. Lastly we evaluate the results with both compatible pairs and altruistic donors enabled. For the 2-cycle, 2-chain configuration the increase in terms of number of final transplantations is equal to 80.27% and a switch to 3-cycle, 3-chain configuration would lead to an additional 2.09% increase.

5.4.3 Quarterly matching frequency

At last the lowest matching frequency has been evaluated. As for 30 and 60 days matching frequency an increase of 12.98% of final transplantations can be observed in case the maximaum cycle length rises from two to three. The inclusion of compatibility pairs would be equal to an increase of 25.04% in case of 2-cycle setting and to 36.42% in case of 3-way one, while the inclusion of altruistic donors would increase the number of matchings up to 53.31% for the 3-cycle, 3-chain one. If both compatible and altruistic donors can join the pool in the 2-cycle, 2-chain configuration the increase in terms of number of final transplantations is equal to 84.64%.

If we then compare the exchange program performances under different frequency configurations we can observe that an increase in terms of time between two consecutive matchings does not substantially affect the number of matching entities. Theoretically the number of overall exchanges should increase together with the increase of the matching pool. That is to say that longer time between matching can lead to bigger pool an then to an increase in the number of potential exchanges. That should be true if we assume that once a pair enter the pool he can leave it only if a transplant has been performed. Nevertheless patients and donors dropouts were usually overlooked in most of the reviewed approaches. Our model instead takes into consideration patients mortality and if we analyze the dynamic impact of this element on the exchange pool we must consider that longer times between two consecutive matchings could lead to higher probability that a patient dropout.

5.4.4 Type distribution of matched entities

In the previous section we analyzed the impact of different policies in terms of overall increase of matched entities. It is important to say that the inclusion of altruistic donors and of compatible pairs in kidney exchange programs has to be evaluated in terms of improvement for incompatible pairs. That is to say that we can assert that a matching policy is more effective than another if we can measure an improvement in terms of number of incompatible pairs matched. Table 5.7 shows the distribution of matched entities under different policy configurations. We can then observe that the inclusion of compatible pairs for the monthly matching 2-cycle policy can guarantee an increase of 8% in terms of incompatible matched pairs, while the inclusion of altruistic donors leads to a 16% rise. The most effective policy with a 30.2% increase is the 3-cycle 3-chain matching with compatible pairs and altruistic donor inclusion. The improvements remain stable with the increase of the time between two consecutive matchings.

Matching frequency	Cycle length	Chain length	Compatible pairs enabled	Incompatible	Compatible	Altruistic
30	2	0	0	388	0	0
30	2	0	1	419	68	0
30	2	2	0	451	0	174
30	2	2	1	453	68	184
30	2	3	0	454	0	218
30	2	3	1	479	67	239
30	3	0	0	452	0	0
30	3	0	1	474	70	0
30	3	3	0	476	0	186
30	3	3	1	505	70	197
60	2	0	0	397	0	0
60	2	0	1	427	67	0
60	2	2	0	444	0	164
60	2	2	1	462	70	185
60	2	3	0	473	0	215
60	2	3	1	477	66	215
60	3	0	0	441	0	0
60	3	0	1	473	70	0
60	3	3	0	472	0	166
60	3	3	1	492	68	171
90	2	0	0	399	0	0
90	2	0	1	429	70	0
90	2	2	0	480	72	185
90	2	2	1	480	72	185
90	2	3	0	471	0	196
90	2	3	1	485	68	210
90	3	0	0	451	0	0
90	3	0	1	475	69	0
90	3	3	0	461	0	151
90	3	3	1	497	71	146

Table 5.7: Number of matchings per entity type

5.4.5 Blood type analysis

The definition of number of matchings under different policy assumptions can be considered as the first factor of analysis while considering kidney exchanges. Nevertheless, some useful insights can be reached analyzing blood type distribution of matched paris. Table 5.8 shows the distribution of blood type among the incompatible patients that have been matched. It is then possible to observe that matched patients blood type distribution, if compared with the one of table 5.3, is relatively composed by an higher number of AB type patients and a reduced number of 0 type ones. This phenomenon has been observed by Sönmez and Ünver [2010] that identifies 0-A, 0-B and 0-AB as under demanded pairs. Observing the blood type distribution of donors, a 0-type patient has a probability equal to 68.88% of being under demanded if he has only one related donor. While if we analyze the AB-blood type patients we can justify the relative increase in terms of blood type composition due to the fact that AB-0, AB-A, AB-B pairs can be classified as overdemanded since the probability of an AB-type patient of having a related donor with 0, A or B blood type is equal to 96.85%. If we analyze the relative blood type distribution under different matching frequencies we can observe that no real improvement can be reached increasing the time between matchings. That is also true if we analyze the impact of compatible pairs inclusion or increase of cycle length. We can then conclude that the simulation of kidney exchange program under a maximum number of transplant optimization policy does not take into account, or properly evaluate, the presence of over demanded and under demanded pairs. It is then interesting to observe that the possibility of tuning the weights as an input parameter during the configuration phase is a feature of great importance in terms of policy evaluation.

Matching frequency	Cycle length	Chain length	Compatible pairs enabled	Blood O	Blood A	Blood B	Blood AB
30	2	0	0	41.53%	37.20%	18.13%	3.15%
30	2	0	1	44.12%	34.30%	18.11%	3.47%
30	2	2	0	42.76%	35.15%	18.86%	3.24%
30	2	2	1	45.09%	33.11%	18.44%	3.36%
30	2	3	0	42.51%	35.06%	19.30%	3.13%
30	2	3	1	45.75%	32.33%	18.89%	3.03%
30	3	0	0	42.21%	35.18%	19.62%	3.00%
30	3	0	1	44.11%	33.33%	19.39%	3.17%
30	3	3	0	43.66%	33.74%	19.83%	2.76%
30	3	3	1	46.27%	31.90%	18.92%	2.91%
60	2	0	0	42.75%	35.87%	18.37%	3.01%
60	2	0	1	44.79%	34.23%	17.64%	3.34%
60	2	2	0	42.94%	35.17%	18.94%	2.95%
60	2	2	1	44.62%	33.82%	18.55%	3.00%
60	2	3	0	42.70%	35.13%	19.23%	2.94%
60	2	3	1	44.92%	33.62%	18.72%	2.74%
60	3	0	0	41.85%	35.40%	19.78%	2.97%
60	3	0	1	44.56%	33.03%	19.35%	3.05%
60	3	3	0	43.26%	33.71%	20.20%	2.84%
60	3	3	1	44.76%	33.22%	19.35%	2.66%
90	2	0	0	41.92%	36.35%	18.64%	3.09%
90	2	0	1	45.92%	34.04%	17.06%	2.97%
90	2	2	0	43.24%	34.82%	19.01%	2.93%
90	2	2	1	44.95%	33.73%	18.23%	3.08%
90	2	3	0	42.39%	34.72%	19.66%	3.23%
90	2	3	1	44.90%	33.55%	18.73%	2.83%
90	3	0	0	41.50%	35.17%	20.08%	3.26%
90	3	0	1	43.81%	33.32%	19.95%	2.92%
90	3	3	0	42.27%	34.18%	20.42%	3.13%
90	3	3	1	45.30%	32.30%	19.42%	2.97%

Table 5.8: Blood type distribution of incompatible matched patients

5.4.6 Waiting time analysis

Blood type distribution and the probability of receiving a kidney can also affect the waiting time of patients. Table 5.9 shows the patients waiting time for each class of blood type. As suggested by the previous analysis on blood type distribution, it is possible to observe a strong difference between the waiting time of 0-type patients and the one of AB-type ones. AB-blood type patients on average has to wait 63% less than 0-type ones. This phenomenon is particularly stressed for the monthly frequency policy, while an increase in the number of days between matching reduces the gap between the two types of patients. The less penalizing policy can then be reached with the inclusion of compatible pairs and altruistic donors under a 3-cycle/3-chain policy. The comparison between policies under a waiting time point of view strongly enforces the importance of increasing the length of both cycles and chains. If we focus on 0-blood type patients and we take the incompatible pool configuration with 2-cycle exchanges and monthly matching policy as the baseline scenario we can observe that the inclusion of compatible pairs would reduce the waiting time before transplant of 6.62%, while the inclusion of altruistic donors under a 2-chain configuration seems to help A, B and AB patients to the detriment of 0-type ones. If we observe the overall waiting times and not just the ones related to the under demanded pairs, we can see that the increase of time between two consecutive matchings has a general negative impact on A, B and AB patients.

Matching frequency	Cycle length	Chain length	Compatible pairs enabled	Blood O (Days)	Blood A (Days)	Blood B (Days)	Blood AB (Days)
30	2	0	0	303.94	84.85	195.77	66.50
30	2	0	1	283.82	83.53	194.26	76.18
30	2	2	0	318.00	37.18	130.62	36.79
30	2	2	1	293.60	39.66	156.71	35.70
30	2	3	0	235.79	40.81	107.61	38.44
30	2	3	1	212.46	41.38	107.00	44.72
30	3	0	0	201.50	71.82	116.20	61.05
30	3	0	1	180.49	68.28	126.17	76.87
30	3	3	0	207.03	40.29	70.38	39.72
30	3	3	1	192.37	42.53	77.67	49.61
60	2	0	0	290.36	106.04	218.16	84.39
60	2	0	1	278.03	101.65	213.61	103.88
60	2	2	0	301.79	62.50	147.78	58.70
60	2	2	1	280.64	64.54	177.33	57.87
60	2	3	0	238.79	66.78	112.69	62.15
60	2	3	1	221.74	67.47	121.17	74.62
60	3	0	0	213.38	97.58	134.12	95.37
60	3	0	1	191.73	95.46	131.21	100.05
60	3	3	0	216.11	69.88	92.50	63.53
60	3	3	1	204.44	71.61	101.56	73.86
90	2	0	0	289.84	122.34	211.51	100.28
90	2	0	1	273.82	123.94	229.64	116.64
90	2	2	0	290.55	83.46	162.71	81.81
90	2	2	1	283.99	87.70	163.29	85.35
90	2	3	0	243.84	93.26	144.97	84.86
90	2	3	1	234.34	93.36	150.89	106.51
90	3	0	0	225.52	121.56	152.33	118.95
90	3	0	1	226.07	125.79	159.64	141.25
90	3	3	0	231.03	100.63	117.41	92.66
90	3	3	1	236.82	105.25	123.52	103.63

TABLE 5.9: Waiting times of incompatible matched patients

5.5 Conclusions

In collaboration with INESC TEC, an R&D center in Porto, Portugal, we implemented a simulation-optimization tool that gives the possibility to a policy maker to test different configurations regarding matching frequency, matching characteristics and pool characteristics. Discrete Event Simulation modeling gives the possibility of representing the key elements of KEP in a simple and straightforward way through the definition of entities and classes of events that can occur during the dynamic evolution of the program. Most of the papers related to KEP model only incompatible pairs, but it is known that nowadays KEP programs have evolved from this first definition of exchange pool. In the UK, as well as in the USA, exchange pools are composed not only by incompatible pairs but also by compatible ones and by altruistic donors. The idea of the implemented simulator is to take into consideration the widest typology of actors that can take part to an exchange pool. Two main classes of events, the arrival of new entities into pool and the matching, can describe KEP dynamics. As we are modeling a policy making DSS tool, entity generation and matching rules can be modified in order to evaluate exchange pool performances with different settings.

As far as entity arrival concerns, the tool allows policy makers to set the classes of entities that will be simulated. It is possible to simulate exchange pools with all possible combinations of incompatible entities, compatible entities and altruistic donors in order to evaluate potential benefits that the inclusion of new classes of actors can bring to incompatible recipients.

Matching rounds can be performed following different rules in order to test multiple policies related to kidney exchange. The simulator invokes an optimization code that, given graph characteristics as input, returns information about matched vertexes. The optimization code can be tuned in order to define the kind of matching that are enabled. We defined an input infrastructure that easily allows the modification of pool characteristics so as to build a tool that can be used to test exchange policies for different regional and national programs. We collected a set of data in order to test the proposed approach under different configurations and we analyzed the results quantifying exchange performance increase both in terms of pool characteristics and of matching policy. We then observed that a lower matching frequency does not increase the number of final transplant because of dynamic dropout modeling. In addition we showed that the inclusion of altruistic donors strongly increases the number of exchanges more than compatible pairs one. Nevertheless the presented work has to be considered just as a first step towards a more complete simulation tool that can simultaneously model different national exchange programs and evaluate their integration in a unified european matching pool. In addition an overtime optimization model has been developed and it will be used in order to compare the implemented policies with the theoretical maximum number of exchanges that could be reached under omniscience hypothesis.

Bibliography

David J Abraham, Avrim Blum, and Tuomas Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 295–304. ACM, 2007.

Itai Ashlagi, Patrick Jaillet, and Vahideh H Manshadi. Kidney exchange in dynamic sparse heterogenous pools. *arXiv preprint arXiv:1301.3509*, 2013.

Pranjal Awasthi and Tuomas Sandholm. Online stochastic optimization in the large: Application to kidney exchange. In *IJCAI*, volume 9, pages 405–411, 2009.

M Beccuti, V Fragnelli, G Franceschinis, and S Villa. Dynamic simulations of kidney exchanges. In *Operations Research Proceedings 2010*, pages 539–544. Springer, 2011.

Yanhua Chen, J Kalbfleisch, Yijiang Li, P Song, and Yan Zhou. Computerized platform for optimal organ allocations in kidney exchanges. *Proc. BIOCOMP*, 11, 2011.

Miguel Constantino, Xenia Klimentova, Ana Viana, and Abdur Rais. New insights on integer-programming models for the kidney exchange problem. *European Journal of Operational Research*, 2013.

Miguel Constantino, Joao Pedro Pedroso, Abdur Rais, Nicolau Santos, Paolo Tubertini, and Ana Viana. Kidney exchange problem: a simulation-optimization approach. Technical Report OR 14-8, LabOR DEI - Università di Bologna, 2014.

John P Dickerson, Ariel D Procaccia, and Tuomas Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *AAAI*, 2012a.

John P Dickerson, Ariel D Procaccia, and Tuomas Sandholm. Optimizing kidney exchange with transplant chains: Theory and reality. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 711–718. International Foundation for Autonomous Agents and Multiagent Systems, 2012b.

John P Dickerson, Ariel D Procaccia, and Tuomas Sandholm. Failure-aware kidney exchange. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 323–340. ACM, 2013.

Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of mathematics*, 17(3): 449–467, 1965.

KM Glorie. Estimating the probability of positive crossmatch after negative virtual crossmatch. Technical report, Erasmus School of Economics (ESE), 2012.

Kristiaan M Glorie, Marry de Klerk, AP Wagelmans, Joris J van de Klundert, Willij C Zuidema, FH Claas, and Willem Weimar. Coordinating unspecified living kidney donation and transplantation across the blood-type barrier in kidney exchange. *Transplantation*, 2013.

Yijiang Li, Jack Kalbfleisch, Peter Xuekun Song, Yan Zhou, Alan Leichtman, and Michael Rees. Optimization and simulation of an evolving kidney paired donation (kpd) program. 2011.

Joao Pedro Pedroso. Maximizing expectation on vertex-disjoint cycle packing. Technical report, DCC, Faculdade de Ciencias, Universidade do Porto, 2013.

Organ Procurement and Transplantation Network. About cpra for patients. URL http://www.http://optn.transplant.hrsa.gov/resources/allocationcalculators.asp?index=75.

Felix T Rapaport. The case for a living emotionally related international kidney donor exchange registry. In *Transplantation proceedings*, volume 18, page 5, 1986.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *The American economic review*, pages 828–851, 2007.

Susan L Saidman, Alvin E Roth, Tayfun Sönmez, M Utku Ünver, and Francis L Delmonico. Increasing the opportunity of live kidney donation by matching for two-and three-way exchanges. *Transplantation*, 81(5):773–782, 2006.

Dorry L Segev, Sommer E Gentry, J Keith Melancon, and Robert A Montgomery. Characterization of waiting times in a simulation of kidney paired donation. *American Journal of Transplantation*, 5(10):2448–2455, 2005.

Tayfun Sönmez and MU Ünver. Market design for kidney exchange. *Oxford Handbook of Market Design, Oxford University Press (to appear)*, 2010.

M U Ünver. Dynamic kidney exchange. *The Review of Economic Studies*, 77(1): 372–414, 2010.

Stefanos A Zenios. Optimal control of a paired-kidney exchange program. *Management Science*, 48(3):328–342, 2002.

Stefanos A Zenios, Glenn M Chertow, and Lawrence M Wein. Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research*, 48(4): 549–569, 2000.

Chapter 6

Emergency Room Management in Lazio, Italy [1]

6.1 Introduction

The Department of Epidemiology of the Regional Health Service of Lazio, Italy (DEP-Lazio in the following), a regional center for Health monitoring and management, is currently involved in a project that aims at defining optimal allocation policies of patients to regional hospital network facilities. The reorganization of health centers in order to deliver services in an effective way safeguarding economic sustainability is a topic of increasing importance for Regional Health Services in Italy. In recent years several inputs have been given, through financial laws, to reorganize hospitals infrastructure in order to decrease inefficiencies. Reorganization policies can be considered, from a strategic point of view, as composed by two main decision elements: the definition of the subset of hospital facilities that should be active within the regional territory and the allocation of demand of services to active facilities. Since the reorganization of a Regional Health system in terms of facility location and service allocation is a task of great complexity Regional mangers decided to focus their attention on Emergency Departments (ED). EDs are a crucial access point to hospital network facilities and as a consequence their management is a critical factor in order to improve system effectiveness and efficiency. In Italy it is possible to state that the role of ED is even more important than in other European countries since, in addition to real emergency and urgency services, they have to face a set of demands that should instead be managed by Primary care units or by General Practitioners.

Emergency Department characteristics. Emergency Departments can be defined as an health facility that is dedicated to the management of emergency and

[1]This chapter is based on Technical Report OR 14-9 (see Leo et al. [2014])

urgency treatments, that is to say to that spontaneous or traumatic pathological conditions that need to be treated within a short period of time. Emergency activities are, for their own nature, non elective and patients can reach ED facilities both by their own or with the support of an ambulance. Due to the impossibility of planning patients arrival, EDs have to provide an initial treatment for a wide number of diseases some of which can be life-threatening. Since the set of patients that ask for treatments is heterogeneous from the pathological point of view, the admission of patients is driven by a priority-based policy. The stochastic nature of arrival times and of pathological conditions can have a strong impact on workload and as a consequence on patient waiting times and quality of care. It is then fundamental that priority assignment is properly managed in order to meet patients' needs according to their critical condition. The process of assigning priorities to patients is defined as triage and it is usually coded at a regional or national level. Triage is a set of procedures that ensure, in the best possible way, that patients with a more critical condition are admitted before the others. The priority level is usually represented by a color code (white, green, yellow and red) that defines the increasing need of care. For each patient the priority is usually defined just after the arrival by a dedicated operator. Triage procedure definition is then fundamental to guarantee an immediate care for the patient, to identify the priority level and the medical area that may treat him and, ranking lower priority patients, to reduce waiting times. Triage activities can directly address the patient to the most appropriate hospital ward in case of complex treatments, for less serious ones the patient can be directly treated by Emergency Room (ER) physicians and discharged. It is then important for health managers to plan ERs so as to meet a set of objectives that can be in some cases conflicting. At first it is fundamental to guarantee quality of care that is composed by treatment timeliness, according to the patient health condition, and appropriateness, according to the patient pathological condition. On the other hand the cost sustained to provide services has to be reduced as much as possible safeguarding a minimum standard of care.

Literature Review. Operations Research has been widely applied to study ER management issues such as capacity planning and patient flows using both optimization and simulation techniques. Literature case studies can be classified according to the set of decisions that are taken into consideration, including capacity planning, staff scheduling and general planning for future development of the facility. Pure capacity planning case studies evaluate the impact of resource resizing on patients waiting times. It is then important to evaluate which is the degree of complexity of a comprehensive simulation model. As an example in Connelly and Bair [2004] the authors consider triage, prioritization and several staff level types as well as imaging studies, laboratory studies, physical examination, nursing activity, consultations, and bedside procedures. The model however does not consider technical resources reducing the potential analysis of supply shortages. In Bagust et al. [1999] the authors show how capacity planning can provide an efficient patient flow by calculating the maximum occupancy level of beds. In Wang et al. [2013] the authors define an analytical model to describe patient flows in ER department taking into consideration scarce resources such as medical doctors, nurses, beds and diagnostic machines. The model is used to evaluate the impact of resource resizing policies. In a similar way in Komashie and Mousavi [2005] the resizing of different resources is compared in order to identify which is the one that mainly influences ER performances. Patients arrival pattern can be also simulated in order to level the peak of resource utilization, leading to a significantly better planning of staff and resources Sinreich and Marmor [2005]. In a similar study Sinreich and Marmor [2004] arrival analysis allowed a reduction of patient turnaround times. ED capacity management can be also analyzed from a different perspective through the evaluation of how budget restrictions and workforce reduction can be faced preserving operational performances Sinreich and Jabali [2007]. In that case study patient flow patterns are fixed and the main goal of the problem is to evaluate how staffing management can influence waiting times.

It is clear that emergency department performances cannot be improved only by means of resource resizing; advanced prioritization models as well as new organizational designs can turn out to be more effective than simple capacity planning. In Konrad et al. [2013] the authors evaluate the introduction of the so-called split-flow concept that is an emerging approach to manage ED processes by a split of the patient flow according to their acuity and enabling parallel processing. The model, applied to a real ED, aims at reducing patients waiting times and system congestion. In Cochran and Roche [2009] a new prioritization model for patients is evaluated taking in consideration patient acuity mix, arrival patterns and volumes and trying to minimize the walk-away for patients waiting for a long time. In Kuban Altιnel and Ulaş [1996] simulation also proved to be of great potential for the evaluation of future expansion of an ED by increasing the understanding of the processes involved. An integration of simulation and optimization techniques is presented in Yeh and Lin [2007]to reduce patient queuing time. Modeling the complex behavior of an ED is a challenging task, due to interaction of human and physical resources. Medical staff, for example, is rarely dedicated to one patient or task. Instead, the staff treats several patients at a time while waiting for other processes. This diversity of process interaction can be described as multitasking, a common feature of ED operations even if rarely considered in planning models (see, e.g., Gunal and Pidd [2006]).

Till now we focused our attention on Emergency Department planning considering this organizational unit as a unique component that is externally influenced only by patient arrivals. It is clear that ED inflow is strongly related to the definition of catchment areas since this department and Hospitals in general usually cover the health needs of a subset of the local population. Through 'covering' we mean that a specific (regional) population cluster has as a reference point for health needs a specific hospital that is usually defined on a distance basis. It is then clear that if we widen the focus of analysis we can develop capacity plans for hospitals and EDs taking in consideration the fact that a reorganization can strongly influence the volume of activities and as a consequence system performances both in terms of patient outcomes and quality of service. As an example, in Chu and Chu [2000] the authors propose a modeling framework to analyze the supply and demand matching of public hospital beds addressing the planning issues of hospital locations and service allocations, which include new service distribution as well as existing service redistribution. In Branas et al. [2000] an optimization model is formulated using integer programming and heuristics, the goal of the case study being to maximize coverage of severely injured patients by locating trauma centers and aeromedical depots. Finally, in Harper et al. [2005] the authors propose a discrete-event geographical location/allocation simulation model for evaluating various options for the provision of services including the location of the service centers, service capacities, geographical distribution of patients, and ease of access to the health services.

Contribution. As already discussed triage is the first activity that is performed when a patient reach the ER, this means that the classification of walk-ins and of patients arrived with an ambulance can only affect the care pathway within the hospital structure without taking in consideration the possibility that a better quality of care or a shorter waiting time could have been reached if the patient would have been sent to another ER.

The objective of the present study is to develop an hybrid model that considers both ED workload and service allocation, evaluating what could be the impact of a remote triage management that, anticipating the patient classification, can address population requests to the first-aid structure, thus assuring the best possible service level. In particular, the final objective of the case study is to develop an allocation policy for Emergency Room requests in order to maximize quality of care and service timeliness. In order to develop a regional allocation approach we must suppose that all requests can be filtered at a regional level. That is to say that walk-in or ambulance referral that have not been screened by the triage management center are not accepted. Clearly, this is only an hypothetical scenario that is, however, very useful to define a reference solution (as well as a reference methodology) in terms of service quality (to be defined below), so as to evaluate, in comparison, new and more sophisticated allocation policies. In other words, the current case study establishes a benchmark solution with respect to which the cost of a completely decentralized and loosely planned allocation is computed.

As mentioned, the first required step is to appropriately define service quality indicators, and we do distinguish two in particular.

1. **Travel time.** An initial version of the model can evaluate how to assign requests to Emergency Rooms in order to minimize the overall time needed to reach the First aid facility. The travel time between the place where the call is made and the hospital is an element of paramount importance since, if the patient has compromised vital functions (consciousness, respiration, heart rate , shock) and is in life threatening conditions, the the time needed to reach the closest hospital can strongly impact on the probability to survive.

2. **Waiting time.** As a second factor the workload of the emergency room, quantifiable as "waiting time", has to be evaluated. Using data from the Health Emergency Information System , it is possible to empirically estimate the workload for the hospital for each day of week and time of day, in this way, the choice of the structure, may be evaluated considering penalty coefficients "proportional " to the estimated waiting time. Finally, each hospital can be classified according to a penalty coefficient based on the quality of care provided, estimated by the indicators of outcome and process of the Regional Program for the Evaluation of Outcomes (see Fusco et al. [2012] and Renzi et al. [2012]). If, at the time of the request, the patient's symptoms are not clearly defined, a summary measure of hospital quality of care (taking into account some of the most relevant indicators and proceeding to their synthesis) is applied. Otherwise, if a patient has more defined symptoms, the penalty coefficient may be applied using specific indicators according to the pathological area.

Chapter Organization. In Section 6.2 we discuss the details of the problem and we introduce the required notation and definitions. In Section 6.3 a Mixed Integer Linear Programming approach (MIP) is proposed and several properties and relaxations are discussed. In Section 6.4 we extensively discuss computational experiments performed by solving the MIP model on real-world instances provided by DEP-Lazio. Finally, in Section 6.5 we draw some conclusions and discuss the use of the proposed optimization approach.

6.2 Notation and Definitions

Given a positive integer τ, the time horizon of our analysis is modeled by discrete ordered set $T := \{1, \ldots, \tau\}$, whose elements represent time slots. Let U be the set of enumeration districts in which the territory under the authority of Lazio Region is partitioned. Let $V \subseteq U$ be the subset of districts in which an emergency room department is located. Let $d(u, v)$ the expected time duration of a trip from $u \in U$ to $v \in V$. We assume $d(u, v)$ is constant over T. Let F be a set of certain first-aid medical treatments that can be supplied by healthcare centers.

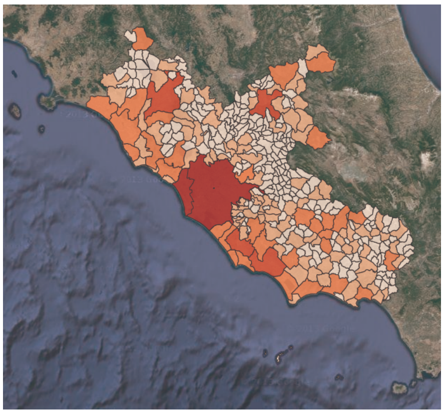


FIGURE 6.1: Municipal districts of Lazio Region.

Emergency room departments. Let S be the set of emergency room departments operating under the authority of Lazio Region. Each $s \in S$ is modeled by a quadruple composed by the following elements: (i) $v_s \in V$ is the district in which s is located; (ii) $F_s \subseteq F$ is the subset of specializations that s can offer; (iii) $w_s : \mathbb{R} \to \mathbb{R}$ is a function returning the expected time a patient has to wait at s before receiving first-aid service; (iv) $q_s \in \mathbb{R}^{|F_s|}$ reports the quality of service for each specific medical treatment in F_s, according to Lazio Region Evaluation Program for medical operations results (PReValE).

Triage codes and pathologies of interest. Let C be the set of emergency room codes that can be assigned by triage diagnosis. Let P be a subset of pathologies that are known to be significant within emergency room management. Each $p \in P$ is characterized by a maximum estimated time $t^{\max}(p)$ that a person suffering p could wait without medical control. Let $f_p \in F$ a specific medical treatment for treating p.

Our analysis focuses on three pathologies, namely ST Elevation Myocardial Infarction (STEMI), Acute Myocardial Infarction (AMI) and Femoral Fracture (FF), which have a remarkable impact on Lazio healthcare management system.

First-aid requests. Let R be the set of first-aid requests arising on the Lazio Region area, during time horizon T. Each $r \in R$ is modeled by a quadruple (u_r, t_r, c_r, p_r), where: $u_r \in U$ and $t_r \in T$ are resp. the district and the time slot in which r arises, $c_r \in C$ is the expected triage code associated with r, namely its presumed emergency level, and p_r is the expected pathology, diagnosed in terms of subjective symptoms.

First-aid requests assignment problem. Let us given a set of emergency room departments S and a set of first-aid requests R arising from a defined geographic area, during a fixed time horizon T. An assignment of first-aid requests to emergency room departments is feasible if the following conditions are satisfied:

a) each request $r = (u_r, t_r, c_r, p_r) \in R$ is assigned to exactly one emergency room $s = (v_s, w_s, F_s, q_s) \in S$;

b) s supplies suitable medical treatment for p_r;

c) the expected duration of the trip from u_r to v_s does not exceed the maximum estimated time for avoiding life-threatening, i.e., $t^{\max}(p_r)$.

The goal is looking for feasible assignments that allow to maximize the overall benefit, in terms of efficiency and effectiveness of supplied emergency room services.

6.3 A Mixed Integer Programming approach

In this section we introduce a basic MIP model for the problem (Section 6.3.1) and we then discuss some interesting and useful mathematical properties (Section 6.3.2) and relaxations (Section 6.3.3).

6.3.1 The Basic MIP model

In oder to define the backbone of the basic MIP model we need assignment variables and constraints. More precisely, we introduce a binary variable x_{rs} with $r \in R$, $s \in S$ for each assignment, such that $x_{rs} = 1$ if and only if request r is assigned to emergency room s. Thus, the following constraints guarantee a feasible assignment.

$$\sum_{\substack{s \in S: \\ f_{p_r} \in F_s}} x_{rs} = 1 \qquad \forall r \in R \tag{6.1}$$

$$x_{rs} \leq 1 - \min\left\{1, \left\lfloor \frac{d(u_r, v_s)}{t^{\max}(p_r)} \right\rfloor \right\} \qquad \forall r \in R, \forall s \in S \tag{6.2}$$

$$x_{rs} \geq 0 \qquad \forall r \in R, \forall s \in S \tag{6.3}$$

Let observe that (6.1) makes each request r to be assigned to exactly one emergency room department that is able to supply the required medical treatment; thus, both conditions (a) and (b) are satisfied. Moreover, (6.2) avoids any assignment that does not respect condition (c).

Evaluating efficiency. Our model allows to evaluate the efficiency of each assignment in terms of waiting time. We identify two time components, namely the expected time a patient takes to reach the location of the assigned emergency room and the expected time a patient has to wait for receiving medical treatment. The first cost contribution is easily given by $d_{u_r v_s}$, whereas the second is given by function w_s. In particular, w_s allows to estimate the needed waiting time for processing all first-aid

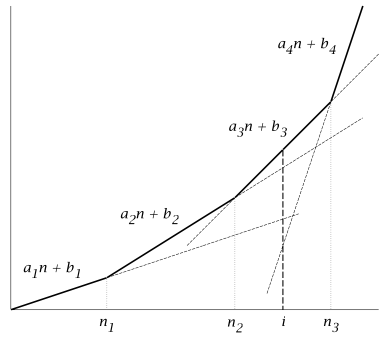


FIGURE 6.2: Example of piecewise linear convex function.

requests assigned to s. In our analysis, we model w_s as a (convex) piecewise linear function because we aim at penalizing emergency room overload situations.

Definition 6.1. Given k_s nonnegative integers $0 < n_1 < \cdots < n_{k_s}$ for each $s \in S$, let

$$w_s(n) := \begin{cases} a_h^s n + b_h^s & n_h \leq n < n_{h+1} \ \forall h = 1, \ldots, k_s - 1 \\ a_{k_s}^s n + b_{k_s}^s & n \geq n_{k_s} \end{cases} \tag{6.4}$$

such that the following conditions hold:

$$a_h^s < a_{h+1}^s \qquad \forall h = 1, \ldots, k_s - 1 \tag{6.5}$$

$$b_1^s = 0 \tag{6.6}$$

$$b_{h+1}^s = b_h^s + \left(a_h^s - a_{h+1}^s\right) n_h. \qquad \forall h = 1, \ldots, k_s - 1 \tag{6.7}$$

In our study, parameters a_h^s have been estimated by analyzing real waiting time data provided by DEP-Lazio. Moreover, since w_s is a piecewise convex function, it is easy to check that the following property holds:

$$w_s(n) = \max_{h=1,\ldots,k_s} \{a_h^s n + b_h^s\}, \tag{6.8}$$

i.e., for any value of (the number of patients waiting for medical treatment) n the slope of the linear segment such that $n_i \leq n < n_{i+1}$ is the leading one in (6.8). This is depicted in Figure 6.2.

Let $\bar{z}$ be the average waiting time of all emergency room departments of Lazio. In order to balance emergency room departments workload, we introduce a fixed (overtime) cost λ for each emergency room whose waiting time w_s exceeds a constant threshold $\bar{z}$. In our computational experience, we discuss results for different value of λ.

Evaluating effectiveness. We evaluate the effectiveness of each assignment by considering the quality of healthcare service for pathologies of interest. For each emergency room $s \in S$, vector q_s gives the quality for each medical treatment supplied by s.

In particular, the quality of care service supplied by s for treating p is denoted by q_{sp} and it is computed according to two indicators, namely the ratio of medical treatments for p over the total number of medical services supplied by s and the ratio of successful clinical interventions for p. In our model, we relate q_s components to time dimension by introducing a suitable parameter γ, which expresses the amount of time a patient is prepared to wait for achieving a one-percentage point improved service.

We are now ready to define the basic MIP formulation. Let n_t^s be a nonnegative integer variable representing the total number of first-aid requests assigned to emergency room s during time slot t. Let $\bar{n}_s$ be a nonnegative integer constant corresponding to the expected total number of patients who have been waiting or receiving medical treatments in s at starting time of first time slot of T. Let $\alpha_t \in [0,1]$ a real-valued constant that reports the expected ratio of patients who have required first-aid services during $t-1$, but are still waiting or receiving medical treatments during t. Let z_t^s the workload of s during t, estimated by waiting time function w. Let y_t^s be a binary variable such that $y_t^s = 1$ if the total workload of s during t exceeds the fixed threshold $\bar{z}$. Thus, a MIP formulation of the problem follows:

(**MIP**)

$$\min \sum_{r\in R}\sum_{s\in S} d(u_r, v_s)x_{rs} + \sum_{t\in T}\sum_{s\in S}(z_t^s + \lambda y_t^s) - \gamma \sum_{s\in S}\sum_{p\in P} q_{sp} \sum_{\substack{r\in R:\\ p_r=p}} x_{rs} \tag{6.9}$$

s.t.

$$n_0^s = \bar{n}_s \qquad \forall s \in S \tag{6.10}$$

$$n_t^s = \sum_{\substack{r\in R:\\ t_r=t}} x_{rs} \qquad \forall s \in S,\ \forall t \in T \tag{6.11}$$

$$z_t^s \geq a_h^s\left(\alpha_t n_{t-1}^s + n_t^s\right) + b_h^s \qquad \forall h \in \{1,\ldots,k_s\},\ \forall s \in S,\ \forall t \in T \tag{6.12}$$

$$z_t^s \leq \bar{z} + M y_t^s \qquad \forall s \in S,\ \forall t \in T \tag{6.13}$$

$$x_{rs} \in \mathcal{A} \cap \{0,1\} \qquad \forall r \in R,\ \forall s \in S \tag{6.14}$$

$$n_t^s \in \mathbb{Z} \qquad \forall s \in S,\ \forall t \in T \cup \{0\} \tag{6.15}$$

$$z_t^s \in \mathbb{R} \qquad \forall s \in S,\ \forall t \in T \tag{6.16}$$

$$y_t^s \in \{0,1\} \qquad \forall s \in S,\ \forall t \in T \tag{6.17}$$

where M is a suitable large real-valued constant and $\mathcal{A} \subset \mathbb{R}^{|R|\cdot|S|}$ is the polytope given by assignment constraints (6.1)-(6.3).

Let us observe that any constraint (6.12) forces the corresponding z_t^s variable to assume the appropriate value of function w by exploiting property (6.8). In particular, z_t^s estimates the total waiting time of s during t by considering all requests assigned at time slot t and the partial number of requests assigned at time slot $t-1$, obtained from ratio α_t.

Moreover, it is easy to check that any constraint (6.13) forces the associated y_t^s to 1 if the total waiting time z_t^s exceeds $\bar{z}$. Let us observe that y_t^s can get value 1 also when the previous condition is not satisfied: in that case, the corresponding solution could be feasible but not optimal because the objective function (6.9) is in minimization form.

6.3.2 Integrality Property of the Assignment Variables

In the following we show how to simplify model (6.9)–(6.17) by exploiting some integrality property of the assignment part of the model. First of all, we need a preliminary result that characterizes the polytope associated with assignment variables and constraints.

Proposition 6.2. *Given $|S|\cdot|T|$ integers ν_t^s, let polytope $\mathcal{P} \subset \mathbb{R}^{|R|\cdot|S|}$ be the intersection between $\mathcal{A}$ and the polyhedron in $\mathbb{R}^{|R|\cdot|S|}$ defined by the inequalities*

$$\sum_{\substack{r \in R: \\ t_r = t,\ f_{p_r} \in F_s}} x_{rs} = \nu_t^s \quad \forall s \in S,\ \forall t \in T. \tag{6.18}$$

Then, $\mathcal{P}$ is integral and the problem of optimizing a linear function over $\mathcal{P}$ is strongly polynomial.

Proof. Let $H(N, A)$ be a digraph with node set $N := R \cup (S \times T)$ and arc set A such that: (i) each node is in bijection with either a request $r \in R$ or a pair $(s,t) \in S \times T$; (ii) each arc is in bijection with an ordered pair $(r, (s,t))$ that satisfies both conditions $d(u_r, v_s) \leq t^{\max}(p_r)$ and $f_{p_r} \in F_s$. Moreover, let us consider the following formulation of $\mathcal{P}$:

$$\sum_{\substack{s \in S: \\ f_{p_r} \in F_s, \\ d(u_r, v_s) \leq t^{\max}(p_r)}} x_{rs} = 1 \qquad \forall r \in R \tag{6.19}$$

$$\sum_{\substack{r \in R: \\ t_r = t,\ f_{p_r} \in F_s, \\ d(u_r, v_s) \leq t^{\max}(p_r)}} x_{rs} = \nu_t^s \qquad \forall s \in S,\ \forall t \in T \tag{6.20}$$

$$0 \leq x_{rs} \leq 1 \qquad \forall t \in R,\ s \in S \tag{6.21}$$

where (6.19) and (6.20) are obtained by combining (6.2) respectively with (6.1) and (6.18). Now, it easy to check that the constraints matrix associated with (6.19)-(6.21), called B, corresponds to the incidence matrix of H, thus B is totally unimodular, so it follows that $\mathcal{P}$ is integral. In particular, $\mathcal{P}$ corresponds to the feasible region of a flow problem associated with digraph H with demands $d_r = -1$ for each $r \in R$, $d_{(s,t)} = \nu_t^s$ for each $(s,t) \in S \times T$. Then, by **citation**, we can conclude that optimizing a linear function over $\mathcal{P}$ is strongly polynomial. $\square$

Now, we are able to define an improved formulation in which the number of integer variables is reduced from $O(|R| \cdot |S| + |S| \cdot |T|)$ to $O(|S| \cdot |T|)$.

Theorem 6.3. *Let* **MIP′** *be the mixed integer program obtained from* **MIP** *by relaxing the integrality of variables* x_{rs}*. Then,* **MIP** *and* **MIP′** *have the same optimum value and an optimal solution to* **MIP** *can be obtained from an optimal solution to* **MIP′** *in strongly polynomial time.*

Proof. Let ω' and ω be resp. the optimal solution values of **MIP′** and **MIP**. In general, $\omega' \leq \omega$ holds since **MIP′** is a relaxation of **MIP**. Let $\chi' := (x', n', z', y')$ be an optimal solution of **MIP′** and let consider polytope $\mathcal{P}$ with $\nu_t^s = n'^s_t$ for all $s \in S$, $t \in T$. Then, let $x^\star$ be an optimal solution obtained by maximizing function $\sum_{r \in R} \sum_{s \in S} d(u_r, v_s) x_{rs} - \gamma \sum_{s \in S} \sum_{p \in P} q_{sp} \sum_{\substack{r \in R:\\ p_r = p}} x_{rs}$ over $\mathcal{P}$. Due to Proposition 6.2, $x^\star$ is integral and it can be computed in strongly polynomial time. Since $\chi^\star := (x^\star, n', z', y')$ is feasible for **MIP′** and its corresponding objective function value is less or equal to ω', we have that $\chi^\star$ is an optimal solution to **MIP′**. Moreover, since $\chi^\star$ is feasible for **MIP**, we can conclude that $\chi^\star$ is optimal also to **MIP**. $\square$

6.3.3 Relaxing Workload Balance

Let $\mathbf{MIP_0}$ be the mixed integer program obtained from **MIP′** by relaxing constraints (6.13) (and assuming $\lambda = 0$). In particular, $\mathbf{MIP_0}$ models the relaxation of MIP (6.9)–(6.17) in which emergency rooms workloads are not required to be balanced. In the following, we present a reformulation of $\mathbf{MIP_0}$ as a generalized min cost flow problem on a suitable network.

Let $D(N, A)$ be a digraph with node set N and arc set A, let $b : N \to \mathbb{R}$ be a demand function associated with nodes, let $l, \mu : A \to \mathbb{R}_+$ and $a : A \to \mathbb{R}$ be respectively capacity, gain and cost functions associated with arcs. A pseudoflow is a function $\varphi : A \to \mathbb{R}$ such that $0 \leq \varphi(i,j) \leq l(i,j)$ holds for all arcs $(i,j) \in A$. The generalized min cost flow problem consists of finding a pseudoflow that minimizes the overall cost $\sum_{(i,j) \in A} a(i,j)\varphi(i,j)$ subject to the generalized flow conservation constraints

$$\sum_{(i,j) \in A} \varphi(i,j) - \sum_{(j,i) \in A} \mu(j,i)\varphi(j,i) = b(i) \qquad \forall i \in N.$$

For each $e = (i, j) \in A$, let $\bar{e} := (j, i)$ be the reverse arc corresponding to e and let $\bar{A}$ denote the set of reverse arcs associated with A. For reverse arcs, gain and cost functions respectively satisfy $\gamma(\bar{e}) = 1/\gamma(e)$ and $a(\bar{e}) = -a(e)/\gamma(e)$. Moreover, given a pseudoflow φ, the residual capacity function $l_\varphi : A \cup \bar{A} \to \mathbb{R}$, is defined as $l_\varphi(e) = l(e) - \varphi(e)$ for each $e \in A$ and $l_\varphi(\bar{e}) = \gamma(e)\varphi(e)$. Then, let $D_\varphi(N, \bar{A}, b, l_\varphi, \gamma, a)$ be the residual network associated with φ. The gain of a cycle belonging to D_φ is the product of the gains of arcs that compose the cycle. A cycle of D_φ whose gain is strictly greater (resp. less) than one unit is called flow-generating (resp. flow-absorbing). A bicycle is composed by a flow absorbing cycle and a flow generating cycle that are arc-disjoint and connected by a path containing at least one node. We recall that a feasible pseudoflow φ is optimal if and only if D_φ does not contain any unit-gain cycle or bicycle. For further details, the reader is referred to Ahuja et al. [1993], Goldberg et al. [1989b].

The generalized min cost flow is a well-known optimization problem that has a wide range of applications in many scientific area, as discussed in Ahuja et al. [1993]. It belongs to the field of generalized flow, so it reduces to min cost flow by assuming $\gamma(e) = 1$ for all $e \in A$. Since generalized min cost flow is a special case of linear programming, it can be solved in polynomial time by ellipsoid method Karmarkar [1984]. In literature, many other polynomial algorithms have been addressed, which are based on linear programming as reported in Kapoor and Vaidya [1986], Vaidya [1989], or explain combinatorial approaches, like in Goldberg et al. [1989a], Wayne [2002]. While min cost flow can be solved in strongly polynomial time Tardos [1985], it is unknown whether the generalized min cost flow problem admits strongly polynomial algorithms. However, in Cohen and Megiddo [1994] it is shown that the problem is strongly polynomial if there is a fixed number of arcs whose gain is either than one unit.

In the following we characterize an instance of generalized min cost flow, denoted by $D(N, A, b, l, \gamma, a)$, which gives a combinatorial description of $\mathbf{MIP_0}$.

Let $K_s := \{1, \ldots, k_s\} \times T$ for each $s \in S$, $R' := R$, $S' := S$ and $T' := T$. Then, let $D(N, A)$ be a digraph with node set

$$N = R \cup S \cup (S \times T) \cup K_1 \cup \cdots \cup K_{|S|} \cup (S' \times T') \cup R' \cup S',$$

and arc set $A = \bigcup_{j=1}^{8} A_j$ such that

$$
\begin{aligned}
A_1 &:= \{(r,(s,t)) : r \in R,\ s \in S,\ t \in T \text{ with } r_t = t,\ d(u_r, v_s) \leq t^{\max}(p_r),\ f_{p_r} \in F_s\} \\
A_2 &:= \{(s,(s,t)) : s \in S,\ t = 1\} \\
A_3 &:= \{((s,t),(k,t)) : s \in S,\ t \in T,\ (k,t) \in K_s\} \\
A_4 &:= \{((k,t),(s',t')) : s' \in S',\ t' \in T',\ (k,t) \in K_s \text{ with } s = s',\ t = t'\} \\
A_5 &:= \{((s',t'),(s,t+1)) : s' \in S',\ t' \in T' \text{ with } t' < |T'|,\ s = s',\ t = t'\} \\
A_6 &:= \{((s',t'+1),r') : s' \in S',\ t' \in T',\ r' \in R' \text{ with } t' < |T'|,\ r'_t = t'\} \\
A_7 &:= \{((s',t'),r') : s' \in S',\ t' = |T'|,\ r' \in R' \text{ with } r'_t = |T'|\} \\
A_8 &:= \{((s',t'),s') : s' \in S',\ t' = 1\}.
\end{aligned}
$$

Furthermore, let $R_t := \{r \in R : r_t = t\}$; then, we are given

$$
b(i) = \begin{cases}
-1 & i = r \in R \\
+1 & i = r' \in R' \\
-n_0^s & i = s \in S \\
+n_0^{s'} & i = s' \in S' \\
0 & i = (k,t) \in K_s \quad \forall s \in S \\
0 & i = (k',t') \in K_{s'} \quad \forall s' \in S';
\end{cases}
$$

$$
l(e) = \begin{cases}
1 & e \in A_1 \\
n_0^s & e \in A_2 \\
n_k - n_{k-1} & e \in A_3 \cup A_4 \text{ with } k < k_s,\ n_0 = 0 \\
|R_t| - n_k & e \in A_3 \cup A_4 \text{ with } k = k_s,\ R_t := \{r \in R : r_t = t\} \\
|R_t| & e \in A_5 \\
\alpha_t |R_t| & e \in A_6 \\
\alpha_t & e \in A_7 \\
\alpha_0 n_0^s & e \in A_8
\end{cases}
$$

$$
\mu(e) = \begin{cases}
\alpha_t & e \in A_5 \\
1/\alpha_t & e \in A_6 \\
1 & e \in A \setminus (A_5 \cup A_6)
\end{cases}
$$

$$
a(e) = \begin{cases}
d(u_r, v_s) - \gamma q_{sp} & e \in A_1 \\
a_k^s & e \in A_3 \\
0 & e \in A \setminus (A_1 \cup A_3)
\end{cases}
$$

The following result states that the generalized min cost problem associated with $D(N, A, b, l, \mu, a)$ is a relaxation of $\mathbf{MIP_0}$.

Lemma 6.4. *For each feasible solution χ to* $\mathbf{MIP_0}$ *there exists a feasible pseudoflow φ_χ associated with $D(N, A, b, l, \mu, a)$ such that χ and φ_χ have the same cost.*

Proof. Let $\chi := (x, n, z)$ be a feasible solution of $\mathbf{MIP_0}$. For each $s \in S$, let h_s the largest integer in $\{1, \ldots, k_s\}$ such that $h_s \leq n_t^s + \alpha_{t-1} n_{t-1}^s$. Then, let φ_χ a pseudoflow associated with $D(N, A, b, l, \mu, a)$ such that, for each $r' = r \in R$, $s' = s \in S$, $t' = t \in T$

$$\varphi_\chi(r, (s, t)) = x_{rs} \quad \text{with} \quad t_r = t \tag{6.22}$$
$$\varphi_\chi((s', t' + 1), r') = \alpha_t x_{rs} \quad \text{with} \quad t_r = t,\, t < |T| \tag{6.23}$$
$$\varphi_\chi((s', |T|), r') = x_{rs} \quad \text{with} \quad t_r = |T| \tag{6.24}$$
$$\varphi_\chi(s, (s, t)) = n_0^s \quad \text{with} \quad t = 1 \tag{6.25}$$
$$\varphi_\chi((s', t'), s') = \alpha_t n_0^s \quad \text{with} \quad t = 1 \tag{6.26}$$
$$\varphi_\chi((s', t'), (s, t + 1)) = n_t^s \quad \text{with} \quad t < |T| \tag{6.27}$$
$$\varphi_\chi((s, t), (k, t)) = n_k - n_{k-1} \quad \forall k \leq h_s \tag{6.28}$$
$$\varphi_\chi((s, t), (h_s + 1, t)) = n_t^s + \alpha_{t-1} n_{t-1}^s - n_{h_s}. \tag{6.29}$$

It is easy to check that φ_χ is feasible: capacity and pseudoflow conservation constraints are satisfied. Moreover, let observe that (6.22) implies

$$\sum_{r \in R} \sum_{s \in S} \left(d(u_r, v_s) - \gamma \sum_{\substack{p \in P: \\ p_r = p}} q_{sp} \right) x_{rs} = \sum_{r \in R} \sum_{s \in S} a(r, (s, t)) \varphi_\chi(r, (s, t)). \tag{6.30}$$

By relation (6.8), the following condition holds:

$$z_t^s = a_{h_s+1}^s \left(\alpha_t n_{t-1}^s + n_t^s \right) + b_{h_s+1}^s \quad \forall s \in S, \forall t \in T. \tag{6.31}$$

Then, by substituting (6.5)-(6.7) in (6.31), it follows that

$$\begin{aligned} z_t^s = a_{h_s+1}^s (n_t^s + \alpha_t n_{t-1}^s) + \sum_{h=1}^{h_s} (a_h^s - a_{h+1}^s) n_h = \\ a_{h_s+1}^s (n_t^s + \alpha_{t-1} n_{t-1}^s - n_{h_s}) + \sum_{h=1}^{h_s} (n_h - n_{h-1}) a_h^s = \\ a((s, t), (h_s + 1, t)) \varphi_\chi((s, t), (h_s + 1, t)) \\ + \sum_{h=1}^{h_s} a((s, t), (h, t)) \varphi_\chi((s, t), (h, t)). \end{aligned} \tag{6.32}$$

Thus, relations (6.30) and (6.32) imply that χ and φ_χ have the same cost. $\square$

In general, the converse is not true, i.e., there exist feasible solutions of generalized min cost flow over $D(N, A, b, l, \mu, a)$ that cannot be mapped into feasible solutions of $\mathbf{MIP_0}$.

However, latter problems are equivalent under certain conditions, e.g., assuming $\alpha_t = 1$ for each $t \in T \cup \{0\}$. In this case, the generalized min cost flow over $D(N, A, b, l, \mu, a)$ reduces to min cost flow problem over $D(N, A, b, l, a)$. Since demands and capacities are integer, there exist integral optimal flows corresponding to feasible solutions of $\mathbf{MIP_0}$, which are optimal by Lemma 6.4. Furthermore, we can show the following result.

Theorem 6.5. *Let us assume*

$$\min_{\substack{r\in R \\ s,\tilde{s}\in S,\, s\neq\tilde{s}}} \{d(u_r, v_s) - d(u_r, v_{\tilde{s}})\} \geq \max_{\substack{s,\tilde{s}\in S \\ s\neq\tilde{s}}} \left\{a^s_{k_s} - a^{\tilde{s}}_1\right\}. \tag{6.33}$$

Then, optimal solution to $\mathbf{MIP_0}$ *can be computed in strongly polynomial time.*

Proof. Let $\varphi^\star$ be an optimal pseudoflow to generalized min cost flow problem associated with $D(N, A, b, l, \mu, a)$. In general, $\varphi^\star$ is not integral. Since the residual network $D_{\varphi^\star}$ corresponding to $\varphi^\star$ does not contain negative cycles, it easy to check that vertices (s, t), (k, t), (s', t') form strictly positive cost cycles for each $s \in S$, $t \in T$ with $s' = s$ and $t' = t$. Thus, it follows that relation (6.32) is satisfied. Moreover, (6.33) ensures that for each $(r, (s, t)) \in A$ such that $0 < \varphi^\star(r, (s, t)) < 1$, there exists at least a null cost cycle in $D_{\varphi^\star}$ that contains arc $(r, (s, t))$ and $((s, t), (k, t))$ with residual capacity greater or equal to $1 - \varphi^\star(r, (s, t))$. Thus, an optimal integer pseudoflow φ' can be obtained from $\varphi^\star$ by saturating $O(|R|)$ null cost cycles. Then, we have that φ' corresponds to a feasible solution χ' of $\mathbf{MIP_0}$, so we conclude that χ' is optimal by Lemma 6.4. Finally, by Cohen and Megiddo [1994], it follows that χ' can be computed in strongly polynomial time. □

6.4 Computational Results

The computational experience focuses on a wide set of instances that are based on real-world data from the Lazio emergency room department system during entire year 2012.

First-aid requests characteristics have been retrieved from the Hospitals Information System (HIS)[2] and the Emergency Room Information System (ERIS)[3] of regional healthcare services authority, by the support of DEP-Lazio. In particular, HIS manages the Hospital Discharge Register (HDR)[4] database, which maintains information of all hospital admissions and discharges, by integrating patients personal details, healthcare services supplied and medical treatment results. Lazio's HDR provides additional medical treatments information for STEMI, AMI and FF, which are the pathologies of

[2]Corresponding Italian acronym is SIO: "Sistema Informativo Ospedaliero".
[3]Corresponding Italian acronym is SIES: "Sistema Informativo per l'Emergenza Sanitaria".
[4]Corresponding Italian acronym is SDO: "Scheda di dimissione ospedaliera".

interest associated with our analysis. The ERIS integrates HIS database by supplying specific and detailed information only for emergency room departments.

The description of regional emergency departments, with associated quality of service information, has been retrieved from statistical studies carried out by DEP-Lazio, which are based on regional and national evaluation programs for medical operations results. For more details, we refer the reader to Fusco et al. [2012], Renzi et al. [2012].

The computational experience has been carried out on a x86-64 GNU/Linux machine (CentOS 6.3) with 8 cores @2GHz and 16GB of RAM. We have generated instances of $\mathbf{MIP'}$ and $\mathbf{MIP_0}$ for each day of year 2012 by considering all 50 operating emergency departments of Lazio. Then, we have achieved optimal solutions for all instances by using IBM ILOG Cplex 12.5.1.

Table 6.1 summarizes computational results for $\mathbf{MIP'}$ instances by reporting average values for each month: in particular, *i*) the second column reports the average number of emergency requests occurred in each day; *ii*) the third column reports the number of infeasible instances, i.e., the number of days of the month in which at least one request could not be correctly assigned according to the constraints of our model; *iii*) the forth column shows the average optimal solution value of each day, while columns fifth and sixth indicate cost contributions of waiting time functions (the sum of the z_t^s variables in (6.9)) and overall penalty time value (the sum of λy_t^s terms), respectively; *iv*) the last two columns report the average Cplex performance (in terms of elapsed real computational time and total branch&bound nodes) that has been observed for solving each (feasible) instance of the month.

Month (2012)	**Mean |R| (day)**	**Infeas. MIPs**	**Optimum (min)**	**Waiting time (min)**	**Penalty time (min)**	**Cplex time (sec)**	**Cplex nodes (average)**
1	4,300.4	8	60,297.63	25,360.77	2,899.85	111.316	3,209.1
2	3,980.9	6	55,943.78	23,570.66	2,838.61	109.878	3,387.1
3	4,323.1	5	61,418.85	25,859.72	2,888.61	122.203	3,231.3
4	4,287.2	5	60,861.64	25,599.72	2,894.28	131.929	3,384.2
5	4,493.5	7	63,015.01	26,568.41	2,924.67	137.239	3,409.3
6	4,590.0	7	64,929.66	27,427.35	2,990.86	148.866	3,371.1
7	4,409.1	8	62,458.13	26,341.63	2,977.74	139.629	3,311.0
8	4,229.0	4	59,870.97	25,180.69	2,932.98	132.760	3,257.4
9	3,996.3	0	56,082.89	23,604.62	2,859.93	113.067	3,305.0
10	4,154.2	10	58,111.27	24,425.29	2,874.46	112.856	3,395.1
11	4,112.7	11	57,930.31	24,405.94	2,902.58	114.509	3,382.1
12	4,025.8	8	56,399.67	23,862.90	2,822.65	119.214	3,295.8

TABLE 6.1: Computational experience.

The results in Table 6.1 show that $\mathbf{MIP'}$ can be solved relatively easily by a sophisticated MIP solver like Cplex 12.5.1. The number of instances that turn out to be infeasible is relatively small, namely around 20%. The influence of the penalty term associated with workload unbalance amounts at 10% of the term associated with the waiting time. In order to evaluate how important is the penalization of such an unbalance we also solved $\mathbf{MIP_0}$ and the results are rather easy to interpret. Because $\mathbf{MIP_0}$ is a relaxation of $\mathbf{MIP'}$, as discussed in Section 6.3.1, optimal values to $\mathbf{MIP_0}$ are on average better of 5.36% than those of $\mathbf{MIP'}$, but at the price of an increased waiting time cost contribution, on average of 0.74%, due to the absence of workload balance. We omit detailed results on $\mathbf{MIP_0}$ instances but it is worth mentioning that they are very easy to solve both by using a combinatorial algorithm for generalized min cost flow or by solving $\mathbf{MIP_0}$ with a general-purpose MIP solver like Cplex. In the latter case, no branching is ever necessary.

As pointed out in the Introduction, the aim of the present study is to have a completely offline picture of the effect of an optimal assignment of requests to Emergency Rooms in Lazio and to use it to evaluate both the state of the art and future reorganization ideas. To achieve this we compare in Table 6.2 real (observed) first-aid request assignments during year 2012 with the optimal solutions of model $\mathbf{MIP'}$. Specifically, Table 6.2 is organized as follows: *i*) second and third columns indicate the number of infeasible assignments with respect to the violation of constraints (6.2) (each patient has to reach an emergency room department within a suitable time according to kind of health emergency) and (6.1) (each request has to be assigned to an emergency department with a suitable specialization that allows to supply appropriate medical treatments), respectively; *ii*) the forth column exhibits the average objective function value for each day, while columns fifth and sixth specify cost contributions of waiting time functions and overall penalty time value, respectively (analogously to Table 6.1); *iii*) the last three columns report the average relative gaps between values of observed assignments and optimal solution for the overall value and cost contributions of waiting and penalty times, respectively.

Month (2012)	Violated constr. (6.2)	Violated constr. (6.1)	Overall value (min)	Waiting time (min)	Penalty time (min)	Overall value (gap%)	Waiting time (gap%)	Penalty time (gap%)
1	2.0	360.4	101,437.07	47,964.39	3,851.36	40.18	46.69	24.58
2	2.0	299.2	91,175.49	42,848.48	3,818.35	38.86	45.01	25.79
3	1.6	362.6	102,101.16	48,415.55	3,870.28	40.05	46.73	25.38
4	2.0	361.1	101,898.65	48,231.42	3,862.21	40.57	47.24	25.13
5	1.5	370.4	105,827.41	50,405.02	3,886.48	40.37	47.19	24.80
6	2.0	383.1	110,125.15	51,671.08	3,923.45	41.27	47.19	23.76
7	1.8	343.5	108,814.64	48,679.28	3,944.40	42.90	46.07	24.47
8	2.5	361.5	109,621.28	46,288.93	3,949.02	45.15	45.39	25.72
9	1.9	301.8	93,799.83	43,309.37	3,874.65	40.19	45.43	26.19
10	2.0	320.5	96,755.77	45,338.37	3,852.34	39.32	45.33	25.32
11	1.7	335.2	95,445.33	45,136.85	3,838.66	40.06	46.79	24.52
12	1.8	357.5	94,010.21	44,578.50	3,818.23	40.00	46.22	26.16

Table 6.2: Comparing observed request allocations with optimized solutions.

The numbers in Table 6.2 immediately show that the solutions naturally obtained without a centralized allocation strategy (for example a remote triage) violate many of the constraints of $\mathbf{MIP'}$, especially constraints (6.1) associated with suitable specialization. This information is especially interesting from a strategic (and practical) standpoint: such a remote triage conducted in an effective way could have a remarkable impact to significantly reduce these violations that correspond to dangerous inefficiency of the system. The rest of the numbers of Table 6.2 are instead interesting but not easy to interpret. In a sense the objective function (6.9) is completely disregarded by the observed request allocation system but maybe the part of it associated with the minimization of the travel time. Thus, the absolute and relative difference of the components of the objective function are less meaningful at this point in time, while they will be more and more so when different reorganization settings will be evaluated.

6.5 Conclusions

The assignment of service requests to Emergency Rooms is of paramount importance both from a life-threatening and an economical viewpoints. In the process of a more general project that aims at defining optimal allocation policies of patients to regional hospital network facilities, the Department of Epidemiology of the Regional Health Service of Lazio, Italy was interested in obtaining a completely offline picture of the effect of an optimal assignment of requests to Emergency Rooms so as to be able to evaluate both the state of the art and future reorganization ideas.

We have implemented and tested with real-world data of all service requests of 2012 a Mixed Integer Linear Programming model that computes such an optimal request allocation by minimizing travel and waiting times and penalize workload unbalance among emergency rooms in the region. Within the development process we have studied special cases and relaxations of the complete model showing interesting mathematical properties that are, in turn, useful from a practical viewpoint.

The present study is a first step in the evaluation process of centralized allocation strategies like remote triage that could have a remarkable impact in making the allocation process much more efficient and effective.

Bibliography

Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ, 1993.

Adrian Bagust, Michael Place, and John W Posnett. Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. *BMJ*, 319(7203): 155–158, 7 1999. doi: 10.1136/bmj.319.7203.155.

C C Branas, E J MacKenzie, and C S ReVelle. A trauma resource allocation model for ambulances and hospitals. *Health Serv Res*, 35(2):489–507, Jun 2000. ISSN 0017-9124 (Print); 0017-9124 (Linking).

S.C.K. Chu and L. Chu. A modeling framework for hospital location and service allocation. *International Transactions in Operational Research*, 7(6):539–568, 2000. ISSN 1475-3995. doi: 10.1111/j.1475-3995.2000.tb00216.x. URL http://dx.doi.org/10.1111/j.1475-3995.2000.tb00216.x.

Jeffery K. Cochran and Kevin T. Roche. A multi-class queuing network analysis methodology for improving hospital emergency department performance. *Computers & Operations Research*, 36(5):1497 – 1512, 2009. ISSN 0305-0548. doi: http://dx.doi.org/10.1016/j.cor.2008.02.004. URL http://www.sciencedirect.com/science/article/pii/S0305054808000257.

Edith Cohen and Nimrod Megiddo. Algorithms and complexity analysis for some flow problems. *Algorithmica*, 11(3):320–340, 1994.

Lloyd G Connelly and Aaron E Bair. Discrete event simulation of emergency department activity: a platform for system-level operations research. *Acad Emerg Med*, 11(11):1177–1185, Nov 2004. ISSN 1069-6563 (Print); 1069-6563 (Linking). doi: 10.1197/j.aem.2004.08.021.

Danilo Fusco, Anna P Barone, Chiara Sorge, Mariangela D'Ovidio, Massimo Stafoggia, Adele Lallo, Marina Davoli, and Carlo A Perucci. P.re.val.e.: outcome research program for the evaluation of health care quality in lazio, italy. *BMC Health Serv Res*, 12:25, 2012. ISSN 1472-6963 (Electronic); 1472-6963 (Linking). doi: 10.1186/1472-6963-12-25.

Andrew V. Goldberg, Serge A. Plotkin, and Éva Tardos. Combinatorial algorithms for the generalized circulation problem. *Mathematics of Operations Research*, 16: 351–379, 1989a.

A.V. Goldberg, É. Tardos, Stanford University. Computer Science Department, and R.E. Tarjan. *Network Flow Algorithms*. Number No. 1252 in Computer Science Department: Report STAN-CS. Department of Computer Science, Stanford University, 1989b. URL http://books.google.it/books?id=mDsZnQEACAAJ.

M.M. Gunal and M. Pidd. Understanding accident and emergency department performance using simulation. In *Simulation Conference, 2006. WSC 06. Proceedings of the Winter*, pages 446–452, Dec 2006. doi: 10.1109/WSC.2006.323114.

P.R. Harper, A.K. Shahani, J.E. Gallagher, and C. Bowie. Planning health services with explicit geographical considerations: a stochastic location,Äìallocation approach. *Omega*, 33(2):141 – 152, 2005. ISSN 0305-0483. doi: http://dx.doi.org/10.1016/j.omega.2004.03.011. URL http://www.sciencedirect.com/science/article/pii/S0305048304000593.

S Kapoor and P M Vaidya. Fast algorithms for convex quadratic programming and multicommodity flows. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, STOC '86, pages 147–159, New York, NY, USA, 1986. ACM. ISBN 0-89791-193-8. URL http://doi.acm.org/10.1145/12130.12145.

N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

Alexander Komashie and Ali Mousavi. Modeling emergency departments using discrete event simulation techniques. In *Proceedings of the 37th Conference on Winter Simulation*, WSC '05, pages 2681–2685. Winter Simulation Conference, 2005. ISBN 0-7803-9519-0. URL http://dl.acm.org/citation.cfm?id=1162708.1163203.

Renata Konrad, Kristine DeSotto, Allison Grocela, Patrick McAuley, Justin Wang, Jill Lyons, and Michael Bruin. Modeling the impact of changing patient flow processes in an emergency department: Insights from a computer simulation study. *Operations Research for Health Care*, 2(4):66 – 74, 2013. ISSN 2211-6923. doi: http://dx.doi.org/10.1016/j.orhc.2013.04.001. URL http://www.sciencedirect.com/science/article/pii/S2211692313000052.

I Kuban Altınel and Esra Ulaş. Simulation modeling for emergency bed requirement planning. *Annals of Operations Research*, 67(1):183–210, 1996. ISSN 0254-5330. doi: 10.1007/BF02187029. URL http://dx.doi.org/10.1007/BF02187029.

Gianmaria Leo, Andrea Lodi, and Paolo Tubertini. Emergency room management in lazio, italy. Technical Report OR 14-9, LabOR DEI - Università di Bologna, 2014.

Cristina Renzi, Chiara Sorge, Danilo Fusco, Nera Agabiti, Marina Davoli, and Carlo A Perucci. Reporting of quality indicators and improvement in hospital performance: the p.re.val.e. regional outcome evaluation program. *Health Serv Res*, 47(5):1880–1901, Oct 2012. ISSN 1475-6773 (Electronic); 0017-9124 (Linking). doi: 10.1111/j.1475-6773.2012.01401.x.

David Sinreich and Ola Jabali. Staggered work shifts: a way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Management Science*, 10(3):293–308, 2007. ISSN 1386-9620. doi: 10.1007/s10729-007-9021-z. URL http://dx.doi.org/10.1007/s10729-007-9021-z.

David Sinreich and Yariv Marmor. Ways to reduce patient turnaround time and improve service quality in emergency departments. *Journal of Health Organization and Management*, 19(2):88–105, 2005. doi: 10.1108/14777260510600022. URL http://dx.doi.org/10.1108/14777260510600022.

David Sinreich and Yariv N. Marmor. A simple and intuitive simulation tool for analyzing emergency department operations. In *Proceedings of the 36th Conference on Winter Simulation*, WSC '04, pages 1994–2002. Winter Simulation Conference, 2004. ISBN 0-7803-8786-4. URL http://dl.acm.org/citation.cfm?id=1161734.1162107.

Éva Tardos. A strongly polynomial minimum cost circulation algorithm. *Combinatorica*, 5(3):247–255, July 1985. ISSN 0209-9683. doi: 10.1007/BF02579369. URL http://dx.doi.org/10.1007/BF02579369.

P.M. Vaidya. Speeding up linear programming using fast matrix multiplication. In *Proceedings of 30th IEEE Annual Symposium on Foundations of Computer Science*, pages 332–337, 1989.

Junwen Wang, Jingshan Li, and PatriciaK. Howard. A system model of work flow in the patient room of hospital emergency department. *Health Care Management Science*, 16(4):341–351, 2013. ISSN 1386-9620. doi: 10.1007/s10729-013-9235-1. URL http://dx.doi.org/10.1007/s10729-013-9235-1.

Kevin D. Wayne. A polynomial combinatorial algorithm for generalized minimum cost flow. *Mathematics of Operations Research*, 27(3):445–459, 2002. doi: 10.1287/moor.27.3.445.313. URL http://pubsonline.informs.org/doi/abs/10.1287/moor.27.3.445.313.

Jinn-Yi Yeh and Wen-Shan Lin. Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department. *Expert Syst. Appl.*, 32(4):1073–1083, May 2007. ISSN 0957-4174. doi: 10.1016/j.eswa.2006.02.017. URL http://dx.doi.org/10.1016/j.eswa.2006.02.017.