UNIVERSITAT DE BARCELONA

Data Driven Approach to Enhancing Efficiency and Value in Healthcare

Richard E. Guerrero Ludueña

Chapter 1

Introduction

This Thesis investigates the impact of Machine Learning, Metaheuristics, Modelling & Simulation, and Data Analytics in increasing efficiency and value in healthcare, analysing strategies to expand the impact of those data-driven approaches. The research is based on solving practical problems for particular context in different levels of healthcare management, with the research drawing upon both theoretical and empirical analysis.

This chapter helps to contextualise the subsequent research, and is organised as follows: Section 1.1 reviews trends in healthcare in Europe. Section 1.2 describes the data-driven healthcare. Section 1.3 reviews supply and demand of data-driven approaches. An outline of the research aims is presented in Section 1.4. Section 1.5 discusses limitations of study. Section 1.6 describes the methods. And an outline of the Thesis is described in Section 1.7.

1.1 Healthcare trends

Despite the wide range of different National Health Systems across Europe, there are shared trends both desirable and detrimental. Among the main issues that countries need to address are: inequality in life expectancy; increase in cancer deaths (since 1985, EU cancer deaths have increased by 12% for men and by 9% for women); increase in diabetes; increase in Alzheimer's cases; growth in incidence and prevalence of chronic disease (related to 70%-80% of healthcare costs). Desirable trends in

European countries are: increased life expectancy; reduced infant mortality; better treatment of life-threatening conditions (such as heart attacks, strokes and cancer); increase in the survival rates for different types of cancer (including colorectal and breast, thanks to earlier detection and better treatment); increase in the number of doctors per capita [58].

1.2 Data-driven healthcare

Healthcare is changing, and a new era of data-driven healthcare organisations is under way. Considerable excitement has been generated in recent years by the progress in policy directives on open data [151, 203]; digitalisation of medical records [6, 117]; technical advances and analytical tools [82, 221, 146, 226]; pharmaceutical R&D data availability [120]; and mobile healthcare applications [227, 216, 144]. All of these initiatives, focus on the needs of patients, healthcare providers, and payers, could help to improve health and reduce costs through increasing efficiency (measure whether healthcare resources are being used to get the best value for money) [161, 64] and value (balance of healthcare spend (cost) and patient impact (outcomes)) [169, 168].

Recently there has been increased interest from the private sector, government, national health services, and other healthcare stakeholders in concepts such as *big data*, *data science*, *predictive modelling* and *simulation techniques*. Researchers from different academic disciplines (e.g., computer science, physics, mathematics), on the other hand, have generated a considerable body of research over the recent decades in these well-known methodologies. Nevertheless, despite some exceptions, the gap in knowledge transfer is still apparent (see for example [135, 137, 149]).

1.3 Data-driven: supply and demand

Several studies have investigated the traditional separation of the producers of research evidence in academia from the users of that evidence [98, 96, 97, 142, 38]. Ward et al [218] identified 28 models to explain the knowledge transfer process and presented five common components of the knowledge transfer process: problem identification and communication; knowledge/research development and selection; analysis of context; knowledge transfer activities or interventions; and knowledge/research utilisation. Mitton et al. [142] studied knowledge transfer between research producers and research users, and found that only 20% of the studies analysed reported on

a real-world application of a knowledge transfer and exchange strategy, and that a small number had been formally evaluated.

In a review of knowledge transfer in healthcare, Huberman et al. [97] suggests that an intensified interaction between researchers and practitioners is an important variable to explain the level of implementation of findings and the establishment of future collaborations.

Researchers have identified similar results when reviewing the evidence of implementation of Operational Research (the discipline of applying advanced analytical methods to help make better decisions [208]) in healthcare [25, 24]. A report presented after the *Festival of Evidence 2015 meeting in London* by Brailsford et al. [22] concluded that despite the demonstrated benefit of modelling and simulation to improve decision-making in healthcare, there is a lack of evidence of successful implementation of modelling interventions beyond the original client organisation. The re-use of simulation models in healthcare was also discussed in [60], as part of a study aiming to develop a generic hospital simulation model. A discussion of re-use of simulation models will in our research be presented in Chapters 3 and 4.

Marshal et al. [135] studied the *Research-in-Residence* model of participation as an approach to move improvement research closer to practice and presented experiences with this model in three different contexts related with healthcare. As noted in [135], *this model positions the researcher as a core member of a delivery team, actively negotiating a body of expertise which is different from, but complementary to, the expertise of managers and clinicians.* The study concluded that this model has the potential to engage both academics and practitioners in the promotion of evidence-informed service improvement, but further evaluation is required before the model should be routinely used in practice.

Artificial Intelligence and Machine Learning (ability to learn autonomously from data) also present unprecedented opportunities for improving the healthcare sector, mainly because the increasing amount of heterogeneous health-related information generated by medical devices; medical apps; electronic medical records; omics data; and many others. However, there are still challenges in the context of healthcare (e.g., volume and variety of data; specific challenges in each step of the learning process; and the limited adoption of technology by healthcare professionals) [223]. A critical standing bottleneck is the lack of data, due to both quantity/quality and challenges with sharing clinical information for research and development [145].

Researchers have identified the same gap between demand and supply across Metaheuristic algorithms and Data mining in healthcare [211, 230, 75, 172, 5, 104]. Gatheri

et al [75] identified the lack of interaction between researchers and physicians, and the unfamiliarity of mathematics and computer science among healthcare professionals as the main factors.

As noted, a range of different participatory approaches has been studied over the years. Nevertheless, in the heath sector such approaches have rarely moved beyond small scale projects and few have been integrated into routine practice [135]. One of the main limitations of these studies is the focus upon one specific data-driven methodology. ***The question remains whether an approach without borders in terms of academic societies and field of study could help to tackle the lack of impact of the data-driven approaches in the enhancing efficiency and value in healthcare?***

1.4 Research aims

Having argued in the previous sections that there is a gap between theory and practice related with data-driven methodologies, this Thesis investigates the impact of Machine Learning, Metaheuristics, Modelling & Simulation, and Data Analytics in increasing efficiency and value in healthcare, through analysing strategies to expand the impact of those data-driven approaches. The research is based on solving practical problems for particular context in different levels of healthcare management, with research drawing upon both theoretical and empirical analysis.

My primary purpose is to demonstrate that multidisciplinary knowledge and cross-discipline approach is required to create value through data-driven methodologies in healthcare.

I will explore this aim through the results of studies focusing on a wide range of different healthcare organisations, tackled with cross-disciplinary approaches.

Based on these studies, the aims of this research may be divided and concisely represented by the following principle objectives:

- Investigate the relations, similarities and differences between Machine Learning, Metaheuristics, Modelling & Simulation, and Data Analytics.
- Analyse activity, capacity and demand of a cancer diagnostic services across a region.
- Predict short-term and long-term demand for Knee Arthroplasty at a national level.

- Improve operational performance and patient experience in a hospital department.
- Solve a complex optimisation problem in the healthcare sector.
- Analyse and extract insight from an organisational email network.
- Develop a new healthcare system performance metric.

A review of these objectives will be presented within the final part of this Thesis (Part VI), both to evaluate their contribution and to present a general point of reference for this research.

1.5 Limitations

In this Thesis the term data-driven is used to describe approaches such as the implementation of Metaheuristics algorithms; Artificial Intelligence & Machine Learning; Operational Research methods such as Optimisation & Simulation; Data Analytics; Data Mining; Data Visualisation; etc. Though there are several excellent general reviews of these methodologies, each to some extent reflects the author's personal research interests and expertise. Due to the pace of development and breadth of research, a truly comprehensive review is probably impossible, and certainly beyond the scope of this Thesis. Chapter 2 presents a brief review of the properties of Machine Learning, Metaheuristics Optimisation, Modelling and Simulation, in addition to the principles of Social Network Analysis.

I will not attempt here to evaluate all aspects of those approaches. My intention is to use real-world examples to demonstrate, based on my experience, how data-driven approaches can help to improve health services.

1.6 Methods

I have based my study on applications of data-driven methodologies to real-world problems in health organisations in Spain and the UK.

These applications are divided into four main sections.

In the first part, a variety of techniques from Modelling & Simulation will be studied and used to analyse current performance and to model improved and more efficient future states of healthcare systems. The focus will be primarily concerned with the analysis of capacity, demand, activity, and queues both at hospital and

population levels. In the second part, Genetic Algorithm will be studied and used to solve a Routing Home Healthcare problem. In the third part, Social Network Analysis will be studied and used to visualise and analyse email networks. In the final, a new healthcare system performance metric will be proposed and implemented using a case study. New frameworks to implement these methodologies in real-world problems are going to be presented through the Thesis.

Several projects developed and implemented in collaboration with the University of Southampton, Wessex Academic Health Science Network (AHSN), and NHS England will be discussed. The work aims to increase early detection of cancer and thereby reduce premature mortality. The research was conducted working closely with NHS Trusts and Clinical Commissioning Groups (CCGs) across the Wessex region in England to produce bespoke service modelling, as well as population level models. At a regional level, a Colorectal Cancer Screening Programme model developed in the South of England, and an analysis of endoscopy activity, capacity and demand across the region will be presented. This works includes the estimation of regional future demand for endoscopy services in five years' time.

A Genetic Algorithm metaheuristic will be implemented and applied in a variant of the Home Health Care Problem (HHCP), focusing on the route planning of clinical homecare.

Estimation of future utilisation scenarios of knee arthroplasty (KA) revision in the Spanish National Health System in the short (2015) and long-term (2030), and their impact on primary KA utilisation will be also described. This work was developed in collaboration with the IMIM (Hospital del Mar Medical Research Institute, Barcelona) and the Agency of Health Quality and Assessment of Catalonia (Agència de Qualitat i Avaluació Sanitàries de Catalunya [AQuAS], Barcelona).

A Social Network Analysis (SNA) project will be also presented. This work was developed in collaboration with the Wessex Academic Health Science Network (AHSN) in the UK. The analysis focuses on organisational email knowledge extraction with SNA and Data Mining.

Finally, a new healthcare system performance metric - based on the Overall Equipment Effectiveness (OEE) measure - will be proposed and evaluated using real data from an UK-based Endoscopy Unit.

To summarise, this work identifies four key techniques to use in the investigation of health data - Machine Learning algorithms, Metaheuristic optimisation, Modelling & Simulation and Data Analytics. Following a review of the different subjects and

its associated issues, these techniques are going to be evaluated and used with an applied-focuses to solve healthcare problems.

1.7 Outline of the Thesis

Chapter 2 presents a review of background information for the Thesis. Chapter 3 describes two studies: Section 3.1 reviews the modelling of colorectal cancer diagnostic service pathways, and Section 3.2 summarises a study focused on predicting the burden of revision Knee Arthroplasty. Chapter 4 includes Modelling & Simulation of an endoscopy department. Chapter 5 describes the implementation of Genetic Algorithms to solve a routing healthcare services problem. Chapter 6 reviews an organisational email knowledge extraction with Social Network Analysis and Data Mining. Chapter 7 proposes a new healthcare system performance metric. Chapter 8 presents a discussion of findings, contributions and implications of the research for policy and practice in healthcare. Finally, Chapter 9 presents conclusions and directions for future research.