

Imperial College London
Imperial College Business School

**ACCOUNTING FOR BEHAVIOURS AND CONTEXT IN
EVALUATIONS OF COMPLEX HEALTH INTERVENTIONS**

Tiago Cravo Oliveira

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy

DECLARATION

I declare that this thesis submitted for the degree of Doctor of Philosophy has not been submitted for any other degree of professional qualification, is my own composition, and that, unless otherwise referenced, the material presented herein is my own original work, undertaken under the supervision of Professor James Barlow and Dr. Steffen Bayer at Imperial College Business School, between January 2010 and January 2014.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Tiago Cravo Oliveira

January 2014

ABSTRACT

Health care systems across developed countries face a perfect storm of rising demand and constrained funding. Systems have relied so far on short-term fixes but the time for incremental piecemeal solutions is passing. To achieve transformational change and fundamental service redesign, policy makers are resorting to ever more complex interventions. Evaluating their effects is far from trivial.

From public health programmes, to integrated and community care services, to electronic health technologies, complex health interventions typically exhibit a large number of components and interactions among them and other parts of the system; involve numerous intricate behaviours by those delivering and receiving the intervention; engage multiple and diverse groups, organisational levels and populations; result in many outcomes, typically with a high degree of variability; and are extensively tailored to local settings and circumstances. Evaluating such interventions is as much about whether they work, as how and why.

In this research, I examine the difficulties in using standard economic evaluation methods to assess complex interventions in the outpatient setting, and develop an approach to evaluation which uses methods and techniques that can explicitly address complexity, incorporate preferences and behaviours of patients and carers, and account for wider contextual influences.

I apply the suggested approach to the evaluation of teleconsultation in Alentejo, drawing on insights from previous theoretical and empirical research, new econometric and statistical studies, and simulation modelling. The application makes contributions to extant research on behaviour and decision making, and has implications for the evaluation of teleconsultation, as well as for broader discussions of how to assess complex interventions.

Complex health interventions have the potential to deliver a revolution in health care, but to achieve it we must be able to identify those that truly work, how and why. It is hoped the approach suggested here will contribute to that objective.

DEDICATION

To my grandfather

ACKNOWLEDGEMENTS

I am indebted to my supervisors, Professor James Barlow and Dr. Steffen Bayer, for guiding my work throughout these last four years, and dedicating much of their time helping me turn concept into reality. It has been a pleasure working with them, learning from their experience and knowing that I could always rely on their support and advice. To James, for believing in what was initially no more than a rough idea, and to Steffen, for being indefatigable in these final months, I am truly grateful.

This research was partly funded by a doctoral grant from *Fundação para a Ciência e a Tecnologia*, Portugal (grant reference SFRH/BD/66305/2009), by the United Kingdom Engineering and Physical Science Research Council's Health and Care Infrastructure Research and Innovation Centre (HaCIRIC), and by the *Associação Portuguesa de Economia da Saúde* (APES). I am grateful for their support.

I am indebted to all those who took part in my primary studies and who must, naturally, remain anonymous. Conducting research in social sciences is not possible without people who give a bit of their time away, expecting nothing in return. They have my appreciation.

I would like to thank all the charming people I have come across in Alentejo, who have made each trip there an absolute pleasure. José António Estrompa, Fernando Miranda, José Robalo, Hugo Quintino, Sónia Martins, Filomena Mendes, Rosa Matos, Maria José Branquinho, Horácio Feiteiro, Penalva Santos, Fátima Ferreira and Fátima Breia, thank you.

I am especially indebted to Luís Gonçalves, for his unwavering support and belief in me, to Manuela Rosado, for her relentless willingness to help and constant smile, and to Manuel Murta, for being such a delightful collaborator. I extend a special thank you to Manuel Grandal, for being such a joyful host.

I am thankful for the expert assistance and advice I have received throughout my research, namely from Professor John Rose, Professor William Greene, Dr. Mylene Lagarde, Dr. Mauro Laudicella, Dr. Katharina Hauck, Dr. Benita Cox, Dr. Céu Mateus and Professor Peter Smith. I am also grateful to the participants of the seminars and conferences where various parts of this thesis were presented¹, especially Dr. Marta Soares for her detailed comments as discussant of my paper at the 4th Workshop of APES.

Parts of this thesis have also been published elsewhere: Oliveira, T. C., Barlow, J., Gonçalves, L., & Bayer, S. (2013). *Teleconsultations reduce greenhouse gas emissions*. Journal of health services research & policy, 18(4), 209-214.; Cravo Oliveira, T., Bayer, S., Gonçalves, L., & Barlow, J. (2013). *Telemedicine in Alentejo*. Telemedicine and e-Health, forthcoming; Cravo Oliveira, T., Branquinho, M. J., & Gonçalves, L. (2012). State of the Art in Telemedicine—Concepts, Management, Monitoring and Evaluation of the Telemedicine Programme in Alentejo (Portugal). In *Perspectives on Digital Pathology: Results of the Cost Action IC0604 Euro-Telepath*, 179, 29.

¹ Parts of this thesis were presented in the following conferences and seminars: HaCIRIC International Conference, Edinburgh, September 2010; Healthcare management seminar, London, September 2010; Annual Conference of the UK Chapter of the System Dynamics Society, London, February 2011; 37th Annual Meeting of the EURO working group on Operational Research Applied to Health Services, Cardiff, July 2011; HaCIRIC International Conference, Manchester, September 2011; 12th National Health Economics Conference (APES), Lisboa, October 2011; VIII Forum Iberico de Telemedicina, Alandroal, November 2011; Annual Conference of the UK Chapter of the System Dynamics Society, London, February 2012; European Conference on Health Economics, Zurich, July 2012; 30th International Conference of the System Dynamics Society, St. Gallen, July 2012; 4th Workshop of the APES, Évora, September 2012; HaCIRIC International Conference, Cardiff, September 2012; IX Forum Iberico de Telemedicina, Badajoz, November 2012; Annual Conference of the UK Chapter of the System Dynamics Society, London, February 2013; 31st International Conference of the System Dynamics Society, Cambridge, July 2013; X Forum Iberico de Telemedicina, Lisboa, October 2013; 13th National Health Economics Conference (APES), Braga, October 2013; Duke-NUS Health Services and Systems Research, Singapore, November 2013.

I am deeply appreciative of the support and collaboration I have received from so many lovely people at Imperial College Business School: Linda, Fawaz, Lucy, Hae-Kyung, Yeyi, May, Alexandra, Elena, Dimitrios, Anupa, Danielle, Theti, Anja and Inger. A warm and special thank you is due to Marisa, Eliana and Bora. And, to my close friends, in London and back home, who have made these last four years wonderful, a fond thank you.

Finally, to my parents Isaura and Anibal, and to my brother André, my affectionate and loving gratitude is due. Your kindness, encouragement and support spur me on. Thank you for giving me so much.

TABLE OF CONTENTS

DECLARATION	III
ABSTRACT	V
DEDICATION	VII
ACKNOWLEDGEMENTS	IX
TABLE OF CONTENTS	XIII
LIST OF TABLES	XVII
LIST OF FIGURES	XIX
LIST OF ABBREVIATIONS	XXIII
1 INTRODUCTION	1
1.1 Motivation	4
1.2 Research problem	4
1.3 Objectives	5
1.4 Scope	7
1.4.1 The outpatient care pathway	7
1.5 Structure of the thesis	10

PART 1 ADDRESSING COMPLEXITY IN HEALTH CARE EVALUATION	13
2 LITERATURE REVIEW	15
2.1 Determinants of patients' decisions	17
2.1.1 The decision to seek care	18
2.1.2 Adherence to and compliance with medical advice	23
2.2 Determinants of physicians' decisions	30
2.2.1 Diagnostic investigations and therapy recommendations	32
2.2.2 Referrals to specialist care	37
2.3 What can we learn from the literature	41
3 DEVELOPING AN APPROACH	47
3.1 Economic evaluation of complex interventions	48
3.2 Building on standard economic evaluation methods	51
3.2.1 Understanding complex adaptive systems	52
3.2.2 Understanding individual decision making	58
3.3 An approach to evaluating complex interventions	64
3.3.1 Simulation modelling techniques	67
3.3.2 Applying the approach in practice	74
PART 2 APPLYING THE APPROACH IN PRACTICE	79
4 THE COMPLEX HEALTH INTERVENTION	81
4.1 The evidence on real-time teleconsultations	83
4.1.1 A comprehensive picture of impact	84
4.1.2 Economic impact and methodological issues	90
5 TELECONSULTATION IN ALENTEJO	95
5.1 Impact of the programme	99
5.1.1 Exploratory interviews	99
5.1.2 Survey	101
5.1.3 Teleconsultation in other regions of Portugal	105
5.2 Evaluating teleconsultation in Alentejo	106
5.2.1 Studies and techniques	108
6 WAITING TIMES AND UTILISATION	111
6.1 Background	112
6.1 A model of referrals	115

6.2 A model of non-completion	117
6.3 The data set	118
6.4 Estimation	123
6.4.1 Poisson and negative binomial regression models	125
6.4.2 Panel models for count data	127
6.4.3 Models estimated	131
6.5 Results	132
6.5.1 Model prediction	137
6.6 Discussion	139
6.6.1 Limitations and future work	140
6.7 Conclusion	140
6.7.1 Implications for the evaluation of teleconsultation in Alentejo	141
7 THE DETERMINANTS OF GENERAL PRACTITIONERS' REFERRALS	143
7.1 Background	144
7.2 Methods	147
7.2.1 Development of attributes and levels	148
7.2.2 Experimental design	151
7.2.3 Recruitment and data collection	152
7.3 Econometric analysis	152
7.3.1 Model specification	153
7.4 Results	155
7.4.1 Estimation results	158
7.4.2 Model prediction	162
7.5 Discussion	166
7.5.1 Limitations and future work	171
7.6 Conclusion	172
7.6.1 Implications for the evaluation of teleconsultation in Alentejo	173
8 THE SIMULATION MODEL	175
8.1 Background	178
8.2 A model of teleconsultation in Alentejo	179
8.2.1 Primary and secondary care capacities	187
8.2.2 Implementation of teleconsultation	188
8.2.3 Continued GP education and referral rates	189
8.3 Specification and parameters	191
8.3.1 Simulation parameters	192

8.4 Model validation and testing.....	197
8.4.1 Validation in the context of system dynamics	197
8.4.2 Validation tests.....	200
8.5 Simulations and scenarios	202
8.5.1 Practice-level simulations.....	202
8.5.2 Regional level simulations	209
8.5.3 Policy scenarios	215
8.6 Sensitivity analyses	222
8.6.1 Sensitivity to distance and equipment costs	222
8.6.2 Sensitivity to effect of continued GP education on referrals	224
8.6.3 Sensitivity to elasticities of demand and non-completion	224
8.7 Discussion	226
8.7.1 Limitations and future work	229
8.8 Conclusion.....	230
8.8.1 Implementation of results and stakeholder acceptance.....	231
9 DISCUSSION AND CONCLUDING REMARKS	233
9.1 Summary.....	236
9.2 Contributions.....	238
9.3 Policy implications.....	240
9.4 Limitations and future work.....	242
REFERENCES.....	243
APPENDIX A INTERVIEW SCHEDULE AND GUIDE	263
APPENDIX B SURVEY INSTRUMENT.....	271
APPENDIX C THE DISCRETE CHOICE EXPERIMENT	277
APPENDIX D THE WAITING TIME DATA SET	281
APPENDIX E SIMULATION MODEL DOCUMENTATION.....	285

LIST OF TABLES

Table 3-1 Choosing the right simulation technique	74
Table 5-1 Advantages of teleconsultation as reported in exploratory interviews.....	100
Table 5-2 Selected results from survey of 200 patients in Alentejo	104
Table 5-3 The right simulation technique to evaluate teleconsultations in Alentejo	109
Table 6-1 Average waiting times for dermatology appointments, 2001-2013	120
Table 6-2 Variable description and summary statistics of practice/municipality data	122
Table 6-3 Static panel negative binomial regression models of REFT	133
Table 6-4 Dynamic panel count models of REFT	135
Table 6-5 Static panel negative binomial regression models of NCT	136
Table 6-6 Dynamic panel count models of NCT	137
Table 7-1 Descriptions of attributes and levels.....	151
Table 7-2 Characteristics of respondents and practices: descriptive statistics	156
Table 7-3 Characteristics of respondents and practices: categorical variables.....	157
Table 7-4 Referral decisions by clinical presentation	158
Table 7-5 Results of random effects logit models of decision to refer / not refer	159
Table 7-6 Results of logit models of choice of priority	160
Table 7-7 Base cases used to determine marginal effects	163
Table 7-8 Marginal changes in probabilities of referral for different conditions	166

Table 8-1 Unit costs and auxiliary costing variables used in the model	193
Table 8-2 Parameters used in both practice-level and regional simulations	195
Table 8-3 Summary of practice-level results before/after teledermatology	208
Table 8-4 Summary of regional level results before/after teledermatology	213
Table 8-5 Impact of joining teledermatology practices on regional-level outcomes	214
Table 8-6 Impact of changes to reimbursement and co-payment of teledermatology.....	218
Table 8-7 Impact of two allocation policies: equitable and expanding	221
Table 8-8 Impact of different elasticities on perceived waiting times and unmet demand..	225

LIST OF FIGURES

Figure 1-1 The outpatient care pathway (based on “ <i>stages in the ‘pathway of care’</i> ” identified in van Ackere and Smith (1999)).....	8
Figure 2-1 The patient’s decision to seek care - from clinical need to actual health care utilisation (adapted and modified version of Figure 1 from Campbell and Roland (1996)) ...	20
Figure 2-2 Determinants of patients’ and physicians’ decisions identified in the literature ..	42
Figure 3-1 Suggested approach to evaluating complex health interventions	75
Figure 4-1 Summary of review of evidence on teleconsultations	89
Figure 5-1 Telemedicine network in Alentejo.....	97
Figure 6-1 Histograms and density plots for referrals per quarter per practice.....	123
Figure 6-2 Histograms and density plots for transformations of total referrals	124
Figure 6-3 Elasticity of referrals with respect to waiting time as a function of waiting time (estimated using a population-averaged negative binomial model)	138
Figure 6-4 Elasticity of non-completion with respect to waiting time as a function of waiting time (estimated using a population-averaged negative binomial model).....	138
Figure 7-1 Determinants of GPs’ referrals identified in the literature.....	146
Figure 7-2 Impact of distance to referral hospital on probability of referral.....	164
Figure 7-3 Impact of average waiting time for a normal priority appointment on the probability of referring to an urgent consultation.....	165
Figure 7-4 Distribution of GP age in sample	169

Figure 8-1 Simplified depiction of patient flows in dermatology outpatient care	180
Figure 8-2 The impact of waiting time on patient flows before the introduction of teledermatology (F2f: face-to-face)	183
Figure 8-3 Dermatology outpatient care after the introduction of teledermatology (Tele: teleconsultation; F2f: face-to-face)	185
Figure 8-4 Primary and secondary care capacities after the introduction of teledermatology	187
Figure 8-5 Implementation of teledermatology at practice-level and destination of referrals	188
Figure 8-6 Impact of continued GP education and learning on referral rate	189
Figure 8-7 Stocks and flows at the regional level (feedback effects omitted for readability)	196
Figure 8-8 Actual waiting times for dermatology appointments	204
Figure 8-9 Perceived waiting times for dermatology appointments	204
Figure 8-10 Referral rates per 1000 inhabitants	205
Figure 8-11 Non-completion rates per 1000 inhabitants	205
Figure 8-12 Referral rate to teledermatology before/after reduction from continued education	209
Figure 8-13 Actual waiting times for dermatology appointments	211
Figure 8-14 Perceived waiting times for dermatology appointments	211
Figure 8-15 Referral rates per 1000 inhabitants	212
Figure 8-16 Non-completion rates per 1000 inhabitants	212
Figure 8-17 Referral rate to teledermatology before/after reduction from continued education, with and without a coordinator who performs all teleconsultations	216
Figure 8-18 Impact of equitable allocation policy on teledermatology waiting time and rates	219
Figure 8-19 Perceived teledermatology waiting time for expanding allocation policy	220
Figure 8-20 Sensitivity to different assumptions concerning the learning effect	224

Figure 9-1 Suggested approach to evaluating complex health interventions	235
---	-----

LIST OF ABBREVIATIONS

<i>A&E</i>	Accident and emergency
<i>ABM</i>	Agent-based modelling
<i>ACSS</i>	Central Administration of the Health System
<i>AIC</i>	Akaike information criteria
<i>ARSA</i>	Alentejo Regional Health Authority
<i>BIC</i>	Bayesian information criteria
<i>CRD</i>	Centre for Reviews and Dissemination
<i>DCE</i>	Discrete choice experiment
<i>DES</i>	Discrete event simulation
<i>GHG</i>	Greenhouse gas
<i>GMM</i>	Generalised method of moments
<i>GP</i>	General practitioner
<i>HESE</i>	Hospital do Espírito Santo de Évora
<i>HIV/AIDS</i>	Human immunodeficiency virus / acquired immunodeficiency syndrome
<i>IID</i>	Independent and identically distributed
<i>INE</i>	Statistics Portugal
<i>IP</i>	Public institution
<i>NB1</i>	Negative binomial 1
<i>NB2</i>	Negative binomial 2

<i>NHS</i>	National health services
<i>NICE</i>	National Institute for Health and Care Excellence
<i>OECD</i>	Organisation for Economic Co-operation and Development
<i>QALY</i>	Quality-adjusted life year
<i>ROC</i>	Receiving operator curve
<i>SD</i>	System dynamics
<i>SSRI</i>	Selective serotonin reuptake inhibitors
<i>UK</i>	United Kingdom
<i>US</i>	United States

Chapter 1

INTRODUCTION

Health care systems across developed countries face a perfect storm of rising demand and constrained funding. As demand for medical care keeps rising due to ageing, chronic diseases, unhealthy behaviours, patient expectations and technological innovation, there are reasons to believe the decades of continued health expenditure growth above gross domestic product have come to an end. The global economic crisis has led to dramatic reversals in health spending in those countries most affected by the downturn, and spending per capita has fallen in a third of Organisation for Economic Co-operation and Development (OECD) countries between 2009 and 2011 (OECD, 2013a). The fact that cuts have focused on prevention programmes and pharmaceuticals, has led to concerns that short-term budget gains will eventually be offset by longer term needs (OECD, 2013b). Indeed, average OECD growth rates for outpatient, inpatient and long-term care (which combined account for more than half of total health spending) were positive between 2009 and 2011 (OECD, 2013a).

Portugal, Greece and Ireland – countries bailed out by the European Central Bank, the European Commission and the International Monetary Fund – cut health spending per capita by respectively 2.2, 6.6 and 11.1 percent between 2009 and 2011. Concurrently, other

sectors of the economy also experienced cuts. Given that health is partly determined by the broader social context, cuts to social welfare spending and increased unemployment may lead to greater health care needs (McKeeargue, 2010, Stuckler et al., 2009). Furthermore, it is the most vulnerable, in low-income groups, that are more likely to forego care, with potentially dire long-term consequences (OECD, 2013b).

Even if the economic crisis had not happened, continued excess health spending above economic growth would eventually need to be compensated by diverting personal income and economic resources from other sectors of the economy to health care (Chernew et al., 2003, Chernew et al., 2009). Public health care services, especially, can no longer expect governments to pay for cost growth as in the past. Countries under economic adjustment programmes, such as Greece, Ireland and Portugal, are no longer able to finance excess spending through borrowing. Portugal, for example, could take at least a decade to address the structural challenges it faces (Wise, 2013). Others, such as the United Kingdom (UK), have told health care executives not to rely on any significant net increase in annual funding from 2016 (Department of Health, 2013a). Faced with unwavering demand, the implications for health care services could be dramatic.

In the UK, for example, it has been suggested that the National Health Service (NHS) could face a decade of austerity and a significant funding gap by 2021 (Roberts et al., 2012). Anticipating this state of affairs, in 2008 Sir David Nicholson, chief executive of the UK NHS, set up the Quality, Innovation, Productivity and Prevention programme to deliver so-called *efficiency gains* (put simply, reductions in excess costs that could be re-invested to make up for funding constraints) in the order of £15-20 billion by 2015, equivalent to 4 percent each year from 2011 (House of Commons Health Committee, 2013). There is considerable scepticism that the UK NHS can achieve this (The King's Fund, 2013), especially as it has never made gains of 4 percent in any one year, much less four years in succession (Lord Norman Warner, 2013).

It is increasingly clear that the time for incremental piecemeal change is passing. Health care systems are “*living on borrowed time*” and need “*major change at scale and pace*” (Lord Norman Warner, 2013). So far, solutions have relied almost exclusively on “*short-term fixes rather than (...) long-term transformations*” (Department of Health (2013a), page 5). What is needed is “*large scale transformational change*” and “*fundamental service redesign*” (Department of Health (2013a), page 5). Solutions must adapt to circumstances and contexts, as “*there are no ‘one size fits all’ blueprints for integrated care and support models*”, no silver bullet approach (Department of Health (2013a), page 12). A revolution in planning, financing and delivery of health care is needed.

Policy makers, physicians and patients across different parts of the system, at different levels, are championing this revolution. Health care initiatives are shifting the location of care, affecting multiple parts of the system simultaneously. They are challenging professional boundaries, roles and responsibilities. Their scope is diffuse and dynamic as a consequence of learning and re-invention. They have broad effects, both clinical and non-clinical, and within and beyond the health care system. They are often driven by necessity and initiative, rather than any evidence of cost-effectiveness. They are flexible and contextual, tailored to meet the needs and preferences of specific populations and settings. They are leading to *transformational change*, *fundamental service redesign*, and *integration*.

There is no doubt that health care systems, notably those financed through public funds such as the Portuguese and UK NHSs, need to change to become sustainable. It is also beyond dispute that change of this magnitude requires fundamentally new ways of planning and delivering care. It is less clear how to assess and evaluate these new ways of working. In this thesis, I examine the challenges involved in evaluating complex health interventions. Building on conventional economic evaluation methods, I propose an approach which explicitly acknowledges that costs and consequences of complex interventions are a direct result of behaviours of those delivering and receiving them, and that the wider context in which those behaviours take place should not be disregarded.

1.1 Motivation

Complex health interventions exhibit a large number of components and interactions among them and other parts of the health care system; involve numerous intricate behaviours by those delivering and receiving the intervention; engage multiple and diverse groups, organisational levels and populations; result in many outcomes, typically with a high degree of variability; and are extensively tailored to local settings and circumstances (Craig et al., 2008). They often challenge existing boundaries and interfaces and are, by nature, new: new tasks and interactions performed in new settings and contexts (Shepperd et al., 2009). Their impact and effects are a direct consequence of the active involvement of those delivering and receiving the interventions (Pawson et al., 2005). Examples include public health programmes, integrated and community care services and electronic health technologies.

Complex health interventions can lead to transformational change and fundamental service redesign and integration. They can also lead to unexpected and unintended consequences (Harrison et al., 2007b, Sveiby et al., 2009), making the problems they are trying to solve, even worse. To achieve the structural change that is needed, at scale and pace, we must be able to identify those interventions that actually work, when and where.

1.2 Research problem

Evaluating complex health interventions is very different from evaluating drugs and medical devices. It is as much about whether they work as it is about how and why they work (Byng et al., 2008). Understanding the mechanisms of change can be as important as assessing final outcomes. While standard methods of economic evaluation tend to treat interventions as black boxes, with little regard for mechanisms and a focus on final impact (Anderson, 2008), assessing complex interventions is precisely about the behaviours of those delivering and receiving them (MacKenzie et al., 2010). Due to their contextual and tailored nature, complex interventions are typically not standardizable (i.e., defined in terms of essential

components which if applied anywhere will produce the same effects). It is likely that findings from one setting will do little to inform impact in other settings, limiting the generalizability of evaluations, and their usefulness for priority-setting.

Because complex interventions often lead to improvements in intermediate outcomes, patient experience, process efficiencies, and benefits to the wider community and economy even, focusing on a single measure of outcome, such as the quality-adjusted life-year (QALY), is too narrow (Craig et al., 2008, Gerard and Mooney, 1993). While health gains are important, health systems have other objectives as well, such as providing an accessible service, promoting equality, valuing patient and provider preferences, and ensuring an efficient use of limited resources (Department of Health, 2013b).

It is increasingly clear that no single method can capture the wealth of effects of complex health interventions. Specifically, the gold standards of evaluation, cost-effectiveness analysis and randomised controlled trials, are unlikely to be as useful for assessing complex interventions as they are for evaluating drugs and medical devices (Drummond et al., 2007, MacKenzie et al., 2010). Agencies which so often have relied on these methods now acknowledge that evaluating complex interventions requires a mix of techniques and approaches, and the use of both experimental and non-experimental data (Drummond et al., 2007, Canadian Agency for Drugs and Technology in Health, 2006, Walach et al., 2006, Hill, 2004). New approaches to the evaluation of complex health interventions are needed.

1.3 Objectives

Complex health interventions have been the subject of much discussion and there is today an understanding of what is required to assess a complex intervention. Behaviours of, and interactions between, stakeholders are important (Anderson, 2008). The impact of a complex intervention is a direct result of active involvement from those delivering and receiving the

intervention, so that knowledge of how stakeholders make decisions is essential to understanding the outcomes (Pawson et al., 2005). Evaluation is as much about whether the intervention works, as how it works (Byng et al., 2008). To achieve such an understanding, evaluations should explicitly account for the interconnectedness and relationships between components and the wider context in which they exist (Craig et al., 2008, Shepperd et al., 2009). Given the likelihood that feedback, delays and non-linearities are important, evaluations should be able to account for these characteristics, signalling situations where counterintuitive behaviour and unintended consequences are a possibility. Ideally, the intervention should be assessed at different organisational levels. Evaluations that incorporate these features are likely to provide a better understanding of how complex interventions work and in what conditions.

The objective of this thesis is to develop a deeper understanding of what the evaluation of complex interventions entails and explore whether an approach which incorporates the requirements identified can be developed. Naturally, given what was written thus far, it would be senseless to now suggest there is a formal step-by-step process by which every complex intervention, in any field of health care, can be assessed. That is not the case. There are, however, as I will show, tools that can be integrated under a general process of evaluation.

There is no silver bullet method to the evaluation of complex health interventions. Unlike the evaluation of many drugs, which can be standardised through very explicit guidelines, evaluating complex interventions is precisely about diversity and flexibility. There is no one-size-fits-all reference case (e.g., see Table 2 in Drummond et al. (2007)) setting out the range of methodologies most appropriate for evaluating any given complex intervention. The external validity of results will always be an issue, and whether findings from one setting can inform decision-making in other settings will invariably have to be explored. The objective of this thesis is to explore whether it is possible to develop an approach which, while general enough to accommodate the need for flexibility, may still

provide insights for a broad array of complex interventions. In summary, this thesis seeks to answer the following questions:

- 1) What are the general requirements for evaluations of complex health interventions?
- 2) Are there tools and methods available to meet those requirements?
- 3) Can an approach be developed which incorporates those tools and methods?

1.4 Scope

Any evaluation of a complex health intervention requires in-depth understanding of the behaviours specific to the context and area of application. In this thesis, I focus on outpatient care as it is the setting for many complex interventions. First, it is at the interface between primary and secondary care, an important interface that is often challenged by complex interventions seeking bigger integration. Second, it is composed of many diverse actors, groups, organisational levels and populations, with potentially different objectives. Third, there is an important behavioural dimension and many interactions. Fourth, as it deals with *frivolous* as well as acute needs, it involves many outcomes. Fifth, there are significant contextual differences. And, finally, it accounts for a significant share of total health care spending.

1.4.1 The outpatient care pathway

Outpatient, or ambulatory, care can be seen as a series of decisions along a pathway (van Ackere and Smith, 1999), from the instant the patient decides to seek care up to the moment she leaves the system. Figure 1-1 is a possible representation of this pathway and the decision points along it. In it, clouds depict where patients enter and leave the system, boxes represent waiting times or lists (i.e., moments when patients await further progress through the pathway), and diamonds denote decision points (i.e., moments when decision makers – patients and/or physicians – must choose between different options).

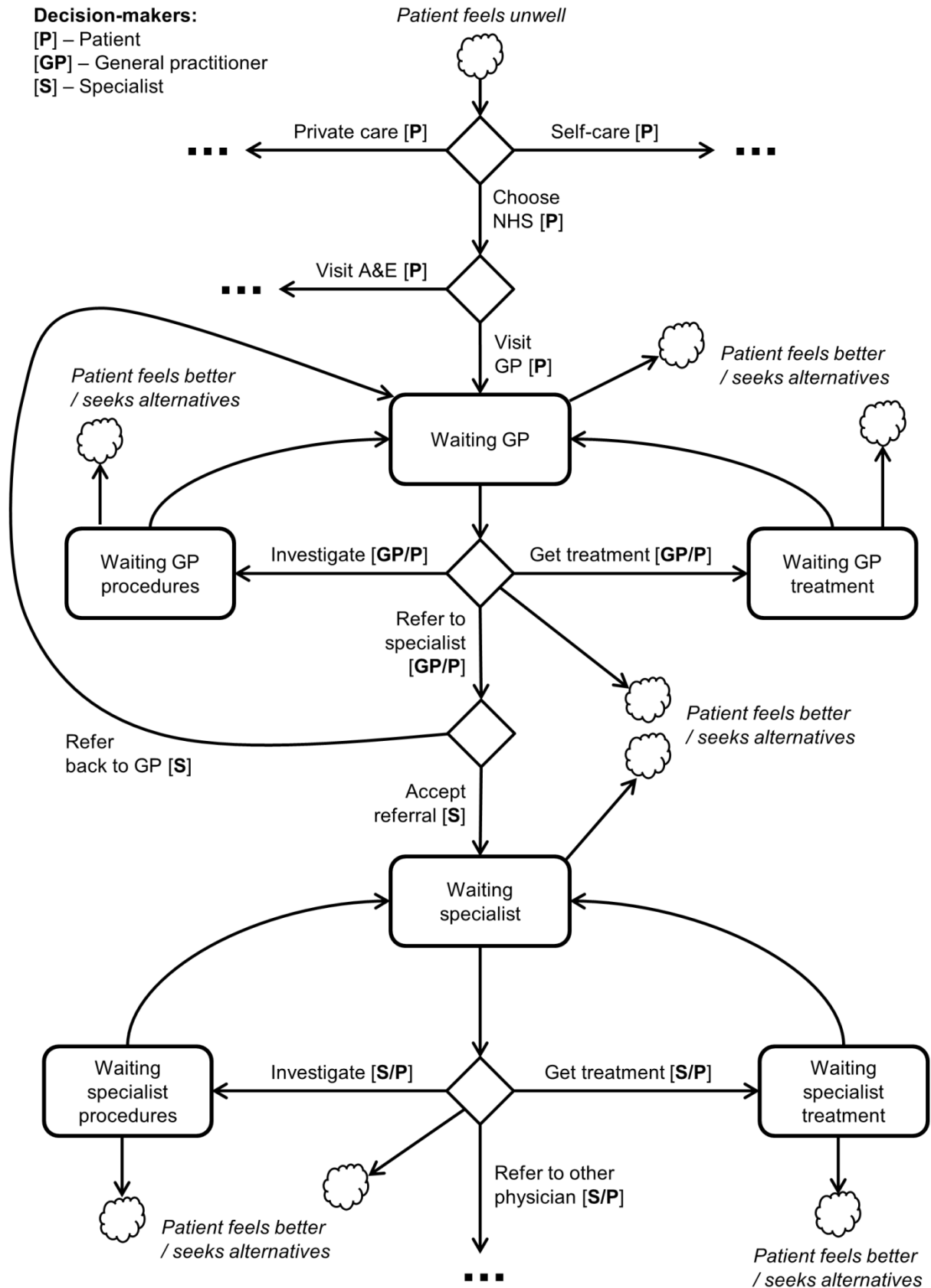


Figure 1-1 The outpatient care pathway (based on “stages in the ‘pathway of care’” identified in van Ackere and Smith (1999))

The first step concerns the choice of care. This decision can be viewed as a two-stage process. The patient first decides whether to seek private or public care. If she chooses the former, then two options are available: attend a private clinic or opt for self-care. A third option is to postpone the decision, which is equivalent to opting for no care (these are included in Figure 1-1 under self-care). If the patient chooses to seek public care (i.e., the NHS, as in the UK and Portugal) then again two general options are available: to visit the accident and emergency department (A&E) or to visit a general practice. For non-urgent care, visiting A&E should not really be an option. However, patients visit A&E for non-urgent complaints for a variety of reasons (see Sempere-Selva et al. (2001) and Singh (1988) for examples). The second option is to go to a general practice. In Figure 1-1, choices other than visiting a general practice (private care, self-care, A&E) lead to subsequent steps which are not illustrated. This is because the focus is on non-urgent NHS care and I assume that patients who visit A&E will be redirected to a general practitioner (GP).

Once a decision has been made to visit a GP, an appointment is scheduled. From this moment onwards, the patient waits for further progress between decision points. A decision point occurs when the patient meets with a physician or when a physician assesses the patient's need and motivation for further progress. In primary care, this happens when the patient sees the GP. A number of options are available. If the GP is uncertain about the diagnosis, then he/she can recommend diagnostic procedures to investigate. Should the patient accept, then a procedure is scheduled, and the patient waits for further progress. The procedure can be performed immediately and the GP assess the results at once, or it may have to be done elsewhere and the patient may need to schedule a new appointment with the GP at a later date. In either case, the decision to investigate eventually leads to a new decision point involving the GP and the patient. Alternatively, the GP may establish a diagnosis and propose treatment. Once again, the patient may accept or reject. If she accepts, she receives treatment and may eventually re-visit the GP for follow up.

The third option is for the GP to refer the patient to a specialist. If the patient accepts, she is added to a waiting list. In all the previous steps, the patient may get better or choose to seek alternative care (e.g., private care, self-care, etc.), in which case she exits the system. While in some cases the GP will know why a patient has left the system (e.g., when a patient is prescribed treatment, comes back for a follow up and informs the GP she is feeling better), in other situations the patient may decide to seek alternatives without even completing the GP's recommendations (there are various names for this, used in different contexts: non-completion, non-adherence, attrition, non-compliance, etc.).

If the patient is referred, the specialist assesses the request for a consultation (i.e., a process frequently called triage), and determines whether to refer the patient back to the GP (e.g., when there is insufficient justification for the referral or important patient data missing) or to accept the referral and book an appointment. If he decides for the latter, and the patient attends the appointment, there is a new decision point, now involving the specialist and the patient. As before, the specialist can propose courses of action, which the patient can accept or reject. The patient may also leave the system, if she feels better or decides to seek alternative care. The discussion in the previous paragraphs applies. Should the specialist decide to refer the patient to a second specialist, the process continues as if the patient had been referred by the GP.

1.5 Structure of the thesis

The thesis is organised into two parts. In Part 1, *Addressing complexity in health care evaluation*, I address the research questions put forth in Section 1.3. Part 1 is composed of two chapters. In the first one, Chapter 2, *Literature review*, I explore the literature on patients' and physicians' decisions in the outpatient care pathway, finding there is a rich theoretical and empirical body of research. While clinical need is an important factor, it is clear that non-clinical factors are also relevant, as well as perceptions and beliefs. There are numerous instances of dynamic complexity, feedback and counter-intuitive behaviour. The

number of factors investigated in the literature is astounding, confirming that to assess complex health interventions it is necessary to take into account the behaviours and preferences of those delivering and receiving them.

In the second chapter of Part 1, Chapter 3, *Developing an approach*, I examine the difficulties in applying standard economic evaluation methods in the context of complex health interventions in detail. I identify analytical tools and techniques which can be used to fill the gaps in previous theoretical and empirical research into behaviours and preferences. Building on conventional economic evaluation methods, I develop an approach to the assessment of complex interventions, which combines a mixture of methods and techniques that can explicitly account for behaviours, as well as the wider context in which they exist.

In Part 2, *Applying the approach in practice*, I illustrate how the suggested approach can be used to evaluate a specific complex health intervention in a specific setting. I explore in-depth the use of teleconsultations, more specifically teledermatology, in Alentejo, a Portuguese region. In Chapter 4, *The complex health intervention*, I review a rich body of literature on the costs and consequences of teleconsultation, and identify gaps in researchers and practitioners' understanding of how and why the intervention works. In Chapter 5, *Teleconsultation in Alentejo*, I discuss how the experience of using teleconsultation in Alentejo compares to previous empirical research. Chapters 4 and 5 provide numerous indications that teleconsultation is a complex health intervention, and that the application of standard economic evaluation methods has been limited and of poor quality. Teleconsultation involves multiple interdependent components, is associated with significant behavioural changes, leads to many diverse effects within and outside the health care system, and is often tailored to specific contexts and settings, at different levels of the health care system. As such, it provides a good case-study to illustrate the application of the suggested approach.

In the following two chapters, Chapters 6 and 7, I undertake two studies to explore patients and physicians' behaviours and preferences, as these are intrinsically related to the costs and consequences of the intervention, and should not be disregarded. In Chapter 6, *Waiting times and utilisation*, I take advantage of recent developments in panel models for count data to explore the impact of waiting times on referrals and non-completion. Given previous evidence that teleconsultations lead to considerable reductions in waiting times, and that elasticities of referrals and non-completion change as a function of waiting times, it is unrealistic to assume that behaviour will remain static. In Chapter 7, *The determinants of general practitioners' referrals*, I use a discrete choice experiment (DCE) to understand GP preferences for outpatient secondary care, and estimate, among other things, the impact of teleconsultations on referral rates due to continued GP education and learning.

Concluding Part 2, in Chapter 8, *The simulation model*, I build on the insights from previous chapters to develop a simulation model of teleconsultation in Alentejo. The model embodies instituted mental models of how and why the intervention works: the literature on physicians and patients' behaviour in the outpatient setting; previous evidence on the impact of teleconsultations; primary studies of teleconsultation in Alentejo such as surveys and exploratory interviews; and econometric analyses of referral and non-completion behaviours. It makes the consequences of existing mental models and worldviews transparent, with behaviours emerging from established assumptions and hypotheses.

Finally, in Chapter 9, *Discussion and concluding remarks*, I summarise the contributions of this thesis, discussing how the suggested approach informs future evaluation studies of complex health interventions. I also consider the limitations of the approach and identify avenues for future research.

Part 1

ADDRESSING COMPLEXITY IN HEALTH CARE EVALUATION

Health care systems are riddled with complexity (Plsek and Greenhalgh, 2001). They typically have fuzzy boundaries, which can encompass different elements at different points in time. Elements react to their environment using dynamic internalised rules and mental models, which may or may not be known by, or logical to, other elements. Systems within the system can combine and influence each other, making it difficult to understand one system without exploring others. System behaviour emerges from interactions among multiple dispersed and decentralised agents with potentially conflicting objectives. Predicting future states of the system is virtually impossible. Achieving transformational change and fundamental service redesign and integration, at scale and pace, faced with all this complexity, is a monumental task. To succeed we must be able to identify those interventions that work and those that will only make the situation worse.

In Part 1 of this thesis, I show that accounting for the behaviours of those delivering and receiving a complex intervention, and explicitly addressing the context in which those

behaviours take place, are key requirements for the evaluation of complex interventions. In Chapter 2, *Literature review*, I explore how patients and physicians make decisions, how they behave and what their preferences are, illustrating the wealth of information that is available to evaluation studies of complex interventions in the outpatient setting.

In Chapter 3, *Developing an approach*, I discuss why it is difficult to use conventional economic evaluation methods to assess complex health interventions. I describe analytical tools and techniques which can be used to fill the gaps in previous research into behaviours and preferences. Building on these tools and the success of economic evaluation methods, I develop an approach to the assessment of complex interventions which explicitly accounts for the role of behaviours, preferences, and the wider context in determining costs and consequences.

Chapter 2

LITERATURE REVIEW

The concept of a complex health intervention is fairly recent. There is no uniform definition of what makes an intervention complex. Researchers have instead described complex interventions in terms of the nature and degree of complexity. Craig et al. (2008) identifies five dimensions of complexity: the number of components that make up the intervention and interactions among them and other parts of the health care system; the number and intricateness of behaviours of those delivering and receiving the intervention; number and diversity of groups, organisational levels and populations targeted; the number of outcomes and their degree of variability; and the extent of tailoring to local settings. Shepperd et al. (2009) describe complex interventions in service delivery as: being delivered across the primary-secondary care interface; being delivered in new settings; and having an added behavioural dimension, with staff performing new tasks in a new context. Context is important because it influences and limits the decisions and interactions of people affected by the intervention (Shepperd et al., 2009). Pawson et al. (2005) characterise complex interventions as active programmes, meaning they work through the active involvement of individuals, so that knowledge of stakeholders' reasoning is essential to understand an intervention's outcomes. Again context is said to be important since the same intervention

can be applied in different ways in different places, and is subject to adaptation and reinvention (Pawson et al., 2005).

A recurring theme in the complex health interventions literature is the importance of actions and decisions made by those delivering and receiving the intervention (Shepperd et al., 2009), which are said to be influenced by contextual factors (MacKenzie et al., 2010). It is hypothesised that individual behaviours and preferences are a significant contributor to the heterogeneity that characterises the complex interventions field. Moreover, only through an informed understanding of the mechanisms underlying a specific intervention's effects can one establish how the intervention works and whether it would work elsewhere. A recent document providing guidance on the development, evaluation and implementation of complex health interventions provides a useful overview (Craig et al., 2008):

Making explicit use of theory to develop an intervention prior to testing, and incorporating insights from the theory into an explicit model of how the intervention might alter behaviour, or affect other links in the causal chain between intervention and outcome, may lead to better-developed interventions, and also to better-designed evaluations (...) The value of a causal modelling approach is that it makes explicit the choice of intervention points and associated measures along the causal pathway. This allows researchers to assess why interventions are effective or not, as well as how effective they are... (Page 17)

Standard methods of evaluation tend to treat interventions as black boxes (Anderson, 2008). While this may be relatively acceptable for drugs and medical devices (i.e., predominantly clinical interventions), it is less so for complex interventions that affect and are affected by behaviours and preferences. Opening the black boxes of behaviours and interactions can provide insights into how and why an intervention works, and whether it could work elsewhere. In the outpatient setting, this means exploring the determinants of the decisions illustrated in Figure 1-1. Why do patients seek care? Why do physicians refer

patients to specialists? Why do patients miss appointments or disregard physicians' recommendations? These are important research questions which have been investigated by previous studies. Answers to these questions provide insights into how an intervention *"might alter behaviour, or affect other links in the causal chain between intervention and outcome"* (Craig et al. (2008), page 17). It is thus important to explore those answers.

In the following sections, I review the literature on how patients and physicians make decisions in the outpatient care pathway and what affects them (in other words, the determinants of their decisions), using a narrative overview approach (Green et al., 2006). Because of the sheer size of the literature on the determinants of patients and physicians' decisions, and my interest in presenting a broad perspective on the topic, the review is not systematic. The focus is on selected relevant theories and studies of behaviour, informed by the decision points depicted in Figure 1-1. I conducted a preliminary search of the literature on Google Scholar using an initial set of keywords (e.g., *determinants, referrals, non-completion*, etc.). The preliminary searches were used to refine the choice of keywords and a snowballing approach used to identify articles from the reference lists of papers.

2.1 Determinants of patients' decisions

Patients are at the heart of health care services (Darzi, 2008, Department of Health, 2013b). Assertions that *"services must reflect, and should be coordinated around and tailored to, the needs and preferences of patients"* (Department of Health (2013b), page 3) acknowledge the fact that it is ultimately patients who decide whether to seek care, what type of care to seek, and whether to adhere to the recommendations of health professionals. As seen in Figure 1-1 in the previous chapter, to a larger or smaller extent, patients are involved in every decision concerning their care. In the following sections, I review relevant theories of how patients make decisions in the outpatient pathway. The first decision is whether to seek care and where to seek it.

2.1.1 The decision to seek care

When a patient feels unwell and decides to seek care, she may elect to: visit a private clinic, postpone care or choose self-care, visit an A&E department, or visit a GP. The first three options are henceforth referred to as alternative care. In this section, I review the literature on the determinants of patients' decisions to visit a GP, occasionally referring to alternative care whenever appropriate.

First and foremost, patients visit their GP because of a perceived medical need. However, that is not the sole reason. The concept of the *illness iceberg* refers to the fact that only a limited number of symptoms and complaints actually reach health care services, and more specifically GPs (Hannay, 1978, Hannay, 1980). The *illness iceberg* (the analogy comes from the fact that typically only one ninth of the volume of an iceberg is above water and thus visible) symbolizes the difference between clinical need and demand for medical care (in the analogy, the visible part of the iceberg). For various reasons, clinical need does not necessarily lead to demand for health care. First, there is uncertainty in the need and demand for medical care and its effectiveness (Arrow, 1963). Patients may not possess the information or competence to ascertain they need medical care, and even if they do establish it, they may not be able or willing to seek care. Imperfect or poor information may also lead them to seek care they would not have sought, if well-informed (Hurley, 2000). Second, health care is merely one of a number of inputs into the production of good health. Grossman (2000) proposes that demand for health care is really a derived demand for good health, a commodity which may be attained by other means besides medical care (e.g., diet, exercise, etc.). Interestingly, it has been suggested that the marginal effect of medical care on health is small, with socio-economic factors and lifestyle being more influential (Thornton, 2002). Finally, even if patients establish a medical need and are able and willing to seek care, there may be factors that limit their capacity to do so. For all this, the relationship between feeling unwell and showing up at the GP is far from straightforward. The outcome is

a result of interactions among multiple factors. These can be grouped into categories, as illustrated in Figure 2-1 (an adaptation and modification of Figure 1 from Campbell and Roland (1996)).

2.1.1.1 Factors affecting the decision to seek care

Campbell and Roland (1996) propose the decision to seek medical care is affected by non-clinical factors as well as medical need. These factors can be grouped into five sets: demographic, socio-economic, socio-psychological, and organisational factors and patient preferences. Demographic factors include age, gender and ethnicity (Campbell and Roland, 1996, Grimsmo and Siem, 1984, McKinlay, 1972, Scott, 2000, Andersen, 1995). Studies have shown that children and the elderly, women, and patients from certain ethnic groups, seek care more often. It has been suggested that women, younger and older patients visit their GP more because they are more likely to require services offered by the health care system (e.g., maternity care, contraception, long-term care, etc.). Differences in utilisation for different ethnic groups may reflect different propensities for disease among certain ethnic groups, or they may reflect non-clinical factors such as language and culture (Campbell and Roland, 1996).

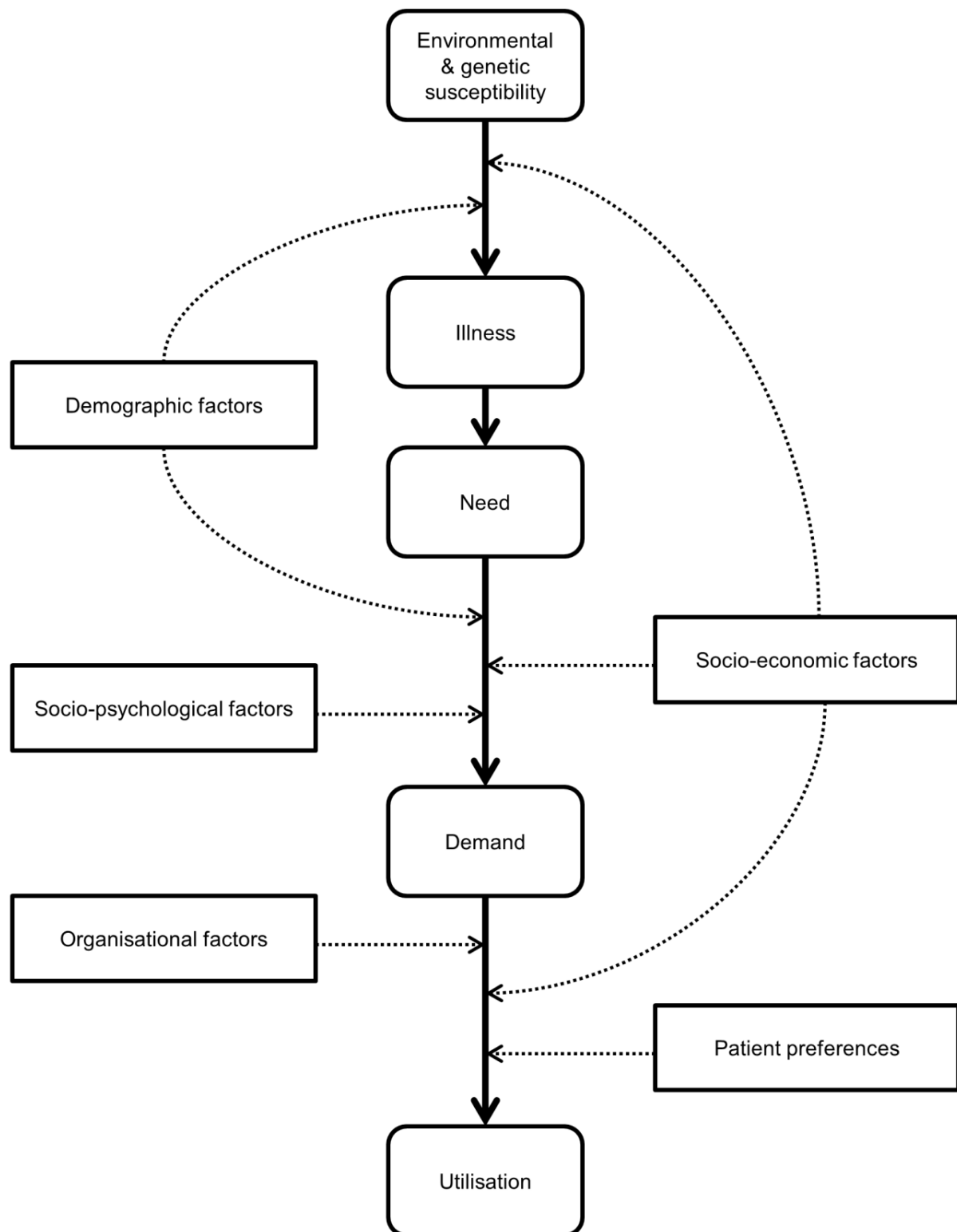


Figure 2-1 The patient's decision to seek care - from clinical need to actual health care utilisation (adapted and modified version of Figure 1 from Campbell and Roland (1996))

Socio-economic factors include social structure and support, education, employment status, income, health insurance, religion, socio-economic status, marital and civil status,

and home ownership (Andersen, 1995, McKinlay, 1972, Ensor and Cooper, 2004, Folland et al., 2006, Ringel et al., 2002, Zweifel and Manning, 2000, Scott, 2000, Grimsmo and Siem, 1984, Campbell and Roland, 1996). Increased utilisation is associated with weaker social structures and limited support (Grimsmo and Siem, 1984). This would explain why the unemployed, tenants, divorced and widowed consult more often (Campbell and Roland, 1996). Besides weaker social structures, these groups of patients may also be more vulnerable to financial constraints (e.g., the recently unemployed). Financial capability seems to be the underlying mechanism through which income, socio-economic status, education and having health insurance affect utilisation. The effect is two-sided. On the one hand, higher socio-economic status is associated with better living conditions, less susceptibility to disease and a greater ability to deal with disease when it is present. All this leads to less need, and consequently demand, for health care. On the other hand, higher income, longer education, and having health insurance have all been associated with increased use of services. The context also matters: when health care services are free at the point of delivery or heavily subsidised (e.g., in the UK and Portuguese NHSs), financial constraints will play a bigger part than financial advantage in determining utilisation. Finally, Ensor and Cooper (2004) point out that religious beliefs may not permit medical care.

Socio-psychological factors are embodied by the health belief model, first developed in the 1950s by social psychologists Godfrey Hochbaum, Irwin Rosenstock and Stephen Kegels, to explain poor participation in health prevention and screening programmes (Strecher and Rosenstock, 1997). The model has been further developed and extensively used since then. Later versions of the model propose that when patients are confronted with *cues to action* such as symptoms or pressure from family and friends, subsequent decisions are a consequence of four factors (Leavitt, 1979): the perceived susceptibility and vulnerability to illness; the perceived severity of the illness either physically or in terms of social functioning; the perceived benefits associated with actions that reduce the vulnerability or the threat; and the perceived barriers to those actions. The emphasis that the

health belief model puts on perceptions is essential towards explaining why people with the same symptoms may choose completely different courses of action. For example, Lydeard and Jones (1989) found that patients visiting the GP with dyspepsia complaints did not have significantly more frequent symptoms than those that did not visit; rather, visitors tended to overestimate the severity and frequency of complaints compared to non-visitors. Both groups had comparable understandings of the condition and anatomy. Anxiety about a symptom and pressure from family and friends to seek care are also important predictors of utilisation (Martin et al., 1991). A potential initial course of action to reduce vulnerability or threat is self-care (Jones, 2000). Anderson et al. (1977) explored differences between users and non-users of GP services. They found that infrequent users perceived themselves as healthier (even after controlling for morbidity), resorted more frequently to self-care, appeared more willing to wait for treatment to take effect, were more willing to deal with discomfort, displayed less inclination to seek care for potentially serious symptoms, and perceived more barriers to visiting the GP (e.g., waiting time, loss of wages, travelling time, etc.). It has also been proposed that current perceptions are affected by past experiences with care (Ensor and Cooper, 2004).

Organisational factors relate to the availability and accessibility of health care services: geographic proximity (which affects travelling time and costs), waiting time, appointment systems, price, number of GPs, regularity of service and availability of alternative services (Andersen, 1995, Anderson et al., 1977, McKinlay, 1972, Ensor and Cooper, 2004, Folland et al., 2006, Ringel et al., 2002, Scott, 2000, Grimsmo and Siem, 1984, Campbell and Roland, 1996). Previous research shows that limited availability and accessibility of services result in lower rates of utilisation. Organisational factors greatly influence the relationship between clinical need and demand for medical care.

Finally, patient preferences are increasingly relevant. Wealth and technology have changed patient expectations of health care services. Patients want to have a bigger role in decisions concerning their own health care and desire a more personal and individualised

experience from health systems (Darzi, 2008). As people learn to manage and live with their chronic conditions, coupled with a profusion of medical information readily available online, patients are becoming experts, at times more knowledgeable than their physicians (Donaldson, 2003). The physician-patient relationship has evolved to one of shared decision making rather than paternalistic agency (Vermeire et al., 2001). The recent application of stated choice methods to health care has allowed researchers to empirically address patient preferences resulting in a wealth of studies in various contexts and settings. These have shown that patients have clear preferences and are willing to trade-off certain aspects of services in return for improvements in others. Turner et al. (2007) estimate the relative importance to patients of continuity of care in general practice. Continuity of care is defined in three ways: informational (the use of past information to determine current care), management (a consistent and coherent approach to managing care), and relational (having an ongoing relationship with one or more providers). The authors found that patients were willing to accept longer waiting times for appointments in return for increased relational and informational continuity. Other studies have found that patients value the choice of doctor and appointment time more than how long they have to wait for an appointment, suggesting a re-evaluation of policy makers' focus on timely access (Rubin et al., 2006, Gerard et al., 2008, Cheraghi-Sohi et al., 2008, Hjelmgren and Anell, 2007).

2.1.2 Adherence to and compliance with medical advice

The decision to seek care, however influenced by numerous factors, is the patient's to make. From the moment the patient first visits the GP, however, physicians become involved in decisions affecting the patient's care. As agents for their patients, GPs and specialists advise patients on the best course of action. It is the patient's prerogative to accept or reject this advice. Health care providers acknowledge that patients increasingly expect to be part of decisions involving their own health care (Department of Health, 2013b). Even so, and despite the fact that medical advice can only have an impact if patients decide to follow it,

some have a paternalistic tendency to blame patients for not adhering to, or complying with, care recommendations (Sabaté, 2003). Donovan and Blake (1992) provide an excellent perspective on this:

The term 'compliance' is itself an interesting one. To comply is to obey, submit, defer or accede to instructions. Much of the work looking at compliance (or non-compliance) carries the implication that patients should comply; indeed that they have little option but to do so, and that in an ideal world, non-compliance would not occur. (...) This ideology portrays non-compliance as deviant behaviour and ensures that the blame for it is directed largely towards patients. It is patients who fail to comply, intentionally or unintentionally, because they are ignorant or forgetful. (Page 507)

The paternalistic view that patients should blindly follow physician advice fails to recognise that non-compliance can be intelligent, and justified, in certain cases. For example, in the presence of misdiagnoses, inappropriate prescribing, adverse reactions or side-effects, or when disease conditions have changed, non-compliance can be beneficial (Vermeire et al., 2001).

Regardless of the terms used to describe the phenomenon or divergences over who is to blame, there is general agreement that non-compliance, non-adherence or attrition constitute a serious, and longstanding, problem for health care services and patients. Hippocrates was aware that patients pretended to take medication thousands of years ago (Vermeire et al., 2001). Despite many centuries having passed, today's statistics are dramatic: adherence to long-term therapy for chronic illnesses in developed countries averages 50 percent, and in developing countries rates are even lower (Sabaté, 2003); between 18 and 37 percent of patients, and 1 in 3 children, fail to attend appointments to which they were referred to (Zuckerman et al., 2011b, Forrest et al., 2007); one third of patients with personality disorder do not make it past the first psychotherapy session

(McMurran et al., 2010). The consequences are inefficient therapies with poor health outcomes and increased health costs (Marinker and Shaw, 2003). In the United States (US) alone, non-compliance with medical treatment has been associated with a \$100 billion yearly cost (Vermeire et al., 2001); non-attendance, no-shows or non-completion of referrals costs the UK NHS an estimated £790 million per year (Finkelstein et al., 2013); even though freed slots from no-shows can be filled by walk-in patients, capacity can remain unfilled denying medical care to patients who would have attended (McMurran et al., 2010). Non-completion affects diagnostic procedures, treatments and drugs, and referrals to specialist appointments.

2.1.2.1 Adherence to diagnostic procedures and therapies

The literature on compliance with pharmacological treatments (i.e., drugs) is prolific. The literature on adherence to non-pharmacological therapies is less rich, and studies on completion of diagnostic procedures are almost non-existent. On this last topic, Dafnis et al. (2005) explored factors related to non-completion of colonoscopies using data from the Swedish Public Health Care system. They found that 19 percent of colonoscopies were not done. Non-completion was more likely for potentially more complex procedures normally associated with anatomy and the size of the colon: women, older patients and patients with diverticulosis.

On adherence to, and completion of, non-pharmacological therapies, van Schoor et al. (2002) reviewed studies of patient compliance with external hip protectors, a shell which is sewn or placed in pockets of special underwear, supposed to shunt energy from falls to soft tissues around the hip and the protector itself, this way preventing hip fractures. Between 37 and 72 percent of patients accepted physicians' recommendations to use hip protectors and 20 to 92 percent complied with use (as with other review studies, the definition of compliance may differ across primary studies). Non-compliance was associated with the protector not being comfortable (e.g., too tight or poor fit), the patient needing extra effort and time to put it

on, urinary incontinence, and physical difficulties or illness (e.g., dementia). Kerins et al. (2011) looked into attrition in phase III of cardiac rehabilitation, the aim of which is to restore quality of life and promote lifestyle changes. It involves a structured programme of 6 to 12 weeks of exercise, education and support conducted by a multi-disciplinary team. According to previous studies, non-attendance and non-completion are associated with difficulties travelling to service, having co-morbidities, suffering from anxiety and depression, and lacking motivation. In the study, 11 percent of patients did not adhere to treatment and 19 percent did not complete (i.e., 30 percent of patients offered treatment did not follow recommendations). Non-attendance and non-completion were more likely for manual unskilled workers, smokers, those who were not interested in the programme, those suffering from common illnesses which prevented them from showing up (e.g., colds and stomach upsets), those working, patients who were readmitted to hospital, and those on holidays. Semi-structured interviews with non-completers revealed dissatisfaction with exercise regime (complaints included regime being either too hard or too easy), and barriers to access such as travelling time, poor communication and relationship with staff.

McMurran et al. (2010) explored adherence to treatments for personality disorder. They found non-completion compromises the cost-efficiency of the service, lowers staff morale, and denies other patients a service they could have attended. Non-completion has been associated with younger patients, the less educated, those in lower occupational levels, the less emotionally distressed and those with lower competence in skills necessary for therapy. Similarly, Barrett et al. (2008) investigate attrition from mental health treatment finding that no-shows can lead to lost revenue, low staff morale and high turnover, besides denying access to other patients who would be willing to attend. Attrition was found to be more likely for younger patients, lower socio-economic status, longer distances from service, longer waiting time, social isolation, having more severe and complex diagnoses, lower tolerance of frustration, poorer motivation, poorer relationships with staff, having limited

health insurance coverage, mismatch between expectations of treatment length and actual duration, and confidence in therapist.

Shieh et al. (2006) investigated predictors of non-completion of treatment for latent tuberculosis infection, involving 6 months of isoniazid. Approximately 66 percent of patients self-discontinued treatment with 5.3 percent due to adverse effects. Non-completion was associated with a concern about venipuncture (the process by which intravenous access is obtained) and a low perception of tuberculosis risk. Claxton et al. (2001) explored the impact of dose frequency and regimen complexity on medication compliance by reviewing studies using electronic monitoring. Across the spectrum of therapeutic areas, the more complex the regimen the less likely patients were to comply. Overall compliance rates were 70 to 80 percent. In a review of 30 years of patient adherence research, Vermeire et al. (2001) identified almost 200 doctor-, patient- and encounter-related predictors of non-compliance. Predictors of non-adherence included the number of symptoms, the costs of therapy, the frequency of dosing, the number of medications prescribed, the duration of treatment, the overall complexity, poor communication with prescribing physician, time between taking drug and feeling its effect, fear of adverse effects, and unresolved concerns about diagnosis and presence of symptoms. Compliance was associated with the quality, duration, frequency of, and satisfaction with, interactions with prescribing physician. In the context of diabetes, lower patient adherence was associated with being depressed, having twice-daily regimens compared to once-daily, and polytherapy compared to monotherapy (Cramer, 2004).

According to Griffith (1990), non-compliance is more likely for the younger, older, unemployed, single, widowed and women (especially mothers of high parity). Other predictors include low income, having mild asymptomatic conditions, suffering from chronic illnesses, conditions for which the consequences of non-compliance are delayed, fear of becoming dependent, and drug route of administration, appearance, taste, size, dosing regimen and duration. Once again, the relationship with the prescribing physician with regards to elicitation of, and respect for, patient concerns, as well as provision of information

were predictors of compliance. The quality of the relationship between patient and prescribing physician was also mentioned by Cameron (1996), as well as the duration and complexity of the regimen.

Although many factors associated with adherence and compliance to procedures and therapies are contextual (e.g., concern about venipuncture), there are similarities across different studies. Non-adherence and non-compliance are more likely for more complex therapies and procedures (e.g., regimen and co-morbidities), when relationships with medical staff are poor (e.g., communication with and confidence in physicians), when there is a mismatch between patient expectations and actual treatment or procedure (e.g., duration of treatment), when there are financial and organisational barriers (e.g., patients' socio-economic status and distance to service), when health beliefs conflict with treatment or procedure (e.g., risk of dependence and support from family and friends), and when treatment or procedures are ineffective or dangerous (e.g., side effects and symptom resolution).

2.1.2.2 Appointment completion and attendance

No-shows, non-attendance or non-completion of appointments account for a significant proportion of total scheduled appointments, both in primary and secondary care, throughout the world: between 5 and 55 percent in US, 30 percent in Saudi Arabia, 36 percent in Israel, and 23 to 34 percent in the UK (George and Rubin, 2003, Finkelstein et al., 2013). A survey carried out in the UK revealed 17 million GP appointments were missed in 2000, costing an estimated £150 million. While it is often possible to fill freed slots with walk-in patients, capacity may still remain unfilled, denying access to patients who would have attended (Moore et al., 2001). Furthermore, the problem leads to perverse incentives and flawed reasoning, such as overbooking, making things potentially worse (Schmalzried and Liszak, 2012). Recently, information and communication technologies have been used to remind patients of coming appointments, with consistently good results (Finkelstein et al., 2013).

In a study of referral non-completion, patients reported they had not attended appointments because their health problem had resolved, they lacked the time, they disagreed with the referring physician over the need for a referral, their insurance plan denied them the appointment, or they were referred by a junior physician (Forrest et al., 2007, Chen, 2001). The likelihood of completing a referral was higher for patients who had requested a referral or who had scheduled the appointment at the time of referral. Longer waiting times increased the probability of non-completion, potentially due to symptoms resolving. Zuckerman and colleagues (2011b, 2011a) have studied the determinants of referral non-completion in the context of paediatrics, finding that a higher likelihood of not attending was associated with low necessity for the referral, older children, lack of chronic conditions, having public insurance as opposed to private, longer waiting times, and the appointment having been re-scheduled. Non-completion was also associated with a divergence in opinions between physicians and parents over the need for a referral and the seriousness of the health problem, even though parents were usually more likely to highlight the need for a referral. The research on referral non-completion also highlights communication problems between referring physicians and consultants. Referring physicians are often unaware that patients do not attend appointments and rarely know the outcomes of the specialist appointment (Forrest et al., 2007, Mehrotra et al., 2011).

No-shows in primary and community care have been associated with younger patients, men, patients of lower socio-economic status, the uninsured, divorced, and widowed, having an appointment on Fridays compared to Mondays, forgetfulness, lack of transportation, resolution of symptoms, not having a sense of urgency, lengthy period of time between scheduling and actual appointment, scheduling on short notice, patients with psychological problems, living in deprived areas, being too ill to attend, being admitted to hospital, having a history of previous failed appointments, poor understanding of purpose of appointment, trouble getting off work, road traffic, family and child care, and disagreements with medical and administrative staff (Moore et al., 2001, George and Rubin, 2003, Lacy et al., 2004).

2.1.2.3 Patient preferences

Stated choice methods have also been used to elicit patient preferences for therapies, especially by showing that patients can sometimes value non-clinical benefits above clinical ones. For example, Ryan (1999) explored women's preferences in the context of in-vitro fertilisation and found that good staff attitudes and continuity of care were respectively more important than a 6 and a 2 percent increase in the chance of taking a baby home. Women were also willing to pay considerably for better attitudes from staff, continuity of contact with same staff, follow up support and less time on the waiting list. In a study of women's preferences for osteoporosis drug treatments, de Bekker-Grob et al. (2008) found there were clear preferences for route and type of administration, and tolerance of side effects such as nausea.

Recently, studies have turned to more sophisticated econometric models to account for respondent heterogeneity (i.e., differences in preferences across respondents). Studies of patient preferences for managing asthma and genetic testing, for example, show that some attributes are perceived as positive by a number of respondents and negative by others (King et al., 2007, Hall et al., 2006). The usefulness of these studies for policy makers is still developing. Even so, they illustrate how little we know about individual preferences regarding health care. Clinicians usually choose treatments that maximise clinical benefits, yet stated choice studies suggest patients value non-clinical benefits as well, with implications for adherence to, and compliance with, therapeutic recommendations.

2.2 Determinants of physicians' decisions

An agency relationship exists when a principal delegates decision making authority to another party, the agent. Recognizing that agents have better information, and are generally more competent to make certain decisions, principals delegate part of that responsibility to them. This is essentially what happens when patients seek medical advice from physicians –

they entrust their wellbeing and health to physicians who have been trained to take on that authority (McGuire, 2000). A perfect agent is one who chooses as the principal himself would choose, had she the same information as the agent. In reality, perfect agency is thwarted by asymmetric and imperfect information. Physicians may be unable, or unwilling, to correctly interpret patient preferences, which are not static and are very likely affected by interactions with physicians. Patients may be unable, or unwilling, to appreciate all the implications of physicians' recommendations. To complicate matters further, uncertainty regarding preferences, the presence and severity of disease and the best and most effective course of action, is ever present (Arrow, 1963). Imperfect and asymmetric information not only renders the notion of a perfect agent virtually unattainable, it also allows physicians to act in their own self-interests, in detriment of the patient's. Whether or not physicians have patients' best interests at heart, they play a notable role in shaping demand for health care services, the ones they themselves provide and those provided by other medical staff (Labelle et al., 1994).

In the 1990s, the doctor-patient relationship started to change as physicians moved away from a paternalistic approach towards a partnership with the patient (Vermeire et al., 2001). The paternalistic approach was characterised by four assumptions (Butler et al., 2001): there is a single best treatment; physicians know the best treatments and apply them consistently; physicians are in the best position to evaluate trade-offs between treatments and make choices; and physicians have a legitimate investment in each treatment decision. As Butler et al. (2001) and Zuckerman et al. (2011b) show (to mention just two examples), these assumptions are frequently invalid. The paternalistic approach has thus evolved into shared decision making. In the following sections, I review the literature on how physicians make decisions about their patients' care showing that clinical aspects are just one, albeit important, factor.

2.2.1 Diagnostic investigations and therapy recommendations

In their article concerning the reasoning processes inherent in medical diagnosis, Ledley and Lusted (1959) provide one possible description of the steps involved:

If a physician is asked, "How do you make a medical diagnosis?" his explanation of the process might be as follows. "First, I obtain the case facts from the patient's history, physical examination, and laboratory tests. Second, I evaluate the relative importance of the different signs and symptoms. (...) Third, to make a differential diagnosis I list all the diseases which the specific case can reasonably resemble. Then I exclude one disease after another from the list until it becomes apparent that the case can be fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." (Page 9)

Throughout this iterative process, diagnostic procedures and tests (not only laboratory tests as Ledley and Lusted mention above) can help the physician identify the principal diagnosis and propose a course of action. The latter may involve therapy or referral to a specialist. While physicians' assessments of clinical need generally provide the main rationale for diagnostic investigations and therapies, two patients presenting with the same condition can be given drastically different recommendations (Chandra et al., 2012).

2.2.1.1 Determinants of diagnostic and therapeutic recommendations

In a recent review, Chandra and colleagues (2012) attempt to understand how treatment choices are made and why they are so diverse. They group factors into three categories: demand-side factors including income, price and patient preferences (I have discussed these in the previous section concerning patients' decisions); supply-side factors such as physician income and professionalism; and situational factors such as availability of treatments and/or equipment. Health economists use the notion of utility to understand how

physicians choose between alternative treatments and procedures. Studies have explored numerous determinants of physician utility: income, leisure, consumption of other goods, workload, patients' utility, welfare and well-being, the interest of society, physician reputation, experience and status, resources and capacity constraints, intellectual satisfaction and autonomy, and whether substitutes for physicians' services are available (Scott, 2000, Chandra et al., 2012).

Much work has focused on income and profits, exploring the impact of different payment mechanisms and changes to physicians' income on the choice of treatments and procedures. Very generally, there are two payment mechanisms: capitation and fee-for-service. Under the former, physicians are paid a fixed amount for each patient (per capita, hence the name) under their care. Capitation is a prospective form of payment, so that although the amount paid to physicians is risk-adjusted for patients in a physician's care, it is independent of the actual quantity and intensity of care provided. At the other extreme, fee-for-service retrospectively reimburses physicians for all services provided. Capitation incentivises physicians to cream-skim low cost and low effort patients from their lists, so as to maximise the limited amount they are paid for each patient. Fee-for-service incentivises physicians to provide more services. Krasnik et al. (1990) have illustrated this by showing that a change in payment from capitation to fee-for-service led GPs to provide more services and refer less. In many health care systems today, physicians are paid through a combination of mechanisms, in an effort to ensure incentives are right. Besides pay for performance approaches, other innovations in payment systems in the US include: non-payment of treatment for avoidable or preventable complications and events, episode-based payment, and shared savings with physicians (Rosenthal, 2008). These innovations have a bigger potential to incentivise physicians to act in desired ways but they may also have unintended consequences that reach far beyond the control of policy makers.

As mentioned, a consequence of asymmetrical and imperfect information is the potential for physicians to shift their patients' demand curve in ways that promote their own

self-interest, rather than the patient's. The physician- or supplier-induced demand theory presupposes two things (McGuire, 2000): the physician must be acting in his own self-interest, and his influence must be on the patient's demand rather than on the quantity supplied. These requirements pose considerable challenges for empirical research and make physician-induced demand one of the most contentious topics in health economics. Regardless, there are indications that physicians react to changes in their income by increasing or reducing services. Several studies have found that after Medicare moved from a retrospective fee-for-service payment system for hospital care to a prospective diagnostic-related group payment system, in 1983, there was a large reduction in inpatient days, apparently substituted to the outpatient setting (Chandra et al., 2012). Another study found that after Medicare changed payments for cardiac bypass surgeries, in 1988, physicians responded by increasing supply in both Medicare and private sectors (Yip, 1998).

Physician specialisation and training are also determinants of treatment recommendations. A patient with localised prostate cancer can receive different treatments depending on whether she sees an oncologist or a surgeon (Chandra et al., 2012). The more specialised a physician gets, the more likely he is to get better at a specific treatment or field in detriment of others, in turn making him more likely to choose those treatments at which he excels. Fear of malpractice can also lead physicians to increase or reduce the number of procedures and treatments provided. This is commonly referred to as *defensive medicine*, and is frequently used to explain excessive use of diagnostic procedures (McGuire, 2000). Other explanations however may include the influence of medical education and training, the technological imperative and physician-directed marketing. Medical training emphasises thoroughness rather than effectiveness, with some physicians arguing that at no point should cost factor into care decisions (Hillman and Goldsmith, 2010, Rosenbaum and Lamas, 2012). The technological imperative proposes that once a technology exists (e.g., a new diagnostic procedure) physicians do not feel they can withhold it from patients, and because patients value their life greatly, new procedures get utilised

heavily (Cutler, 1995). Finally, physician-directed marketing provides physicians with incentives to prescribe and recommend treatments and procedures. In the US, pharmaceutical marketing alone (not including medical device companies) is worth more than \$7 billion a year (Emanuel and Fuchs, 2008).

Despite extensive training and education, physicians are susceptible to bounded rationality and cognitive biases. Chandra et al. (2012) identify three biases and misperceptions: the availability heuristic, status quo bias and the confirmation bias. The availability heuristic states that a recent occurrence of a salient event (e.g., having just diagnosed a patient with influenza) is a strong predictor of future behaviour (e.g., diagnosing the next patient that shows up with a cold as influenza). The status quo bias refers to the fact that people tend to persist in their behaviour unless there is a strong reason to change it. Finally, the confirmation bias refers to how people tend to selectively interpret information so that it confirms their beliefs and expectations. These biases and heuristics are all the more important when the cost of collecting new information (e.g., on the patient's condition, on available treatments, etc.) is considered too high. One explanation for the high variability in treatment choices that remains after accounting for the main factors (e.g., income, specialisation, etc.) is an inability to adequately account for these cognitive biases and heuristics (Chandra et al., 2012).

The notion that all physicians are profit-maximising individuals who react to lower incomes by recommending unnecessary – and potentially dangerous – procedures and treatments is inaccurate, not to say cynical. Physicians value their patients' benefit and well-being as well, if not above all else. It is extremely difficult to determine empirically that physicians are acting in their own best interests to the detriment of their patients', and because of asymmetrical and imperfect information and Arrow's uncertainty (Arrow, 1963), even if physicians are self-serving, their actions may still contribute to their patients' best interests. Physician-induced demand, defensive medicine, medical training and specialisation, the technological imperative and even physician-directed marketing, may all

promote utilisation that ultimately benefits the patient. However, I need not speculate as there is ample evidence that physicians are influenced by their patients' interests, expectations and beliefs.

Across the US, Canada and Europe, people are generally more interested in medical innovations than non-medical technologies (Kim et al., 2001). In the US, direct-to-consumer pharmaceutical marketing is associated with increased patient requests for drugs (Emanuel and Fuchs, 2008). In general, people have come to expect more from health care systems and are increasingly willing and able to make their wishes known to physicians (Darzi, 2008). It has been shown that patients' expectations affect physicians' prescribing, and that paediatricians can prescribe antibiotics with no clinical indication when parents pressure them to do so (Belongia and Schwartz, 1998, Butler et al., 2001). Another study found that the strongest predictor of a drug being prescribed was the physician's perception that the patient expected one, even after controlling for clinical presentation (Cockburn and Pit, 1997). A physician's perception of patient expectations was a better predictor than actual patient expectations, with physicians being right only 65 percent of the time.

Finally, physicians' recommendations are dependent on availability and ease of access to health care resources and capabilities. For example, the UK National Institute for Health and Care Excellence (also known as NICE) defines which drugs and treatments are available in the NHS. In many countries, the latest imaging and radiological devices tend to be available only in larger cities. Even when a procedure or treatment is available, barriers to access may limit the physicians' ability to recommend it. Influenced by health insurance coverage, suppliers of medical innovations and providers of health care services can change the structure of their offerings to reflect insurance conditions (Zweifel and Manning, 2000). Even if financial barriers are not important, waiting lists can limit access. In the absence of a market price mechanism, waiting lists act as a rationing device in which clinical need determines how quickly patients exit, conditional on available resources (Siciliani and Hurst, 2005). It has been shown that physicians reduce additions to the list in reaction to longer

waiting times (Goddard and Tavakoli, 1998, Gravelle et al., 2002, Martin and Smith, 1999, van Ackere and Smith, 1999). It has also been suggested that physicians working both in the public and private sectors have a preference for longer lists, since they can direct patients to their private services (Iversen, 1997).

Whereas accessibility may limit physicians' choices, the mere availability of a technology or service has been associated with increased use. Studies have found that treatments and procedures may be used for purposes for which there is no clinical indication, and often no clinical benefit (Ayanian, 2006, Knauf and Aronson, 2009). Others have found that new procedures and treatments intended to substitute older technologies are used as complements (Baker et al., 2010, Cutler and Huckman, 2003). Even though the reasons (e.g., learning by doing, technological imperative, etc.) for these unexpected patterns of use are poorly understood, several researchers now associate technological advances with greater utilisation (Baker et al., 2003, Bosanquet, 2009).

2.2.2 Referrals to specialist care

Besides undertaking diagnostic investigations and proposing treatments, physicians can also refer patients to other doctors, usually specialists who are better able to provide care in a specific area. Research around physicians' decisions to refer has been prolific. Since the 1960s, it has evolved from merely reporting referral rates for different GPs, practices or specialties (Noone et al., 1989) to exploring the reasons behind referral decisions (Newton et al., 1991) and determining whether referrals were appropriate and timely (Fertig et al., 1993). Researchers have identified a comprehensive number of factors thought to influence referrals. O'Donnell (2000) groups them into four categories: patient characteristics, GP characteristics, practice characteristics, and secondary care factors.

2.2.2.1 Patient characteristics

The nature of the presenting problem, and the perceived clinical need for a referral, are, as expected, important predictors of the decision to refer. Various clinical factors have been identified in the literature: burden of morbidity and perceived severity (Mehrotra et al., 2011, O'Donnell, 2000, Sullivan et al., 2005, Newton et al., 1991), diagnostic category and case-specific factors (Forrest et al., 2006, Bertakis et al., 2001, O'Donnell, 2000, Starfield et al., 2002, Wilkin and Smith, 1987), number and burden of co-morbidities (Forrest et al., 2006, Chen et al., 2005, Bertakis et al., 2001), number of medications (Bertakis et al., 2001), number of previous consultations (Carlsen et al., 2008), and general health status (Bertakis et al., 2001). A decision to refer is associated with more severe conditions, higher number and bigger burden of co-morbidities, a larger number of medications, and a poorer general health status. Previous evidence suggests that clinical factors account for the majority of variation in referral decisions, but that, even after adjusting for case-mix and diagnostic categories, variation persists.

Non-clinical determinants identified in the literature include age and sex (Bertakis et al., 2001, Carlsen et al., 2008, Forrest et al., 2006, Newton et al., 1991, O'Donnell, 2000, Sullivan et al., 2005, Wilkin and Smith, 1987), social class (Carlsen et al., 2008, Newton et al., 1991, O'Donnell, 2000, Wilkin and Smith, 1987), race, education, income, marital status (Bertakis et al., 2001) and insurance status (Forrest et al., 2006). Evidence for these determinants is mixed. These studies have associated a decision to refer with older patients, higher Jarman scores (a measure of deprivation), and having health insurance. Findings regarding gender, race, education, income and marital status are inconclusive. Other non-clinical factors include patient expectations, needs, values and preferences (Carlsen et al., 2008, Forrest et al., 2006, Mehrotra et al., 2011, Newton et al., 1991, O'Donnell, 2000, Ridsdale et al., 2007, Webb and Lloyd, 1994), patient anxiety and pressure to be referred (Morgan et al., 2007, Newton et al., 1991, O'Donnell, 2000, Armstrong et al., 1991, Little et al., 2004), and the patient's ability to assert her views and feelings (Newton et al., 1991).

Patient expectations, anxiety and pressure to be referred are important influences, with anxiety and pressure shown to be especially relevant in studies of referral for headache (Ridsdale et al., 2007, Morgan et al., 2007, Zuckerman et al., 2011b).

2.2.2.2 General practitioner characteristics

As with patients, evidence on whether GP age and gender affect referrals is mixed (Carlsen et al., 2008, Forrest et al., 2006, O'Donnell, 2000, Franks et al., 2000, Wilkin and Smith, 1987). A factor which is highly correlated with age is years of experience (Carlsen et al., 2008, Franks et al., 2000, Mehrotra et al., 2011, Newton et al., 1991, O'Donnell, 2000). Other related factors include the type of training, special interests, degree of specialisation and experience with certain conditions or specialties. The impact of experience and specialisation on the likelihood of a referral being made is not clear, with previous studies reporting opposite findings. Membership of professional organisations for GPs does not provide a good predictor of referral decisions (O'Donnell, 2000). Workload and consultation rates have been associated with higher referrals rates, suggesting that GPs 'share the burden' with specialists (Newton et al., 1991, Wilkin and Smith, 1987). Other GP characteristics explored in the literature include risk attitude and aversion, technical and psychosocial orientation of care, fear of malpractice, confidence, and tolerance of uncertainty. There is insufficient evidence that these factors affect referral decisions; only risk aversion has been shown to be statistically significant (Forrest et al., 2006).

It has been shown that individual referral rates exhibit a high year-to-year correlation, giving support to the notion that GPs have a characteristic referral threshold, a stable trait resulting from a combination of the factors in the previous paragraph (Forrest et al., 2006). Because different GPs have distinct combinations, their thresholds, even though relatively stable, are different. Under this view, variation in referral rates is a result of inter-physician variability, and not a product of GP-specific time-varying factors (e.g., experience, specialisation, etc.).

2.2.2.3 Practice characteristics

Research into the impact of practice characteristics on referral decisions has focused on location, size and financing. It has been shown that urban practices have higher referral rates than rural ones, however it is not clear whether the urban/rural categorisation reflects differences in distances or other factors (O'Donnell, 2000, Mehrotra et al., 2011). Longer distances to specialists have been associated with lower referral rates (O'Donnell, 2000, Carlsen et al., 2008). There is mixed evidence concerning the impact of practice size, whether measured by list size, number of partners or population (Mehrotra et al., 2011, Carlsen et al., 2008, O'Donnell, 2000, Wilkin and Smith, 1987). Finally, payment systems and financing schemes, such as capitation and fundholding have also been studied. Capitation was discussed previously. Fundholding was introduced in the UK NHS in 1991 and abolished 7 years later. Practices under the scheme had freedom to negotiate and procure secondary care services directly and keep any surpluses generated (Kay, 2002). As capitation provides GPs with an incentive to increase their patient lists, this in turn is expected to lead to more referrals (Iversen and Lurås, 2000). Evidence on fundholding suggests it has little impact on referral rates (Scott, 2000, O'Donnell, 2000).

2.2.2.4 Secondary care factors

The most important secondary care factor identified in the literature is the supply and availability of specialists (Carlsen et al., 2008, Forrest et al., 2006, Mehrotra et al., 2011, Morgan et al., 2007, Newton et al., 1991, Noone et al., 1989, O'Donnell, 2000, Scott, 2000). This in turn is related to waiting times and lists for specialist appointments, which have both been associated with lower referral rates (Newton et al., 1991, O'Donnell, 2000). Other factors relate to the relationship between GP and specialist: prior interactions and previous experience, whether the specialist is known to return the patient to the GP, quality of communication, and medical skill of specialist as perceived by the GP (Kinchen et al., 2004, Mehrotra et al., 2011). Studies have proposed that increasing communication and

collaboration between GPs and specialists can benefit referrals. For example, telemedicine (the use of video-conferencing equipment to remotely connect GPs, patients and specialists) and joint consultations (appointments in which the specialist and the GP are both physically present) improve the quality and appropriateness of referrals (Vierhout et al., 1995, MacFarlane et al., 2006). There are two reasons for this. First, GPs get direct feedback from specialists on the quality and appropriateness of their referrals. Second, they take part in the consultation eventually learning how to treat some of the more common conditions in primary care, without having to refer patients.

2.3 What can we learn from the literature

The literature on determinants of patients and physicians' care decisions is extensive. While clinical need is an important factor, it is clear that non-clinical factors can also be relevant. Hindered by imperfect and asymmetrical information and uncertainty, decisions are as much a product of tangible factors as they are a result of perceptions and beliefs. The range, nature and interconnectedness of determinants pose significant challenges to empirical studies. As Figure 2-2 shows, the number of factors investigated in the literature is astounding. For example, Vermeire et al. (2001) identified almost 200 doctor-, patient- and encounter-related predictors of non-compliance alone, not including the determinants of attendance and physician recommendations.

Given the extent of previous research, illustrated in Figure 2-2, it seems unrealistic that complex health interventions can be assessed without exploring patients and physicians' behaviours, when it is precisely their actions and choices that determine whether and how a given intervention works. There is a wealth of theories on physician and patient behaviour which is so often locked within black boxes in evaluation studies.

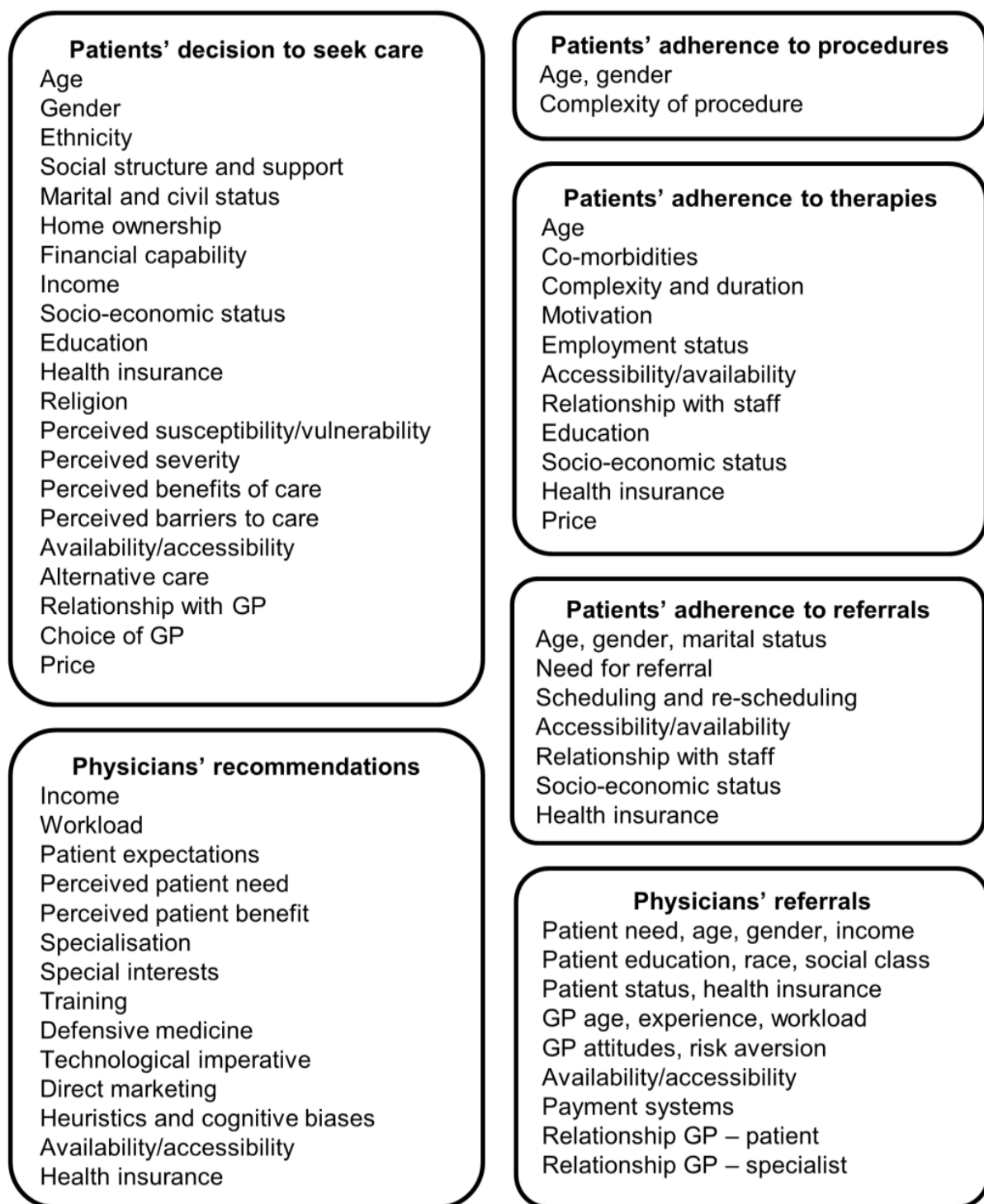


Figure 2-2 Determinants of patients' and physicians' decisions identified in the literature

The list in Figure 2-2 provides just a sample of potentially important factors to take into account when exploring the impact of a specific intervention. Naturally, not every factor in this list will be important in the context of a particular evaluation, in the same way that some factors which could be important may not be in the list. When assessing a particular

intervention, analysts should keep in mind that much has been written on patients and physicians' behaviour in articles that do not directly relate to the impact of that specific intervention. This will be clearer in Part 2 of this thesis, where I explore how a number of factors from this chapter's literature review can inform the evaluation of a specific complex intervention, despite these factors not having been mentioned by studies on the intervention's impact. It is also useful to conduct preliminary and exploratory studies, such as interviews with key stakeholders.

It is important to note there are significant challenges to empirical research in this area. These should not be overlooked as they may bias assessments of whether and how an intervention works. The difficulties are numerous. To begin with, it is undisputed that decisions involving health care are affected by a broad range of factors, and that these interact. For a number of these factors and interactions, operational measures or proxies will be available, but not for all. Even if measures and constructs are available, the potential for confounding is considerable. While recent studies have moved away from univariate analyses of effect, even multivariate approaches can be misleading. In order to adequately control for confounding and alternative explanations, one would need to include all the possible determinants identified previously, while simultaneously guaranteeing that explanatory variables were not measuring the same thing. For example, a dummy variable for patient employment may simultaneously capture an income effect, a social support effect, and perhaps some other unknown confounding effect.

It is also plausible that unemployment will have different effects for the wealthy compared to the less fortunate. It is possible that employment does not really matter; it is actually relative changes in income that are important. Even then, identifying the direction and relative strength of the effect is not trivial. As people value their health, an increase in income may lead to higher health care expenses but it may also lead to better health through better diet, exercise and lifestyle. The net contribution is ambiguous. Contextual factors and individual idiosyncrasies are also important. Developing constructs that adequately capture

specific socio-psychological factors, while minimising the potential for confounding, is far from straightforward. For all these reasons, findings from empirical studies should be interpreted with caution and a good understanding of uncertainties. It is important to be aware that a number of results may be driven by statistical artefacts rather than actual observed behaviours and decision rules.

The fact that certain factors (e.g., income, malpractice, etc.) can have opposing effects may also be an indication that relationships are not linear, nor unidirectional, calling into question the appropriateness of the ubiquitous classical linear regression model. Throughout the literature, researchers acknowledge the presence of feedback and dynamic effects, showing that patients' and physicians' decisions are affected by previous decisions, both directly and indirectly. For example, Siciliani and Hurst (2005) and van Ackere and Smith (1999) talk of feedback effects from waiting times to the quantities of health care demanded and supplied. Cutler and Huckman (2003), and Van de Wetering et al. (2012), caution that the instantaneous effects of a medical intervention can differ substantially from impact at a later date. Schmalzried and Liszak (2012), Butler et al. (2001), and Skinner et al. (2006) all provide examples of well-intentioned actions by patients, physicians and policy makers that have unexpected and unintended outcomes. There is a recognition that feedback and dynamic complexity are present, relevant, and may lead to unexpected and unintended consequences.

It seems that a number of behaviours are not static and probably non-linear. Tools that can explicitly deal with these issues have rarely been applied in the empirical studies reviewed here. A noteworthy exception is provided by articles from the field of *health economics* where more sophisticated econometric methods are frequently employed. Examples include the use of two-stage regressions and instrumental variables to deal with simultaneity (Martin and Smith, 1999), the use of generalized method of moments to deal with temporal correlation (Windmeijer et al., 2005), and the use of individual-specific

regression terms to deal with unobserved heterogeneity and respondent preferences (Hall et al., 2006). Some of these methods are discussed in more detail in Part 2 of this thesis.

Locking patients and physicians' behaviours into black boxes in standard evaluation methods is not inevitable. There are tools that can be used to open these black boxes. In the following chapter, I discuss how standard evaluation methods can provide the basis for a more flexible approach to the evaluation of complex health interventions, one which incorporates, rather than controls for, behaviours and context.

Chapter 3

DEVELOPING AN APPROACH

The literature review shows there is a rich theoretical and empirical understanding of patients' and physicians' decisions. Moreover, it contains numerous instances of dynamic complexity, feedback and counter-intuitive behaviour. The assumption that actions and decisions made by patients and physicians can be put into a black box, assessing solely outcomes, is unrealistic when evaluating complex health interventions. Understanding whether a complex intervention works is as important as how, and why, it works (Byng et al., 2008). It is possible to start opening the black boxes and explore what implications the wealth of studies on decision making has for the evaluation of complex interventions.

Exploring past approaches to evaluation is not only informative; it provides the basis for developing an approach more suited to the assessment of complex interventions. Recently, evaluation of health programmes has come to rely increasingly on economic evaluation, most notably cost-effectiveness, cost-utility and cost-benefit analyses. These types of economic evaluation have gained traction with decision makers because they provide a common framework to compare value for money of different interventions. Their use as priority-setting instruments makes them very appealing to those allocating scarce

resources across a virtually unlimited pool of technologies, services and programmes. Australia and Canada were the first countries to require that manufacturers of health care technologies provide evidence of cost-effectiveness when submitting proposals for funding by health care payers (Claxton et al., 2002). Today, many countries do so from Portugal, Sweden and Finland, to the UK (Sculpher et al., 2004). In this last country, the NHS is legally obligated to fund or provide interventions which have been recommended by technology appraisals of the NICE. A cornerstone of the appraisals is evidence of cost-effectiveness. It is no surprise then that such methods have had an exponential growth in recent years, with cost-effectiveness and cost-utility analyses at the helm. Yet, while economic evaluations of drugs and medical devices have been prolific, their application to complex interventions has been limited.

3.1 Economic evaluation of complex interventions

There are a number of explanations for the scarcity of economic evaluations of complex interventions. To begin with, both cost-effectiveness and cost-utility analyses are based on the extra-welfarist view that the most important benefit of any intervention is improved health (Hauck et al., 2003). The focus is on final health outcomes, with health gains measured as QALYs in cost-effectiveness analysis and natural units in cost-utility analysis. This consequentialist view is often adequate for drugs but it is inappropriate for most complex interventions. While health gains are important, health systems have other objectives: providing a timely and accessible service, promoting equality, valuing patient and provider preferences, and ensuring an efficient use of limited resources (Department of Health, 2013b). Complex interventions often lead to improvements in intermediate outcomes and patient experience, process efficiencies, and benefits to the wider community and economy even. The QALY has been criticised as being narrow since it fails to take into account these wider benefits (Gerard and Mooney, 1993). With complex interventions, focusing on a range of measures may be more informative than identifying a single outcome of interest (Craig et

al., 2008). Economic evaluation methods which are based on classical welfarism allow this, since the focus is on utility rather than solely health outcomes (Hauck et al., 2003). Cost-benefit analysis, for example, uses willingness-to-pay methods (such as contingent valuation or conjoint analysis) to assign a monetary value to outcomes, be it health outcomes which are difficult to value using a QALY, intermediate outcomes, or process efficiencies. However, this approach is fraught with difficulties both in terms of ethical issues to do with valuing health in monetary terms (e.g., willingness-to-pay depends on income and ability to earn it) and in terms of methods to assign those monetary values.

Regardless of whether cost-effectiveness, cost-utility or cost-benefit analysis is used, there are considerations to be made concerning all three. Methodological choices made by the analyst or evaluation team can greatly impact the results of the analysis. Choices pertaining to the scale at which the intervention is being assessed, the perspective used in the evaluation, the time frame and the comparator (i.e., the alternative the intervention is being compared to) can make the difference between a positive result and a negative one. Even with the same scale, perspective, time frame and comparator, results can differ considerably from context to context. Generalizability of recommendations based on economic evaluations is a key issue. Factors affecting the applicability of results to other settings (within and across countries) include epidemiology of disease and demographics; funding, infrastructure and basket of services; remuneration of health care professionals and institutions; relative prices and costs; target populations and subgroups; interactions between interventions and economies of scope; and finally, baseline or initial conditions (Hauck et al., 2003). In the case of complex interventions, to the above we must add the degree of flexibility and tailoring of the implementation, and its delivery in different settings (Craig et al., 2008).

Some of the allure of cost-effectiveness, cost-utility and cost-benefit analyses is their usefulness in helping set priorities across health care interventions, assisting policy and decision makers in resource allocation. Regardless of their different philosophical bases, and

how they go about it, all three approaches allow results from evaluations of different interventions to be compared and priorities set. Hauck et al. (2003) state that if the decision to invest in treatment for people with HIV/AIDS, for example, has already been made, then these approaches can be useful in informing priorities among different treatments, but if the scope is broader, involving whether to choose among treatments for HIV/AIDS or childhood diseases, then methodological issues limit the usefulness of these analyses.

In the context of complex interventions it is arguable whether economic evaluation can be useful even at the level of specific treatment priorities, since interventions can be tailored to specific settings. With complex interventions, what works in one setting may not work in another, making it difficult to establish priorities based on a single evaluation study. Cost-effectiveness, cost-utility and cost-benefit analyses can only be useful as priority setting tools in the context of complex interventions if applied to each setting where the intervention is being used, which may be unfeasible. Another type of economic evaluation may be helpful in these cases: the cost-consequence analysis. In this approach, costs and outcomes are listed separately without aggregating, weighing, or valuing the components (Canadian Agency for Drugs and Technology in Health, 2006). Because it provides a picture of the impact of the intervention, it has been suggested that a cost-consequence analysis be performed for all public health interventions (which intersect with complex health interventions), prior to a valuation of outcomes (Drummond et al., 2007). Because cost-consequence analyses do not explicitly address the issue of priority-setting, they allow any method to aggregate, weigh and value outcomes to be used.

Economic evaluation methods tend to regard randomised controlled trials as the gold standard, the best method to gather evidence of clinical effectiveness (Claxton et al., 2002). Randomised controlled trials are based on the idea that if contextual factors are removed (i.e., controlled for) and patients are randomised to alternative treatment groups, then differences in outcomes can be attributed to differences in treatment received (Blackwood, 2006). A key assumption is that interventions can be standardised, i.e., defined in terms of

essential components which if applied anywhere will produce the same effects. This is hardly the case with complex interventions (MacKenzie et al., 2010). Contextual factors are precisely what make complex interventions work. Controlling for contextual factors in the assessment of complex interventions is tantamount to controlling for the whole intervention. When behaviours are dynamic, learning is present, and tailoring and flexibility are relevant, standardisation is unreasonable (see, for example, the discussion by McCulloch et al. (2002) on the use of randomised controlled trials in surgery). An especially controversial assumption is patient and provider equipoise: the idea that patient and provider do not have a preference for treatment (Walach et al., 2006). When providers so often champion complex interventions and patients expect to be given what they perceive as the best available care, it is hard to make the case for equipoise. In the evaluation of complex health interventions, it is unlikely that a single method will provide comprehensive answers (Walach et al., 2006). Although the use of multiple methodologies and approaches to collecting evidence of effectiveness (clinical and otherwise) can be time-consuming and philosophically challenging, the opportunity to triangulate results should lead to a more comprehensive and robust understanding, emphasising the strengths and mitigating the weaknesses of different approaches (McQueen and Anderson, 1999).

3.2 Building on standard economic evaluation methods

Despite the challenges described above, the growing use of economic evaluation methods has undoubtedly led towards more structured, transparent and systematic decision making in health care. Current guidelines for economic evaluation of technologies, services and programmes are extremely useful and, when applied correctly, can result in complete high quality evaluations. Furthermore, agencies such as NICE and the Canadian Agency for Drugs and Technologies in Health have started work towards ensuring that guidelines are adapted to more complex cases (Canadian Agency for Drugs and Technology in Health, 2006, Drummond et al., 2007, Hill, 2004). Although randomised controlled trials remain the

method of choice in these guidelines, it is now acknowledged that complex interventions require a mix of methodologies and approaches, and the use of both experimental and non-experimental data (Drummond et al., 2007, Canadian Agency for Drugs and Technology in Health, 2006, Walach et al., 2006). There is a growing interest in modelling as a means to make the most of limited data and extrapolate into the future (Craig et al., 2008). Modelling is especially promising with regards to analysing hypothetical scenarios in the early-stages of development and evaluation. It is also useful for identifying sources of uncertainty and potentially important causal mechanisms for which data is sparse, informing future data collection (Canadian Agency for Drugs and Technology in Health, 2006).

Economic evaluation methods are unquestionably the basis for any approach to the evaluation of complex interventions. They provide an invaluable framework on which to build. The work done using cost-consequence analysis is perhaps the most useful with regards to complex interventions. The fact there is growing interest in multi-method approaches and modelling is also encouraging. There is much evidence that goes unnecessarily into black boxes, hindering our ability to understand how a complex intervention works in a given setting and whether it could work in another one. Rather than controlling for context, evaluations should acknowledge it, address it explicitly. There are tools that allow us to start opening those black boxes, and take steps forwards in the evaluation of complex health interventions. In the following sections, I describe some of these tools. I divide these into two groups: those that help us deal with complex systems and those that aid our understanding of individual decision making.

3.2.1 Understanding complex adaptive systems

Very generally, a system can be defined as a group of interacting, interrelated, or interdependent elements forming a collective entity. Determining what makes a system complex seems to be itself a complex task. With one study finding 45 different definitions of complexity, some defining the term in a specific context (e.g., organismic complexity), it is

clear the term means different things, to different people, in different settings (Northrop, 2010). I adopt here Northrop's (2010) description of what makes systems complex. In general: having multiple inputs and multiple outputs, exhibiting non-linear relationships that may cause disproportionate system-wide reactions to small localised changes, and reorganising or adapting over time. Some examples include intracellular metabolic pathways, embryonic development, the human central nervous and immune systems, the human immunodeficiency virus, avian influenza, economic and transportation systems, legislative bodies, and governments (Northrop, 2010). Another characteristic of complex systems is that they are frequently composed of – and interact with – other complex systems. The issue of drug development for neglected diseases is a good example of different complex systems coming together to create a very intricate and complex public health problem requiring action at very different levels, from better understanding of tropical diseases to incentives for private-sector involvement (Trouiller et al., 2002).

In mid-September 2001, Plsek and Greenhalgh (2001) kick-started a series of four papers on the challenges of complexity in health care in the British Medical Journal's *Education and debate* section. The series discussed the implications of greater complexity for clinical practice (Wilson and Holt, 2001), leadership and management (Plsek and Wilson, 2001), and medical education (Fraser and Greenhalgh, 2001). In the first paper, Plsek and Greenhalgh (2001) lay out some of the properties of complex adaptive systems in health care. First, these systems typically have fuzzy boundaries, which can encompass different elements at different points in time. Second, elements react to their environment using dynamic internalised rules and mental models, which may or may not be known by, or logical to, other elements. Third, systems can combine and influence each other, making it difficult to understand one system without exploring others. Fourth, system behaviour emerges from interactions among multiple dispersed and decentralised agents with potentially conflicting objectives. Finally, all the previous properties make it virtually impossible to predict future states of complex adaptive systems.

The literature review on patients and physicians' decisions in the outpatient care pathway provides instances of all of the properties discussed by Plsek and Greenhalgh (2001). The unpredictability and non-linearity of some aspects of outpatient care mean systems are subject to the law of unintended consequences that affects any complex system (Northrop, 2010): actions by patients, physicians and policy makers can lead to unexpected and unintended outcomes, which can be adverse, fortuitous or simply inconsequential. Various examples of this law are available in the literature review:

Overbooking has actually been associated with increased waiting time resulting in ever higher no-show rates. (Page 716, Schmalzried and Liszak (2012))

Pressure of time is frequently cited as a reason for prescribing antibiotics, but in the long term it can lead to more frequent consultations. (Page 438, Butler et al. (2001))

There was little improvement in survival [of heart attacks] after 1996, despite continued growth in costs [of treatment] (...) with regional gains that were (if anything) negatively related to costs. (Page W43, Skinner et al. (2006))

The law of unintended consequences is closely related to our poor ability to deal with complexity (Forrester, 1971). We make sense of the world through both qualitative and quantitative information, the interpretation of which affects and is affected by our mental models (Johnson-Laird, 1983). Mental models emerge out of previously obtained knowledge and observations, and in turn influence the acquisition of new knowledge (Vosniadou and Brewer, 1994). They influence decision rules, organisational strategies, cultural norms and institutional structures (Sterman, 2000). Actual decisions are a consequence of both interpreted information about the world and intrinsic decision rules. Those decisions result in actions that alter the state of the world, providing new information. It is through this iterative process that we learn in and about the world.

Our ability to make the best out of this iterative learning process in complex adaptive systems is hindered in a number of ways (Sterman, 1994, Brehmer, 1987). First, much of the information we use is imperfect, obtained through processes that are susceptible to distortions, delays, biases and errors. Moreover, when analysing data we are generally overconfident, typically failing to generate and consider alternative explanations, and detach our judgements from the way the information is presented. The literature review provides numerous examples of how data collected and analysed through imperfect processes hinders our understanding of patients and physicians' behaviour (see Section 2.3). Unfortunately, while we may be aware of some of these issues, others will not be known to us, and some may even be unknowable. Second, in dynamic and interconnected systems, it is virtually impossible to separate the impact of our decisions and actions from other concurrent changes. This is one of the reasons why randomised controlled trials are considered inappropriate in the context of complex interventions.

Third, we are easily overwhelmed by complexity, resorting to heuristics and rules of thumb such as *educated guesses*, *what we did last time* or *gut feelings* (Oldridge, 1996). These cognitive shortcuts are often based on flawed mental maps that fail to account for feedback, non-linearity and dynamics. For example, physicians may prescribe medications for which there is no clinical indication as a way to deal with time pressures, in the long term increasing the likelihood of adverse events and new demands on the health care system, in turn causing renewed time pressures (Butler et al., 2001). Fourth, even with complete and perfect information, our ability to design strategies and decision rules that have never been tried before, and for which there is no data, is poor. Finally, our decisions may be poorly implemented, again limiting our ability to discern whether outcomes are due to our interventions, or due to poor implementation efforts. For example, the wide scope for variation and tailoring in the way complex interventions are implemented and delivered, limits our ability to attribute results to the intervention rather than concomitant contextual factors (Craig et al., 2008).

Despite all these difficulties, there are ways to improve our ability to learn in and about complex systems. In the same way that we construct mental models of the world, we can develop explicit models that help answer specific questions, and inform our understanding of particular problems (Johnson-Laird, 1983). No model can contain all of reality (if it did, it would be the real world, not a model). Yet, it is precisely a model's imperfection, and its simplified nature and objectiveness, that make it a useful learning tool (Sterman, 2002). By simplifying reality, models allow us to explore a limited number of causal explanations at a time, and identify those mechanisms which are dominant, which most contribute to observed outcomes. As understanding develops, boundaries can be challenged and complexity increased. Furthermore, the modelling process itself can stimulate learning. With complex multidisciplinary problems, the sharing and discussion that go into developing models can help individuals achieve a common understanding of a specific issue (Mathieu et al., 2000, Dodgson et al., 2007a, Dodgson et al., 2007b). In those situations, models act as *boundary objects*: artefacts that promote the sharing of knowledge between differentiated communities of practice (Sapsed and Salter, 2004, Star and Griesemer, 1989, Carlile, 2002). The process of modelling then becomes as important as the end-result itself (Bayer et al., 2010, Zagonel, 2002). In the context of complex health interventions, the model building process can help different groups reach a shared understanding of how an intervention works, and improve communication across organisational boundaries.

Notwithstanding the value of models and the modelling process, flawed models are unlikely to be useful if faults are not identified and understood. Although something can be learned through simply modelling a system, it is through simulation that the assumptions and assertions of models can be tested and challenged, and learning further stimulated.

3.2.1.1 Simulation modelling and microworlds

When models are simulated, they become virtual worlds, or microworlds, where experimentation is possible (Sterman, 2000, Resnick, 1994, Rieber, 1992). Here, I focus on

computational simulations. Virtual worlds have numerous benefits. Even though they can be time-consuming to set up, once they are operational, simulated experiments can be significantly cheaper than real life experiments, especially given the costs of potential disruption and havoc in the real world (Pidd, 1992). Virtual worlds allow time and space to be manipulated, simulating months, years or even centuries in mere seconds. Simulated time can also be halted, allowing those experimenting to pause and reflect. None of this is possible in the real world, where the benefit of running a decade-long experiment is rarely available and probably unfeasible. Virtual worlds allow experiments which can be perfectly replicated. In complex and adaptive social contexts, replication is not possible as systems will have changed in ways the experimenter cannot predict, control or potentially even recognise. In a simulated world, actions can be repeated under the same or different conditions. Virtual worlds are also safe places where experiments have simulated, rather than real, impacts. Dangerous, risky or questionable strategies can be explored without the potential for real world trouble. In simulated worlds, extreme conditions or even disaster can be repeatedly experienced, stimulating learning about rare or one-time events (e.g., anaphylactic shock in surgical anaesthesia). Most importantly, virtual worlds provide immediate high-quality feedback, making them extremely useful playing grounds for decision and policy making.

Good virtual worlds are internally valid, consistent, and well-bounded (Davis et al., 2007). Furthermore, they can make implicit and unconscious behaviours and decision rules, explicit to decision makers (Gilbert and Troitzsch, 2005). They can be tested and their implications explored through simulation. As *Harrison et al. (2007a)* put it:

Traditional approaches to theory development are limited in their ability to analyze multiple interdependent processes operating simultaneously. Even when the individual processes are well understood, analyzing their interdependent behaviour poses difficulties, because the processes involved may interact in complicated and unforeseen ways. And because the interactions typically

produce nonlinear system behaviour with feedback, empirical analysis using the general linear model has limited value, especially when (as is typical) samples are sparse in the regions of greatest interest. (...) we believe simulation (...) has unique advantages in this respect. (Page 1229)

Simulation models are not without fault. Due to their contextual and simplified nature, models may have limited external validity (Davis et al., 2007), especially when they are developed to answer very specific contextual questions. Simulations can require significant amounts of data, some of which will be sparse. Drawing inferences from simulation findings using rare system conditions for which there are limited data, can be counterproductive (Harrison et al., 2007a). The process of setting up the simulation itself may introduce errors that affect the results. Even when simulation is error-free, there are alternative explanations for the same phenomena, essentially different models of the same reality. Different formulations should be explored as they are likely to increase the usefulness of simulation. Presenting and communicating models and simulations is challenging. While models should be parsimonious, even the simplest models will involve numerous assumptions, formulations, parameters, and conditions. Reporting simulated outcomes, as well as learning throughout the modelling and simulation processes, is far from easy, especially when documentation is poor (Harrison et al., 2007a). The biggest limitation of simulation, however, is that it is not immune to all the cognitive and learning impediments discussed in the previous section. Unscientific reasoning and biases as well as defensive routines can undermine the value of virtual worlds.

3.2.2 Understanding individual decision making

The literature review highlighted the breadth of information on patients and physicians' decisions, behaviours and preferences. Because of imperfect and asymmetrical information, and uncertainty, decisions are a product of tangible factors as well as perceptions and beliefs. In the context of complex health interventions, outcomes emerge from these

decisions, behaviours and preferences, and understanding them is crucial to achieve a meaningful assessment of impact.

As illustrated in the literature review, there is a rich body of research from which to build on, when evaluating the impact of complex health interventions. Previous studies can inform model building and provide empirical data for simulations. Even though small models, with limited or no data, can lead to important insights (Ghaffarzadegan et al., 2011), the bulk of simulation modelling tends to require considerable parameterization and substantial amounts of data. These can come from a variety of sources, depending on the system being modelled and the availability of data. Some of the parameters will come from straightforward statistical analysis (e.g., simple counts or averages) while others will require more sophisticated econometric analysis (e.g., instrumental variables regression). The latter tend to be required when exploring determinants of individual behaviour and decision making. Very generally, such studies use either revealed preference data or stated preference data. Revealed preference data are generated in the real world while stated preference data are elicited in hypothetical scenarios. As Louviere et al. (2000) put it, revealed preference data imparts information about the world as it is, while stated preference data shows the world as it could be.

Revealed preference data are useful when exploring preferences for goods and services which are traded in existing markets and technological constraints. While suitable for studying current market equilibrium processes, revealed preference data are limited in important ways. For one, the product of interest may not be traded within a real economic market. Imperfect and asymmetric information, agency, uncertainty, insurance, and externalities all drastically restrict the valuation of health care goods and services using revealed preference data. For example, volunteers in hospitals provide support and companionship to patients, and assist with outings, social programmes and shopping, actively promoting quality of care while reducing staff workload (Handy and Srinivasan, 2004). While approaches to monetizing volunteer hours are debatable they are undeniably

less controversial than methods of valuing quality of care (Handy and Srinivasan, 2004, Brown, 1999). Revealed preference data on quality of care simply do not exist.

Even when market data are available, there may be so little variation in explanatory variables that there is insufficient information regarding decision making outside those limited boundaries. For example, average yearly waiting times for different types of routine surgery in the UK NHS from 1991 to 1998 ranged from 73 to 161 days (Martin and Smith, 2003). Using this data, one can only infer how patients would react to waiting times lower than 73 days or higher than 161 days. Any conclusion will be based on statistical inference rather than actual observation of behaviour. This brings me to the next point: there are no revealed preference data for new goods and services with new attributes and features. Revealed preference data cannot inform decision making outside the boundaries of existing markets and technological frontiers. For example, when planning the introduction of a co-payment for a certain health care service, it would be useful to have an idea of how demand would react to the increase in price. However, if the service was previously free at the point of delivery, then there is no historical data to inform the potential impact of the co-payment on demand. In this example, it is the attribute or feature of the service (i.e., the price) which is new, but in some cases whole new products can be created. In the mid-1980s, patients suffering from depression were treated either with psychotherapy or tricyclic antidepressants (Lieberman, 2003). The introduction of selective serotonin reuptake inhibitors (SSRIs, of which the most famous is arguably Prozac) revolutionized the treatment of depression (Cutler and McClellan, 2001). Naturally, revealed preference data on SSRIs was very limited prior to their use in clinical practice.

Finally, collecting useful and good quality revealed preference data from observational studies, and especially randomised controlled trials, can be time consuming, expensive or even unfeasible. Data on events that occur rarely, for example, provide limited statistical power and usefulness. Notwithstanding these limitations, the use of revealed preference data in empirical studies of decision making far surpasses the use of stated preference data,

as evident in the literature review in Chapter 2. From cross-sectional to panel data, observational and experimental revealed data have been analysed using a multitude of approaches, far too many to discuss in appropriate detail here. Naturally, previous research provides some indication of what techniques are appropriate to answer a specific question. For example, studies of waiting times and waiting lists have used instrumental variables to account for endogeneity of waiting times and referrals (Martin and Smith, 1999, Martin and Smith, 2003, Windmeijer et al., 2005).

3.2.2.1 Stated preference data and hypothetical worlds

Stating that no revealed preference data on SSRIs existed prior to their introduction is a bit like stating the obvious. Yet, SSRIs were not introduced overnight. From as early as the 1970s, there were developments in our understanding of how SSRIs worked and what their effects were (Lieberman, 2003). While there were no revealed preference data on their use, it would still be possible to collect data on stated preferences using the information available at the time. Hall et al. (2002) provide an example, exploring preferences for varicella vaccination in Australia, prior to its introduction. Stated preference data can be used to explore hypothetical situations such as these. Preferences can be elicited for new goods and services with new features and attributes, outside the boundaries of existing markets and technological constraints (Louviere et al., 2000). In the context of complex interventions, this can be extremely useful. Because many of the behaviours and actions involved in complex interventions are being performed for the first time, there is generally a lack of revealed preference data. When revealed data are either sparse or non-existent, stated preference methods can fill the gap. Another important advantage of stated preference data over revealed preference data is the information provided on non-demanders (Ryan and Skåtun, 2004). In revealed data sets of utilisation (e.g., referrals and appointments), non-demanders are effectively absent and no information is available on them. In stated preference methods, it is possible to model non-demanders. When exploring why patients and providers behave in a certain way, information on non-demanders can be very important.

Collecting stated preference data can also be cheaper and less time-consuming than collecting observational data. For decisions that occur only rarely, it may be the only feasible way to collect enough data to model behaviour. Stated preference data are naturally not without faults. First and foremost, the hypothetical nature of the data makes it difficult to determine its external validity. The very purpose of stated preference data is to investigate situations for which there are no revealed preference data, yet the latter is essential for testing external validity. Researchers should keep a *“healthy scepticism about relying on what consumers say they will do compared with observing what they actually do”* (Page 20, Louviere et al. (2000)). A recent review of DCEs (discrete choice experiments, a technique used to collect stated preference data) focused on, among other issues, the validity of responses (de Bekker-Grob et al., 2012). Only one study in 114 tested for external validity by exploring respondent behaviour using revealed preference data. Tests of internal validity are more common, exploring theoretical validity (i.e., whether attributes or features of goods and services have the theoretically expected effect on choices) and determining whether the axioms of consumer theory (e.g., non-satiation, transitivity, reflexivity, etc.) hold. These axioms, put simply, state that consumers are rational (i.e., maximise utility), perfectly informed, and consistent in their choices. However, it has been shown that individuals that failed internal validity tests could have rational reasons for doing so, and that removing them from studies might not be advisable (de Bekker-Grob et al., 2012). Furthermore, what is important is that respondents act in the same way as they state they do, not whether their choices are rational. Other work has focused on making sure that consumers understand hypothetical choice options.

There has been a marked increase in the number of applications of stated preference methods to health care (de Bekker-Grob et al., 2012, Ryan et al., 2008, Ryan and Gerard, 2003). A number of these studies has been mentioned in the literature review (Cheraghi-Sohi et al., 2008, de Bekker-Grob et al., 2008, Gerard and Lattimer, 2005, Hall et al., 2006, Hjelmgren and Anell, 2007, King et al., 2007, Rubin et al., 2006, Turner et al., 2007). Other

studies have looked at providers' preferences, for example regarding treatment choices (Mantovani et al., 2005) and job preferences (Ubach et al., 2003, Lagarde and Blaauw, 2009). While simple descriptive statistics and crosstabulations can be useful, more sophisticated methods are usually used to analyse stated preference data. The literature has largely moved away from conjoint analysis towards DCEs. While the former focuses on respondents' rankings of the products or services under analysis, the latter explores the choices respondents make when comparing a subset of said products or services (Louviere et al., 2010). In terms of econometric techniques, most studies have used random effects binary probit and multinomial logit (de Bekker-Grob et al., 2012). Recently, there have been efforts to model taste or preference heterogeneity using the mixed logit or random parameter logit approaches, which estimate individual parameters for each respondent (see Hall et al. (2006) for a practical example of the use of mixed logit).

In practice, deciding which type of preference data to use will often be determined by availability. Revealed preference data are often the obvious choice when available and of good quality (this is visible in the number of studies using revealed preference data in the literature review compared to stated preference data). However, when revealed preference data are not available, be it because of difficulties in collecting data or because the product or service of interest is simply not available yet, attention turns to stated preference methods. Whenever possible, it is useful to triangulate results from analyses of stated and revealed preference data. There are also methods that integrate the two types of data. The actual mechanics are far from simple (Louviere et al., 2000), and only recently have researchers started to explore this strategy in health care (de Bekker-Grob et al., 2012). There are numerous analytical challenges with only one type of data, much less two. For example, it is now accepted that the way preferences are elicited can itself lead to variability, making it difficult to attribute results to actual tastes rather than data collection (Louviere et al., 2000).

3.3 An approach to evaluating complex interventions

The lack of economic evaluations of complex health interventions is problematic. Given our limited ability to deal with complexity (Sterman, 1994, Brehmer, 1987), without high-quality economic evidence to guide us it is likely that decisions concerning complex interventions will be poor. It is increasingly clear that cost-effectiveness analysis and randomised controlled trials have limited applicability when assessing complex interventions (Drummond et al., 2007, MacKenzie et al., 2010). Agencies which have so often relied on these gold standards of economic evaluation now have guidelines or papers that address methods for assessing more complex interventions, such as public health or community interventions (Canadian Agency for Drugs and Technology in Health, 2006, Hill, 2004). The focus is shifting towards combining the strengths of different methods, rather than relying on a single best method (Walach et al., 2006). Instead of controlling for contextual factors, evaluations of complex interventions should embrace them, clarifying how they affect outcomes. A focus on standardisation should be replaced with an attention to diversity, tailoring and flexibility, all present in complex interventions. Instead of relying on a single measure of outcome, such as clinical effectiveness or utility, evaluation should assess as many outcomes and effects as possible, within and outside the health care sector, whenever possible (Jun et al., 1999, Günal and Pidd, 2010). Evaluations should make room for different scales of implementation, time-frames and choice of comparators, especially when there is uncertainty as to how these choices affect the end-result. Given that generalizability is limited, valuation of outcomes for purposes of priority-setting – such as incremental cost-effectiveness or cost-benefit ratios – should be an accessory rather than a defining step of evaluation.

Evaluations should explicitly address the *complex* in complex health interventions. Behaviours and contextual influences should be explored, as the impact of a complex intervention is a direct result of active involvement from those delivering and receiving the intervention, and knowledge of how stakeholders make decisions is essential to

understanding the outcomes (Pawson et al., 2005). Evaluation is as much about whether the intervention works, as how it works (Byng et al., 2008). To achieve such an understanding, evaluations should explicitly account for the interconnectedness and relationships between components (Craig et al., 2008, Shepperd et al., 2009). Given the likelihood that feedback, delays and non-linearities are important, evaluations should be able to deal with these features, signalling situations where counterintuitive behaviour and unintended consequences are a possibility. Ideally, the intervention should be assessed at different organisational levels. Evaluations that incorporate these features are likely to provide a better understanding of how complex interventions work, and in what conditions.

Incorporating all these features in one evaluation is not only desirable, it is possible. Simulation modelling, informed by previous theoretical and empirical research on decision making and behaviours, can explicitly address complexity in evaluations of complex health interventions. As seen in previous sections, simulation models simplify complexity, helping provide answers to specific questions and problems. Besides their potential as boundary objects (Bayer et al., 2010, Zagonel, 2002), simulation models can lead to the identification of areas which warrant further quantitative work, and on which future data collection should focus. They allow feedback, delays and non-linearities to be modelled explicitly, rather than discussed as an afterthought. They can incorporate a large number of outcomes and effects, which can be assessed in isolation or combined. They can be informed by data from multiple sources and methodologies. The sensitivity of their results to assumptions and parameters can be tested. Simulations can be performed at different scales and time-frames. They can include multiple comparators. They can be tailored to specific contexts, using different formulations, parameters or both. When coupled with stated preference studies, they can generate insights about rare events, and regions where existing data are sparse. And finally, they do not limit the choice of valuation method, allowing the calculation of cost-effectiveness and cost-benefit ratios besides the use of other decision-making aids, such as multi-criteria decision analysis (Santos et al., 2002).

Based on the characteristics of complex health interventions, and the challenges involved in applying standard economic evaluation methods to the evaluation of interventions, programmes and services with those characteristics, I propose an approach to evaluation which builds on an understanding of complex systems, and individual preferences and behaviours. It consists on the use of simulation models, informed by studies of behaviour, to gain an understanding of how an intervention works and what are the likely effects of implementing it elsewhere. The approach uses the techniques previously described to make explicit the relationships between the outcomes of the intervention and the behaviours of those delivering and receiving it. Model building – from initial development, to parameterisation, to validation – is informed by previous theoretical and empirical research into decisions and behaviours deemed relevant to the intervention under analysis. Simulation experiments can then help identify the important causal mechanisms at play, besides providing quantitative impact data. It is these mechanisms, and not necessarily the contextual quantitative findings, which are potentially generalizable. Under this approach, evaluations have both policy and theoretical implications: they inform policy and decision makers of the intervention's impact in a specific setting, but they also build on previous research of how an intervention, or set of interventions, works. Although desirable, generalizability is not guaranteed, as only through the identification of common causal mechanisms, can general insights be attained. The final results mimic those of a cost-consequence analysis, with all outcomes being presented in a disaggregated manner: a comprehensive picture of impact. Different approaches to the valuation of outcomes can then be used, taking into account the preferences and objectives of decision makers.

It is essential to note there is no silver bullet method to the evaluation of complex health interventions. Unlike the evaluation of many drugs, which can be standardised through very explicit guidelines, evaluating complex interventions is precisely about diversity and flexibility. There is no one-size-fits-all reference case (e.g., see Table 2 in Drummond et al. (2007)) setting out the range of methodologies most appropriate for evaluating any given

complex intervention. The external validity of results will always be an issue, and whether findings from one setting can inform decision making in other settings will invariably have to be explored. The suggested approach, while general enough to accommodate the need for flexibility, may still be a useful contribution and a step in the right direction. Given the potential of simulation modelling to produce comprehensive evaluations of complex interventions, it is unfortunate that it has not been used more often. Simulation modelling is still gaining traction with decision makers, who often consider the models too opaque and complicated to understand (Cooper et al., 2006), and some analysts may simply not be aware of simulation modelling techniques (Barton et al., 2004). In the following section, I briefly describe three simulation modelling techniques – discrete event simulation, system dynamics and agent-based modelling – which can be used in the context of complex health interventions, and discuss how to choose between them, when assessing a specific intervention.

3.3.1 Simulation modelling techniques

Approaches to simulation modelling in health care are numerous and include spread sheets, decision trees, Markov models, Monte Carlo simulation, discrete event simulation, system dynamics and agent-based modelling, among others (Cooper et al., 2006, Fone et al., 2003, Brennan et al., 2006, Bayer et al., 2009). Before 1980, simulation modelling was mostly used to address hospital scheduling and queuing, but following advances in the 1990s, applications to a wider range of settings and problems flourished (Fone et al., 2003). Today there is considerable interest in integrating different approaches (Heath et al., 2011). There is a significant history of using models to guide public health policy, especially in vaccination and screening programmes when direct evidence and data are scarce and experimentation is difficult (Weinstein et al., 2001). In economic evaluation and health technology assessments, the majority of studies have employed either decision trees or Markov models (Karnon, 2003, Cooper et al., 2006). The use of other approaches such as discrete event

simulation and system dynamics has been significantly more limited (Karnon, 2003, Cooper et al., 2006).

The question of which approach to use in which situation has been much discussed (Brennan et al., 2006, Barton et al., 2004, Cooper et al., 2006). The choice is bound to be determined by both technical and non-technical factors. Technically, the decision depends on the characteristics of the phenomenon being modelled and the questions being asked. In practice, the decision is also influenced by the modeller's preferences, the technique's ease of use and its general acceptance (Cooper et al., 2006). In the context of complex health interventions, an essential issue relates to the number and complexity of interactions between components of the intervention (Craig et al., 2008). According to Barton et al. (2004) and Brennan et al. (2006), decision trees and Markov models are not the ideal choice in these cases, with system dynamics and discrete event simulation being better equipped. As such, I focus here on system dynamics and discrete event simulation, as well as third more recent approach, agent-based modelling.

Very generally, simulation approaches differ in three ways: how time is handled, whether the system is stochastic or deterministic, and whether change is discrete or continuous (Pidd, 1992). With regard to the flow of simulated time, two options are available: time-slicing or next-event. Because computers only use discrete values, time-slicing attempts to reproduce a continuous flow of time by updating the system every $(t + dt)$, where dt is the chosen time-slice or time step (e.g., an hour, day, week, etc.). On the other hand, next-event techniques update the state of the system only when an event (i.e., state change) occurs. If the system being simulated is deterministic, then its behaviour can be predicted, whereas the behaviour of a stochastic system cannot. It is important to note that deterministic systems may be predictable and yet still be complex (Morecroft and Robinson, 2005). In other words, the outcome may be pre-determined but, due to the cognitive limitations discussed above, we may not be able to determine it at the simulation outset. Finally, the system can change in a discrete or a continuous manner. For example, in a

waiting list, a patient being added to the list would be a discrete change, while changes to the average waiting time for all the patients in the list would be continuous in nature.

Discrete event simulation is a stochastic technique (individuals or entities move through the system according to probability distributions) associated with queuing systems, with state variables changing at discrete points in time (Brailsford et al., 2010). It is a highly flexible technique which allows for sophisticated and complex rules and coding. The modeller can give individual entities different characteristics determining how they move through the activities or services in the model (Brailsford and Hilton, 2001). Due to their stochastic nature, multiple simulation runs are required to obtain statistically significant results (Brailsford et al., 2010). The flexibility and detail usually come at a cost, both in development and in run times (Brailsford and Hilton, 2001).

Discrete event simulation is so pervasive in health care that simulation modelling is sometimes still synonymous with discrete event simulation (Heath et al., 2011). Over the years, there have been numerous reviews of discrete event simulations in health care (Güenal and Pidd, 2010). Jun et al. (1999), for example, show that a large body of work has focused on patient flow (e.g., scheduling and admissions, routing and flow schemes, and the relationship with resource availability) and resource allocation (e.g., number of beds, number of rooms, staff). A review of more recent work has identified applications to the A&E department (perhaps the most popular area for discrete event simulation studies), inpatient care (frequently involving bed occupancy and length of stay), outpatient clinics (often focusing on scheduling and capacity planning), specific hospital units (e.g., intensive care units, laboratories, operating rooms, pharmacies, etc.) and finally whole hospital simulations (Güenal and Pidd, 2010).

System dynamics was established by Jay Forrester in the 1960s, initially in the form of industrial dynamics: the application of feedback concepts to social systems (Forrester, 1961). Its fundamental principle is that the structure of a system (i.e., how its components

are related and affect each other) determines its behaviour over time (Sterman, 2000). System dynamics modelling usually begins with causal loop diagrams, which illustrate the causal influences among variables by connecting them with arrows and identifying the feedback loops present. They are useful in the initial stages of modelling since they elicit and capture mental models and hypotheses about system behaviour (Morecroft, 2007). Simulations use stock and flow diagrams. Stocks are accumulations representing the difference between inflows and outflows. Differences in the rates of flow create delays and provide the system with memory, giving decision makers information which moulds their actions (Lane, 1999). Stocks and flows and feedback loops constitute the pillars of system dynamics thinking.

System dynamics applications reflect the approach's focus on aggregate strategic level processes. Studies have focused on the flow of patients and the use of interdependent resources and units. There have been applications to community and social care (Wolstenholme, 1999, Wolstenholme, 1993, Bayer et al., 2007, Homer et al., 2004), hospital waiting lists (van Ackere and Smith, 1999, Smith and van Ackere, 2002), epidemiology of specific diseases (Roberts and Dangerfield, 1990, Jones et al., 2006), public health and prevention programmes (Homer, 1993, Homer and Hirsch, 2006, Evenden et al., 2005b, Evenden et al., 2005a) and emergency and acute care (Brailsford et al., 2004, Lattimer et al., 2004, Lane and Husemann, 2007, Lane et al., 2000).

Finally, agent-based modelling emerged from the literatures on complex adaptive systems and artificial life (Macal and North, 2010). It has been used to model systems composed of interacting and autonomous agents. An agent is a well-bounded, self-contained, and uniquely identifiable individual (Heath et al., 2011). A typical agent-based model has three components: agents with pre-defined behaviour rules (e.g., adaptive, goal-directed or heterogeneous); relationships and methods of interaction that govern how agents relate to each other in a network; and the environment in which agents are based and interact (Macal and North, 2010). Agent-based modelling is arguably the most technically

demanding approach out of the three paradigms discussed here, potentially involving significant coding. On the upside, it is a highly flexible approach, allowing time-slicing and next-event time handling, stochastic and deterministic behaviour, and discrete and continuous change. In that sense, it stands apart from discrete event simulation and system dynamics. Even though agent-based modelling applications have been limited, there is growing interest in both academia and industry (Siebers et al., 2010). There have been applications to emergency care (Laskowski and Mukhi, 2009, Stainsby et al., 2009), social influences on drinking behaviour (Gorman et al., 2006), epidemiology (Auchincloss and Roux, 2008) and cellular processes (Bailey et al., 2007, Wakeland et al., 2004).

There is growing interest in integrating different simulation paradigms so as to maximise the advantages of each approach while minimising its limitations. The combination of discrete event simulation and system dynamics, for example, has been referred to as a holy grail (Brailsford et al., 2010), yet efforts have extended to integrating agent-based modelling with discrete event simulation and system dynamics as well (Heath et al., 2011, Borshchev and Filippov, 2004, Schieritz and Grobler, 2003, Wakeland et al., 2004). While significant progress has been made in cross-paradigm or hybrid simulation modelling, especially with regards to software, the philosophical issues that need to be tackled to achieve a genuine methodological integration are substantial (e.g., the way different approaches handle time) (Heath et al., 2011). Regardless, cross-paradigm or hybrid simulation modelling is only expected to increase in importance (Heath et al., 2011).

3.3.1.1 Choosing which technique to use

The choice of simulation technique (or techniques since various models can be developed to answer different questions) depends on numerous technical and non-technical factors. Because interaction between individuals and organisational groups is important, decision trees and Markov models are not discussed here (Barton et al., 2004). As mentioned previously, many researchers have sought to compare different simulation modelling

approaches and provide insights into which is more appropriate for different circumstances (Morecroft and Robinson, 2005, Brailsford and Hilton, 2001, Brailsford et al., 2010, Cooper et al., 2006, Heath et al., 2011, Brennan et al., 2006, Barton et al., 2004). Most work has focused on comparing discrete event simulation and system dynamics. The more recent agent-based modelling has received significantly less attention. This may be because of its newness, or because it is philosophically quite different from the other two approaches.

Before I discuss which technique is more appropriate for which situation, it is important to note that creative and inventive uses of all three techniques are possible. While certain criteria can help decide which technique to use, there are probably questions and problems which can be answered using any of the three techniques discussed here. Even so, there are important philosophical differences between these different simulation approaches that allow simple heuristics when choosing between them. To begin, one should determine whether behaviours can be aggregated, or should be handled at the level of the individual (Barton et al., 2004). If the former is true, then system dynamics is likely to be more appropriate. If the latter is true, then both discrete event simulation and agent-based modelling can be used. The second issue has to do with rules of behaviour and movement of agents or entities. When the rules of behaviour are at the level of the system and are independent of interactions between entities and the environment, then discrete event simulation is probably more useful (Heath et al., 2011). If agents have dynamic rules within them, dependent on interactions with other agents and the environment, then agent-based modelling is better. Finally, if variability is important, and the nature of the problem is stochastic, then discrete event simulation and agent-based modelling are more apt than system dynamics (Wakeland et al., 2004, Pidd, 1992). These three criteria – level of aggregation, locus of behaviour rules, and variability – are sometimes enough to determine which technique to use. However, the decision is not always so clear-cut.

Many other criteria have been proposed to help choose a simulation technique. If the scope of the question or problem is more operational than strategic, then discrete event

simulation is likely to be more useful than system dynamics (Brailsford and Hilton, 2001). If important changes to the system happen over time, rather than at specific moments, then system dynamics' time-slicing is better than discrete event simulation's next-event (Heath et al., 2011). If the time-frame of interest is long then system dynamics is probably better (Brailsford and Hilton, 2001). If the number of entities or individuals is large, then system dynamics, or agent-based modelling, are likely to be more appropriate than discrete event simulation. While all three techniques can deal with feedback, system dynamics is the only one which addresses it explicitly, with other approaches *hiding* it within the models' logic (Morecroft and Robinson, 2005). Finally, system dynamics is more apt for gaining understanding, and for long-term policy making and planning, while discrete event simulation is better for short-term optimisation and prediction (Bayer et al., 2010, Brailsford and Hilton, 2001).

The most important difference between system dynamics and discrete event simulation is each technique's assumptions concerning system behaviour. The former sees feedback structure as the main source of behaviour, while for the latter it is randomness that leads to behaviour (Morecroft and Robinson, 2005). Agent-based modelling stands somewhere in between (Macal and North, 2010). As mentioned, creative uses of all three approaches are possible. While feedback is not explicit in discrete event simulation it can be modelled. Conversely, even though randomness is not a big part of system dynamics, it can be incorporated (e.g., using lookup functions). Different techniques can be used to model the same phenomenon (see, for example, Wakeland et al. (2004) models of cellular receptor dynamics). Techniques can also be combined, using two or more models to inform each other (Heath et al., 2011). Naturally, apart from all the more technical criteria above, the choice of simulation approach also depends on ease of use, the speed of model development, the process of validating results, and general acceptance from decision makers (Cooper et al., 2006). Table 3-1 provides support in choosing a simulation technique, summarising much of what was discussed in the previous paragraphs.

Table 3-1 Choosing the right simulation technique

	Yes	No
<i>Can behaviour be aggregated?</i>	SD	DES & ABM
<i>Are dynamic rules of behaviour within individual?</i>	ABM	DES
<i>Is variability/randomness important?</i>	DES & ABM	SD
<i>Do changes to the system happen over time?</i>	SD & ABM	DES & ABM
<i>Is the time-frame of interest long?</i>	SD	DES & ABM
<i>Is the number of entities large?</i>	SD & ABM	DES
<i>Is feedback important?</i>	SD	DES & ABM
<i>Is the scope more operational rather than strategic?</i>	DES & ABM	SD
<i>Is understanding more important than optimisation?</i>	SD	DES

SD – system dynamics; DES – discrete event simulations; ABM – agent-based modelling

3.3.2 Applying the approach in practice

Figure 3-1 summarises the discussions in this chapter. Certain health interventions, programmes and technologies have multiple dimensions of complexity, with implications for how we evaluate their impact. Standard economic evaluation methods are limited in their ability to handle complexity and provide insights into how complex interventions work. And that is the overarching requirement of any evaluation of a complex intervention: it should be about how and why it works, not just whether it works. The approach suggested here brings together tools which can explicitly deal with complexity, answering how and why questions.

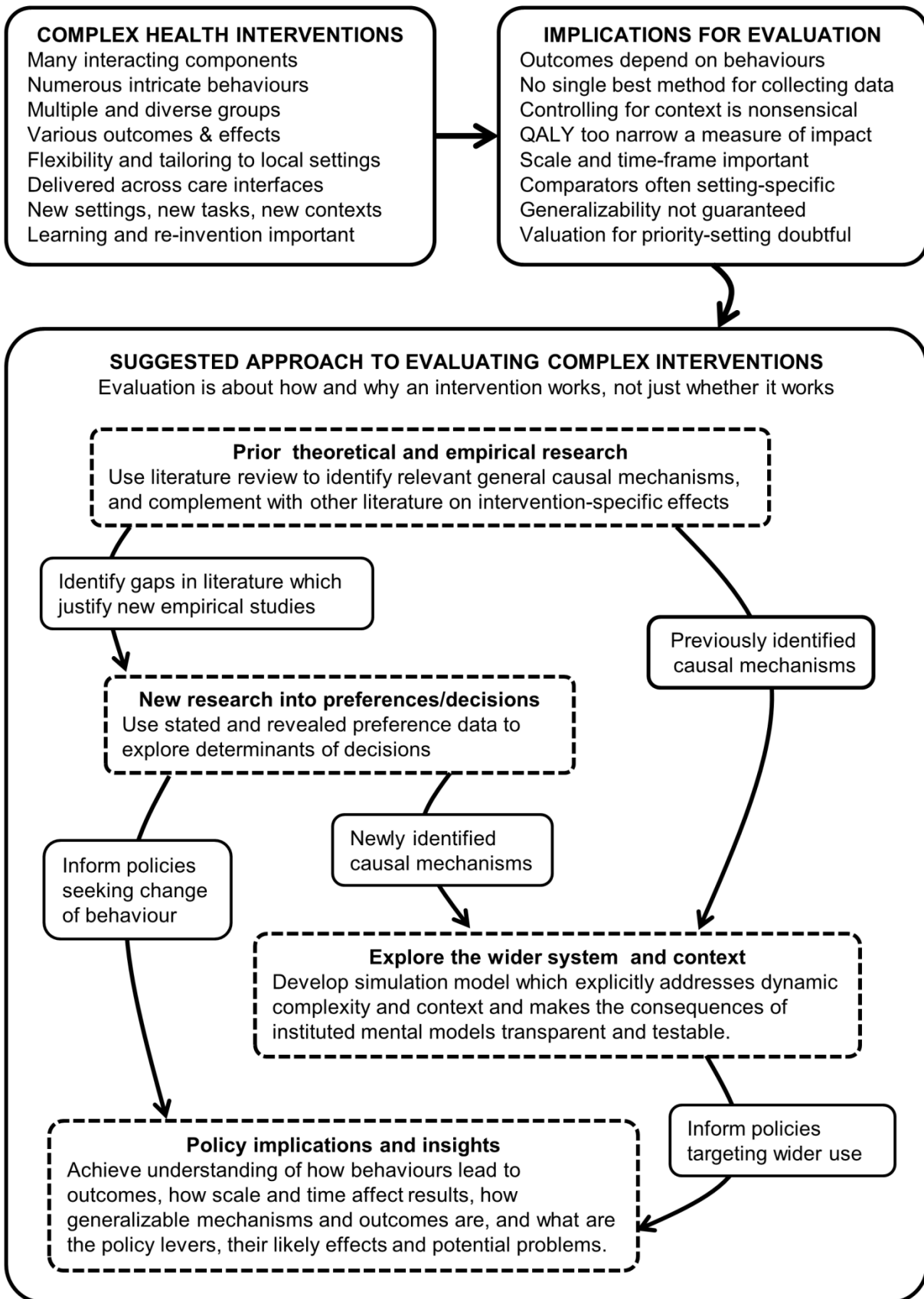


Figure 3-1 Suggested approach to evaluating complex health interventions

To assess the impact of a complex intervention, one must understand how its costs and consequences emerge from the decisions and behaviours of those delivering and receiving it. In other words, the causal mechanisms behind the observed outcomes must be made explicit, and their system-wide impact explored. As with other evaluation methods, one should begin by reviewing previous evidence on the impact of the intervention. The literature review in Chapter 2, notably the list in Figure 2-2, provides a starting point for identifying potentially important factors and causal mechanisms. However, as previously discussed, a number of these factors will not be important in evaluating specific interventions, while some factors that are not in the review may be important.

Naturally, it must be complemented with a review of the evidence on the intervention being assessed. Reviewing previous knowledge is important as it provides a picture of instituted mental models about how and why the intervention works, and often numerous statements about whether it would work in other settings. For the reasons discussed in Section 3.2.1, the lack of a holistic systems perspective greatly limits the value of these statements, which frequently go untested.

Reviewing prior studies also leads to an identification of the gaps in empirical and theoretical research. These can be gaps in the literature about behavioural mechanisms and preferences, which have implications for a wider range of interventions, or they can be gaps in our understanding of intervention-specific mechanisms. In either case, new studies can be undertaken. Data can be collected on revealed or stated preferences, depending on whether the former is available, and whether the latter can provide insights into behaviour and decision making. Numerous econometric and statistical methods can then be used depending on the characteristics of the data and the questions being asked. While some of the new studies will only inform the evaluation of the specific intervention, others may lead to wider contributions to existing knowledge. Some of the insights produced by undertaking these studies may also be generalizable to evaluations of other complex interventions.

The causal mechanisms identified in the literature review, and in the new studies of individual preferences, can then inform the development of a simulation model which takes into account the wider system and context in which individual actions take place. Such a holistic systems perspective is expected to enrich individual-level studies, by explicitly acknowledging that those delivering and receiving the intervention operate in a broader context and affect, and are affected by, the system around them. Simulation modelling allows previous assertions of how and why interventions work, and whether they would work in other settings, to be tested. It demonstrates clearly the consequences of existing mental models. By explicitly addressing dynamic complexity and context, simulation models allow causal mechanisms identified previously to be incorporated in the evaluation, rather than controlled for, or simply contained in black boxes.

Figure 3-1 should not be seen as a linear unidirectional approach, from the literature review down to the policy implications. New studies of behaviour and decision making can re-focus the scope of the evaluation, requiring new literatures to be explored. Simulations and sensitivity analyses can inform whether mechanisms identified in qualitative studies are quantitatively important. Some mechanisms may only be identified at the end of the evaluation process, requiring a re-iteration. All this will be clearer in Part 2 of this thesis, when the approach suggested here is applied in practice.

The end-product is a greater understanding of how and why an intervention works; what are the dominant causal mechanisms, preferences and behaviours at play; how contextual differences, scales and time-frames affect the outcomes; and what are the likely effects of implementing the intervention elsewhere. The reasons behind counter-intuitive behaviours and unexpected outcomes should also be clearer. Furthermore, it is expected that insights from the simulation modelling exercise, and the studies of preferences, contribute to previous theoretical and empirical research, informing future work.

In the remaining part of this thesis, I illustrate how the approach proposed here can be applied in practice to the evaluation of a complex health intervention, in a specific setting, clarifying some of the discussions and statements contained in this chapter.

Part 2

APPLYING THE APPROACH IN PRACTICE

Complex health interventions typically exhibit many interdependent components, a strong behavioural dimension, extensive tailoring to local settings, and numerous and diverse outcomes within and beyond the health care sector (Craig et al., 2008, Shepperd et al., 2009). Under these conditions, the gold standards of economic evaluation – cost-effectiveness analysis and randomised controlled trials – are very difficult to apply, perhaps even impossible. An approach to evaluation which can explicitly deal with complexity was proposed in the previous chapter. In Part 2 of this thesis, I use a case-study of a complex intervention in a specific setting to illustrate how the suggested approach can be applied in practice.

The case-study is the use of real-time teleconsultation in Alentejo, a region in Portugal. Real-time teleconsultations are outpatient appointments using video-conferencing equipment to connect a consultant, in a hospital, with a GP, and a patient, in a primary care practice. In Chapter 4, *The complex health intervention*, I describe the intervention, review the evidence on its impacts, identify the main gaps in the literature, and discuss previous evaluation efforts. In Chapter 5, *Teleconsultation in Alentejo*, I introduce the setting and describe why

the region started using teleconsultation. I then present some initial indications of impact, discussing these in the wider context of the literature on teleconsultation.

Chapters 4 and 5 provide the basis for the subsequent studies undertaken in this thesis, namely the studies of preferences and decision making in Chapters 6 and 7, and the simulation model in Chapter 8. In Chapter 6, *Waiting times and utilisation*, I use revealed preference data to explore how changes in waiting times lead to changes in referral rates and non-completion of secondary care appointments. In Chapter 7, *The determinants of general practitioners' referrals*, I take a different approach to the study of GPs' referral decisions by using stated preference data. In Chapter 8, *The simulation model*, I explore what implications the causal mechanisms identified in previous chapters have for a holistic system-oriented evaluation of teleconsultation in Alentejo. I conclude by discussing how the suggested approach contributes to a greater understanding of how, and why, real-time teleconsultations work, and whether they could work outside Alentejo.

Chapter 4

THE COMPLEX HEALTH INTERVENTION

Real-time teleconsultations (henceforth simply teleconsultations) are outpatient appointments using video-conferencing equipment to connect a consultant, in a hospital, with a GP, and a patient, in a primary care practice. They are included in a larger group of interventions, technologies and services which all begin with the prefix *tele*. Telemedicine, telemonitoring, telecare, telehealth, teleconsultation, and others (the list is extensive) all involve the use of information and communication technologies to deliver health care at a distance. Many of these terms are used interchangeably in scientific papers, reports, practice, etc. For example, real-time video-consultations in dermatology can be referred to as telemedicine, teleconsultation or teledermatology. Moreover, all the previous terms can be used to describe store-and-forward consultations in dermatology, which are markedly different from real-time appointments in that they do not actually involve live interaction between specialist, GP and patient. Store-and-forward dermatology teleconsultations and real-time dermatology teleconsultations are very different interventions. Even real-time teleconsultations may differ as, for example, the patient may or may not be present.

The fact that the same term can mean different things to different people is common with complex health interventions and a symptom of bigger problems than mere terminology. A parallel can be drawn between teleconsultation and MacKenzie et al. (2010) discussion of Keep Well, an anticipatory care approach to cardiovascular care launched by the Scottish government in 2006, and pioneered in the Netherlands and Wales. To begin with, stakeholders had different understandings of what the intervention was, what were its defining components, and at what level it should be managed. These differences of opinion led to diverse practical approaches in different areas of Scotland. Pilot schemes used diverse approaches to the implementation, delivery and monitoring of the intervention. What was supposed to be one intervention had become a number of heterogeneous interventions, limiting comparability. As MacKenzie et al. (2010) put it, *“crucially, evaluators have little or no control over these types of policy refinements”* (Page 401, MacKenzie et al. (2010)).

Teleconsultations have been introduced in numerous countries and regions throughout the world, from Republic of Ireland (MacFarlane et al., 2006) to Canada (Masino et al., 2010), the UK (Gilmour et al., 1998), Scotland (Dorrian et al., 2009), Italy (Massone et al., 2006), India (Solberg, 2008), Brazil (Alkmim et al., 2012), states of Kentucky (Shannon and Buker, 2010) and Ohio (Uhlenhake et al., 2009) in the US and Australia (Lim et al., 2001), to name a few. In all these places, and even within them, teleconsultations have been developed, implemented, managed and monitored in different ways.

In the following section, I broadly review the evidence on the effects of teleconsultation, building on the rich body of literature that has developed over the years in all these places. Even though the primary interest is real-time teleconsultations, I explore a wide range of studies, including different definitions, components, approaches and perspectives. The objective is not to provide a definitive summary of all the evidence on teleconsultations (given what was said in the previous paragraphs, it is questionable whether that is even possible). Rather, the purpose of the next section is to capture the essence of a diverse and rich literature marked by significant heterogeneity, but in which commonalities

can still be found. As with the literature review in Chapter 2, I used a narrative approach here as well. I conducted a preliminary search of the two leading journals in the field (Journal of Telemedicine and Telecare, and Telemedicine and e-Health, mentioned below) as well as PubMed (Medical subject heading – MeSH – terms were used, such as *remote consultation* or *teleconsultation*). The preliminary searches were used to identify key articles (highly cited) and complemented using a snowballing approach.

4.1 The evidence on real-time teleconsultations

In 1995, the first issues of two journals were coming out: the Journal of Telemedicine and Telecare and the Telemedicine Journal (today known as Telemedicine and e-Health). Since then, interest in telemedicine and other *tele* interventions grew steadily, motivated particularly by the promise of such interventions to increase access and reduce costs. With growing interest came increasing numbers of applications, pilot studies and data. Using the keyword *telemedicine* in PubMed returns 355 papers published in 1995, rising to more than three times that number in 2013 alone. In an article examining the occurrence of the terms *telemedicine*, *telehealth* and *e-health*, Fatehi and Wootton (2012) found close to 12,000 papers from as early as 1972.

Methodological approaches evolved from feasibility studies and anecdotal accounts to the use of randomised controlled trials. The number and diversity of effects investigated in the literature multiplied. Studies focused on specific specialties (e.g., cardiology), conditions (e.g., diabetes) and effects (e.g., patient satisfaction). The field was so prolific that other journals, from health services research, to health policy, to health informatics, started to publish articles on *tele* interventions. Systematic reviews of systematic reviews were published. May (2006) provides a good description of this evolution:

For more than thirty years, telemedicine systems have been advocated as a means to secure rapid and responsive access to health care for populations

that are under-served by specialist services because of structural or spatial inequalities in service provision [70,71]. During this period, a large body of literature has grown up describing experimental services and demonstration projects and their evaluation. Recent systematic reviews have emphasized the clinical effectiveness and advantages of telemedicine systems [72-81]. However, other reviews have questioned the response of service users and the cost effectiveness of services, and have pointed to the methodological poverty of much work in these areas [80]. Despite significant support from clinicians, health service managers and policy-makers in many countries, telemedicine services seem to have failed to penetrate wider patterns of service provision.²

As May (2006) points out, research is increasingly focusing on evidence of cost-effectiveness and the methodological quality of previous studies, since both of these areas have been found lacking. It is interesting that despite the scarcity of high-quality methodological approaches to the evaluation of teleconsultation (and most *tele* interventions for that matter) there is unwavering support from those using them. In the following paragraphs, I provide an overview of the evidence on teleconsultation, finishing with a brief discussion of the strengths and limitations of evaluation and evidence of economic impact.

4.1.1 A comprehensive picture of impact

There are many different approaches to the implementation of teleconsultations. In store-and-forward teleconsultation, the GP sends the details of a patient's complaint electronically to the specialist. The specialist reviews the information and proceeds by asking for: more information, a real-time teleconsultation, a face-to-face appointment, or care to be provided in primary care. When only real-time teleconsultations are available, the specialist can ask for a face-to-face appointment or make treatment recommendations. In some cases, the

² References "[70,71] (...) [72-81] (...) [82]" are cited in MAY, C. 2006. A rational model for assessing and evaluating complex interventions in health care. *BMC Health Serv Res*, 6, 86.; they are not cited in this thesis. For more information on these references, please refer to *ibid.*.

patient can choose between a teleconsultation or a face-to-face appointment. These different approaches to implementation can lead to different outcomes. The literature on teleconsultation has explored a multitude of topics: the impact of the intervention on utilisation (e.g., number of appointments), the duration of the consultation, clinical outcomes and effectiveness, access as measured through waiting times, distances and time-prices, patient and provider satisfaction, opinions and views, the continued education of GPs, the carbon footprint and greenhouse gas emissions, and the costs and savings. Many studies have explored how teleconsultation affects the number and type of appointments provided. Teleconsultations are not a perfect substitute for face-to-face appointments, rather they are complements. In other words, there will always be a need for face-to-face appointments (e.g., when a physical examination is required).

Movement between the face-to-face pathway and the teleconsultation pathway leads to unpredictable patterns of use, and potentially very different outcomes and costs. There is also the potential for inefficient use of resources (e.g., conducting a teleconsultation and a face-to-face appointment, when the latter would have been sufficient), although studies of the appropriateness of subsequent face-to-face appointments following a teleconsultation are virtually non-existent. Evidence on store-and-forward teleconsultations suggests they may be a good triage tool, with between 18 to 75 percent of patients avoiding referrals to secondary care (Knol et al., 2006, Martínez-García et al., 2007, Moreno-Ramirez et al., 2005, Eminovic et al., 2009, Eminović et al., 2003, Shapiro et al., 2004, Whited et al., 2002, Griffiths, 2010). After real-time teleconsultations, between 3 and 56 percent of patients have a subsequent face-to-face appointment (Edison et al., 2011, Wootton et al., 2000, Loane et al., 2000a, Loane et al., 2000b, Wallace et al., 2002, Lamminen et al., 2001, Nordal et al., 2001, Granlund et al., 2003, Burgiss et al., 1997, Oakley et al., 2000, Whited, 2006, Scalvini et al., 2009, Baldwin et al., 2003, Leggett et al., 2001). Interestingly, the rates of subsequent appointments following face-to-face appointments are not much different from those for teleconsultations (Wallace et al., 2002, Loane et al., 2000b, Oakley et al., 2000). A recurrent

assertion is that a number of patients could be managed in primary care, without having to be referred to specialist appointments, and that teleconsultation is a means to achieve that (Granlund et al., 2003, Burgiss et al., 1997, Jarvis-Selinger et al., 2008).

Another way in which teleconsultations affect utilisation patterns is by providing care where previously there was none (Hjelm, 2005). In a number of remote locations, there is actually no alternative to teleconsultation as face-to-face care is non-existent (Nordal et al., 2001). Given a baseline of no service, it is no wonder that utilisation increases (Cameron et al., 1998, Martínez-García et al., 2007). Studies have reported that 26 to 30 percent of patients would not attend a face-to-face appointment if teleconsultation was not available (Stensland et al., 1999, Hicks et al., 2003). The literature shows unquestionably that the use of teleconsultation affects the number of outpatient appointments provided. Because changes in utilisation can have an impact on the costs associated with an intervention (Cutler and McClellan, 2001), it is important to understand how teleconsultations affect utilisation and what are the implications for evaluations of economic impact.

A number of studies have explored the impact of teleconsultations on continued medical education and learning by GPs. Although both store-and-forward (Moreno-Ramirez et al., 2005) and real-time (Whited, 2006) teleconsultations have been associated with increased learning, the understanding is that due to the active participation of GPs in real-time teleconsultations, they are less isolated from peers (Nordal et al., 2001) and better able to identify patients who need specialist care and those that can be managed in primary care (Shapiro et al., 2004). Teleconsultations effectively lead to a knowledge transfer from specialists to GPs (Jarvis-Selinger et al., 2008, Loble, 1997). Studies have suggested reductions in referrals in the order of 10 to 25 percent (Wootton et al., 2000, Loane et al., 2001, Taylor, 2005) and in subsequent appointments (Lamminen et al., 2001).

A big part of the literature has focused on patient and provider satisfaction with teleconsultation. Studies of patient satisfaction tend to find high satisfaction, mostly

associated with shorter waiting times, distances and lower costs (Moreno-Ramirez et al., 2005, Hicks et al., 2003, Kvedar et al., 1999, Burgiss et al., 1997, Whited, 2006, Jaatinen et al., 2002, Yogesan et al., 2006, Edwards and Patel, 2003, Deshpande, 2008, Taylor, 2005, Mielonen et al., 2000, Ekeland et al., 2010). Although some patients would rather be seen face-to-face, they would prefer a teleconsultation if waiting time was lower (Collins et al., 2004). Provider satisfaction is also generally high, especially for GPs, although some mention concerns about increased workload (Edwards and Patel, 2003, Demartines et al., 2000, Mielonen et al., 2000, Nordal et al., 2001, Kvedar et al., 1999, Whited, 2006).

The evidence on access measures such as waiting time and distance is overwhelmingly in favour of teleconsultation. There are numerous examples of reductions in waiting times for specialist outpatient appointments following the introduction of teleconsultation, both store-and-forward and real-time (Shapiro et al., 2004, Whited et al., 2002, Krupinski et al., 2002, Nordal et al., 2001, León et al., 2011, Yogesan et al., 2006). Why waiting times for teleconsultations are shorter than for face-to-face appointments is not clear. The duration of the appointments might provide an explanation: if teleconsultations are faster than face-to-face appointments then capacity is effectively increased, *ceteris paribus*. However, findings are mixed. Some studies find teleconsultations are longer than face-to-face appointments (Wootton et al., 2000), some find no difference (Nordal et al., 2001), and some find they are shorter (Oakley et al., 2000, García Rada, 2011, León et al., 2011, Baldwin et al., 2003). In contrast, there is clear evidence that distances travelled are significantly reduced (Oakley et al., 2000, Jaatinen et al., 2002, Sicotte et al., 2004, Dorrian et al., 2009, Masino et al., 2010, Lewis et al., 2009, Merrell and Doarn, 2009). As a consequence, teleconsultations have been associated with considerable reductions in greenhouse gas emissions (Masino et al., 2010, Merrell and Doarn, 2009, Dorrian et al., 2009, Lewis et al., 2009).

Much effort has gone into understanding the impact of teleconsultation on clinical outcomes. Many studies have focused on diagnostic agreement, i.e., whether diagnoses

made in teleconsultations are equivalent to diagnoses made in face-to-face appointments. Diagnostic agreement is generally around 80 percent (Edison et al., 2008, Kvedar et al., 1999), with one study finding a 97 percent rate (Leggett et al., 2001) and another 51 percent (Eminović et al., 2003). In a recent review of the effectiveness of various teleconsultation interventions, there was mixed evidence (Ekeland et al., 2010). This may be due to the fact that effectiveness depends on the specific specialty or condition under study. For example, Hersh et al. (2002) found the best results were for psychiatry and dermatology. A growing number of studies report there is no difference in clinical outcomes between teleconsultations and face-to-face appointments (García Rada, 2011, Deshpande, 2008, Taylor, 2005, Wade et al., 2010, Armstrong et al., 2007). Some researchers have also sought to determine whether patients that attend teleconsultations have similar complaints to those attending face-to-face appointments. While one study found no difference (Krupinski et al., 2002), two other studies suggest teleconsultation patients are healthier and tend to present with less urgent conditions (Mofid et al., 2007, Taylor et al., 2001). The latter studies did not explore why these differences occurred.

Finally, researchers have studied whether teleconsultations are cheaper than face-to-face appointments. Older studies tend to associate teleconsultation with higher costs (Loane et al., 2001, Whited et al., 2003, Whited, 2006, Stensland et al., 1999, Sicotte et al., 2004), while more recent studies report lower costs (Eminović et al., 2010, Armstrong et al., 2007, Griffiths, 2010, Deshpande, 2008). Since initial investment in equipment is an important cost (Wootton et al., 2000, Lamminen et al., 2001, Jarvis-Selinger et al., 2008), the positive findings of more recent studies may simply reflect changes in prices of technology (Doolittle et al., 2004). Another explanation for the different findings is that costs and savings differ depending on the perspective taken. A number of studies propose that whilst teleconsultation may increase costs to providers, for patients and families there are significant savings (Cameron et al., 1998, Sicotte et al., 2004, Jennett et al., 2003, Wade et al., 2010, Whited et al., 2003). In fact, some have suggested that teleconsultation effectively

shifts costs from patients and families (e.g., travelling) to providers (e.g., equipment) (Hakansson and Gavelin, 2000, Sicotte et al., 2004). Another aspect that is less frequently mentioned is the fact that GPs take part in teleconsultations, which should have an impact on their workload and productivity, in turn potentially leading to higher costs (Wallace et al., 2002, Wootton et al., 2000). There is a wealth of research into the impact of teleconsultation; the main findings are summarised in Figure 4-1.

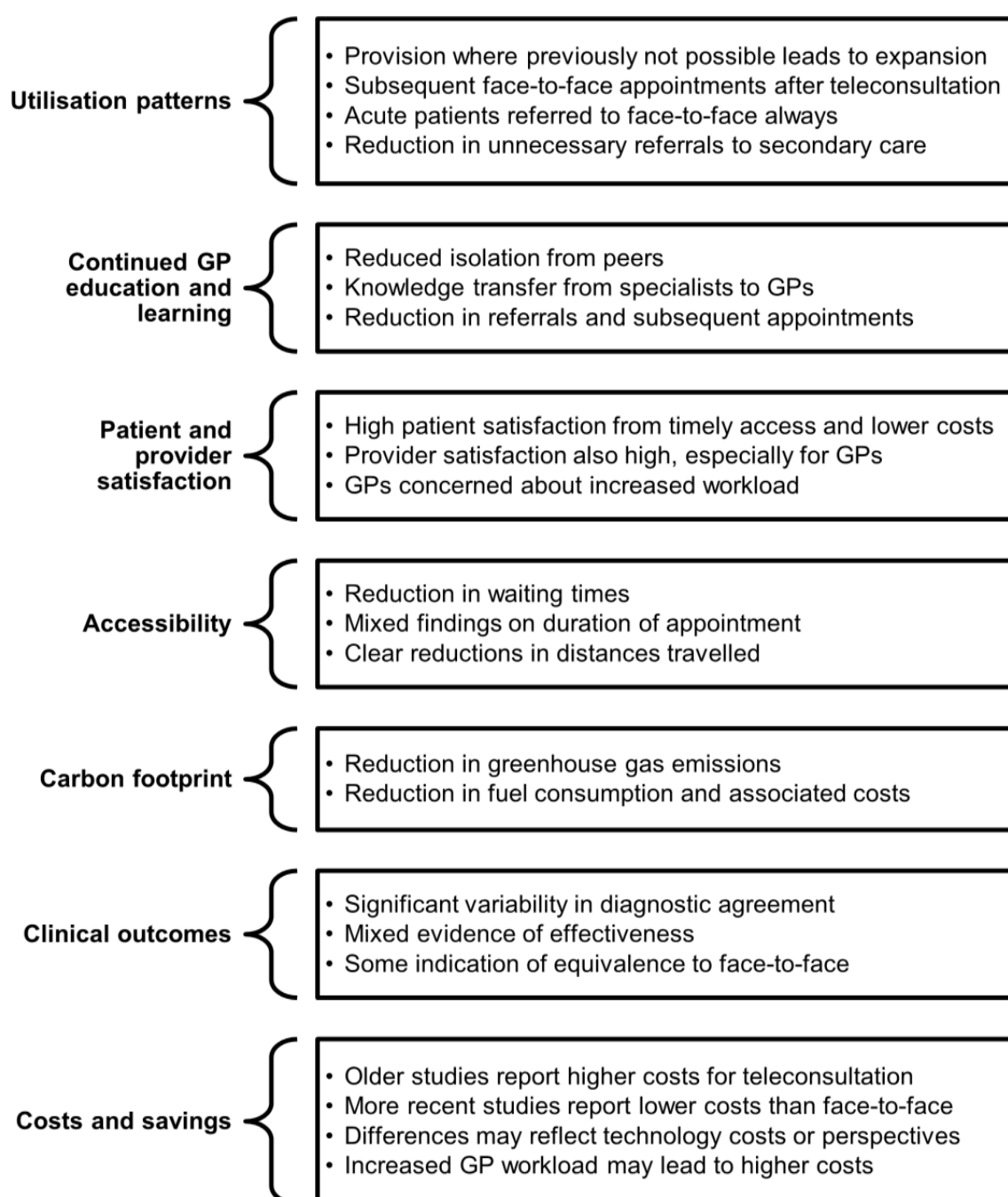


Figure 4-1 Summary of review of evidence on teleconsultations

As is typical with complex interventions, impact can be measured and assessed across several dimensions, and from various perspectives. Furthermore, there seem to be important behavioural dynamics at play with regards to the intervention's impact on utilisation patterns. Unfortunately, most studies focus on end-results rather than mechanisms. For example, waiting time is lower for teleconsultation yet no study discusses the reasons behind that finding in more detail. In summary, teleconsultations tend to be associated with lower waiting times, shorter travelling distances, reduced greenhouse gas emissions, lower patient costs, high patient and provider satisfaction, and continued GP education resulting in lower referral rates. For some time now, researchers have started incorporating all these findings into evaluation studies. These have evolved from simple break-even analyses to more sophisticated cost-effectiveness studies. However, as May (2006) points out, many questions remain regarding the quality of these evaluation efforts and the methods used.

4.1.2 Economic impact and methodological issues

The total costs of an intervention (TC_i) can be defined as the sum of the fixed costs (FC_i , for example the video-conferencing equipment) plus the variable costs (VC_i , for example physician time) times the number of episodes of service (N , i.e., the number of teleconsultations):

$$TC_i = FC_i + VC_i \times N \quad (4-1)$$

Older studies of economic impact have used break-even analysis to determine what would be the number of teleconsultations (N_{BE}) at which the total costs of teleconsultations (TC_{tele}) would equal the total costs of face-to-face appointments (TC_{f2f}):

$$TC_{tele} = TC_{f2f} \Rightarrow N_{BE} = (FC_{tele} - FC_{f2f}) / (VC_{f2f} - VC_{tele}) \quad (4-2)$$

For example, Lamminen et al. (2001) found that 92 patients would need to be seen every year in order to achieve the same total costs in both pathways. Various break-even points can be calculated by changing underlying assumptions. For example, Wootton et al. (2000) determined the break-even point (N_{BE}) for different values of distance and initial investment. There are a number of problems with break-even analyses. First, a break-even analysis presupposes that variable costs of teleconsultations are lower than the variable costs of face-to-face appointments. As Sicotte et al. (2004) shows, that may not be true. Second, break-even analysis is completely dependent on the unit costs included, their quality and generalizability. A break-even point of 92 patients calculated in Tampere, Finland, may differ substantially from a break-even point calculated in another country, or even Finnish city, where unit costs are different. For example, costs will be different if a nurse presents the patient or if a GP does (Stensland et al., 1999). Third, break-even analyses tend to aggregate costs from various stakeholders, so that the break-even point usually reflects a societal perspective. For example, Tsilimigaki et al. (2001) suggest that 18 months of paediatric cardiology teleconsultations would be sufficient to achieve savings, but as Stensland et al. (1999) discuss, providers bear the bulk of the costs with patients and employers getting the majority of the benefits. Ideally, one would calculate different break-even points for different perspectives. Even then, however, break-even analysis can be misleading if it compares costs that should not be compared. For example, when teleconsultations are used to provide care where previously no care existed, teleconsultations may be cheaper, but the hypothetical face-to-face service is not really an option, so the comparison is nonsensical. Finally, and perhaps most importantly, break-even analyses are cost analyses, they do not focus on outcomes (beyond cost savings).

Recent efforts have turned to more sophisticated economic evaluation methods. A number of these assume the effectiveness of teleconsultations is equivalent to that of face-to-face appointments, and thus conduct cost-minimisation analyses. For example, Armstrong et al. (2007) found that teleconsultations were cheaper for providers, while Eminović et al.

(2010) used a Monte Carlo simulation to show that in only 11 percent of cases was teleconsultation cheaper for society. Other studies have conducted full economic evaluations (i.e., assessed both costs and consequences). All three cost-benefit analyses reviewed by Whited (2006) found teleconsultations to be more expensive than face-to-face appointments, although it is important to mention that many of the effects mentioned in the previous paragraphs were not included in these analyses (e.g., reductions in greenhouse gas emissions). A number of articles have proposed to review the evidence on cost-effectiveness of telemedicine (Mistry, 2012, Whitten et al., 2002, Mair et al., 2000). It is important to note that these are not reviews of cost-effectiveness analyses. Rather, they review various types of economic evaluations including, but not limited to, cost-effectiveness analysis. More recent reviews have made this distinction clearer by dropping the term *cost-effectiveness* in favour of the more general *economic evaluation* (Bergmo, 2010, Dávalos et al., 2009, Wade et al., 2010, Bergmo, 2009). Measures of effectiveness in cost-effectiveness analyses have included both final clinical outcomes, as measured by QALYs (Mistry, 2012) or disability-adjusted life years (Johnston et al., 2004), and intermediate clinical outcomes such as number of consultations (Agha et al., 2002). Other more questionable measures of effectiveness have also been used, such as journeys avoided (Sicotte et al., 2004).

While there have undoubtedly been many studies incorporating some measure of economic impact (Whitten et al. (2002) found 600 a little over ten years ago), only a few have actually conducted full economic evaluations (Bergmo (2010) found only 33 studies measuring both costs and clinical outcomes). And those few are generally considered to be poor, according to best practice guidelines (Bergmo, 2010, Bergmo, 2009, Dávalos et al., 2009, Wade et al., 2010, Centre for Reviews and Dissemination, 2012). Limitations include poor generalizability, disparate estimation methods, scarcity of data from high-quality methods such as randomised controlled trials, and over-reliance on short-term small-scale pilot projects. The quality of the data is a recurring complaint (Whitten et al., 2007). The most recent report on the quality of economic evidence on teleconsultations stated there were “no

economic evaluations of relevance” and *“little robust evidence for the effectiveness and cost-effectiveness of teleconsultation*” (pages 6 and 1 respectively, Centre for Reviews and Dissemination (2012)). While the same report went on to state that absence of evidence of effectiveness is not necessarily evidence of absence of effectiveness, it could not recommend the use of teleconsultations based on the existing studies. Despite this lack of official backing, telemedicine applications are present in all seven continents (Whitten et al., 2007, Hyer, 1999, Moser et al., 2004) and, at least in Alentejo, there is significant interest in promoting wider use.

Chapter 5

TELECONSULTATION IN ALENTEJO

Alentejo is a region in Portugal. It represents approximately one third of the country's continental territory (27,329 square kilometres), but has only 5 percent of the national population (499,038 inhabitants). There are 18.3 people per square kilometre, a sixth of the national average (Instituto Nacional de Estatística I.P., 2010). A quarter of the inhabitants are over 65 years old and only 5 percent have a higher education degree. Mean monthly earnings are around 10 percent less than the national average. All these factors interact to create considerable social and economic challenges for local health policy makers and regional health authorities. To make matters worse, the supply of health care services is severely constrained. There is a shortage of health care professionals, especially hospital consultants, and in a number of specialties these are dispersed around the region's hospitals (Figure 5-1 provides a map of the Alentejo region and its health care units). For example, there are only two dermatologists (both in Évora) and one neurologist (in Beja) working for the NHS in the whole region (Administração Regional de Saúde do Alentejo I.P., 2012). The low population density is reflected in the dispersion of dwellings, and local health care units can be hundreds of kilometres away from the region's hospitals. The public transportation network is poor and, in several places, there are only a couple of bus services per day to

district capitals, where hospitals and specialists are located. It is not unusual for patients to take an entire day off to attend a hospital appointment. Furthermore, due to the characteristics of the population, patients are frequently accompanied by family members, often at the expense of a day's work.

In 1998, the regional health authorities introduced the Telemedicine Programme to maximise the impact of limited resources on meeting the needs of the population. The programme is based on the fundamental idea that patients should be seen, diagnosed and treated as close as possible to where they live and work. Its main objectives are to: increase the accessibility of specialist health care services through a reduction in distances between primary and secondary care providers, as well as a reduction in waiting times; ensure equity in the access to the best available care for all patients, through increased support for health care workers and populations living in remote underserved locations; and reduce the costs associated with health care services through, for example, avoiding unnecessary trips to hospitals.

Initial experiences involved the *Hospital do Espírito Santo de Évora* and a primary care practice in Odemira. Odemira is 165 km away from Évora, approximately 2 hours and 10 minutes by car one-way. There is only one bus service between Odemira and Évora per day. The closest hospital is Litoral Alentejano, which lacks a number of medical specialities, such as dermatology and neurology. In 2012, Odemira was home to just under 26,000 people, while Litoral Alentejano's hospital catchment area included close to 100,000 inhabitants (Instituto Nacional de Estatística I.P., 2013). Numbers were similar in 1998. According to documentation from the Central Administration of the Health System (a public institute that manages the financial and human resources, facilities and equipment, and systems and information technologies of the Portuguese NHS, and is also responsible for the definition of policy, regulation and planning of health, along with the regional health authorities), 50 to 60 appointments can adequately serve the dermatological needs of 1,000 inhabitants in any given year, with 35 to 40 percent of those being first appointments (Administração Central do

Sistema de Saúde I.P., 2008). That would mean between 1,750 and 2,100 first appointments for Litoral Alentejano. From 2002 to 2011, first appointments in dermatology in Évora ranged from 695 (in 2008) to 2,002 (in 2012), barely enough to cover Litoral Alentejano, much less other areas of Alentejo. Although anecdotal, this shows the extent of problems with supply of specialised care, notably dermatology, in the region.

Currently, there are twenty six telemedicine platforms distributed among twenty primary care units and five hospitals, covering four districts (see Figure 5-1).

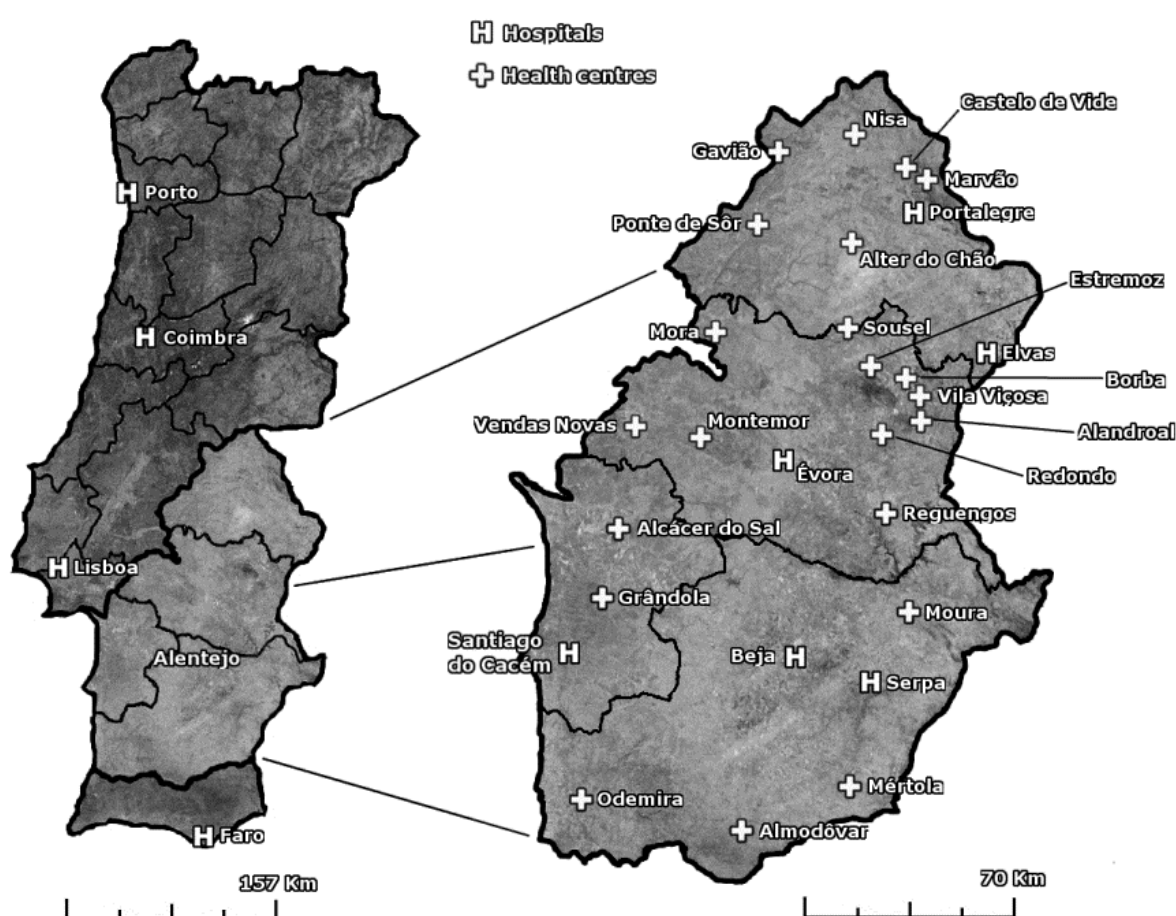


Figure 5-1 Telemedicine network in Alentejo

Platforms communicate over the Health Information Network (*Rede Informática de Saúde*), a high speed virtual private network managed by the Ministry for Health. The network is provided free of charge to health care units in the Portuguese NHS. The telemedicine platforms include high resolution video-conferencing equipment, access to

electronic patient records and picture archive, and peripherals such as electronic dermatoscopes and stethoscopes³ (availability varies between units). The most specialised care is provided in Évora. There are other hospitals (so-called district hospitals) in Beja, Portalegre, Elvas and Litoral Alentejano. The availability of specialised services is poor and dispersed (dermatology is only available in Évora and neurology in Beja).

Teleconsultations are available in a number of specialties and clinical areas: diabetes, traumatology, orthopaedics, general and paediatric surgery, respiratory medicine, urology, gastroenterology, clinical oncology, cardiology, dermatology, physical medicine and rehabilitation, pain, neurology, thyroid and obesity. Besides real-time teleconsultations, the video-conferencing platforms are also used for teleradiology (computed tomography and x-rays), ultrasound telemedicine and tele-education. Tele-education courses have covered such topics as informed consent, infection control and bronchial asthma. Provider satisfaction with tele-education courses has been high (Cravo Oliveira et al., 2012).

In Portugal, GPs act as gatekeepers: to see an NHS specialist, a patient must be referred by a GP. Primary care units which are part of the telemedicine network must refer their patients to teleconsultations. Subsequent face-to-face appointments may then be arranged. Exceptionally, patients that present with potentially severe conditions can be referred directly to face-to-face appointments, bypassing teleconsultations. Teleconsultations are real-time with the presence of a GP. Some primary care units appoint a coordinator: a GP who is present in every teleconsultation for a specific specialty; in others the patient's own GP is present. Treatment is usually prescribed by the GP, except when special rules apply (e.g. neurological drugs which must be prescribed by a qualified neurologist). Prescriptions written by specialists are mailed to patients. Besides managing schedules, support staff operate the video-conferencing equipment during teleconsultations, so that physicians can focus on the patient.

³ An electronic dermatoscope is a medical device with a magnifying lens and a light source, used to conduct non-invasive painless skin examinations and record images to be digitally analysed. An electronic stethoscope is a device which can be used to amplify and record heart and lung sounds.

5.1 Impact of the programme

Very few data are available before 2002. It is thus difficult to achieve an accurate picture of what health care was like in Alentejo before the introduction of teleconsultation in 1998. It is however possible to look at some of the more recent statistics to gain an understanding of what the baseline was, especially as a number of practices started using teleconsultation in different years. It helps that the programme developed as a pilot project between 1998 and 2002, and thus teleconsultations represented only a small percentage of all outpatient appointments.

In 1998 there were a total of 30 teleconsultations. A decade later, this number was up to 3,717, and by the end of 2011 a cumulative 27,852 teleconsultations had been performed in Alentejo (Administração Regional de Saúde do Alentejo I.P., 2012). A third of all teleconsultations were in dermatology, followed by neurology and cardiology. In recent years, the number of teleconsultations in physical medicine and rehabilitation and respiratory medicine has risen significantly, with urology and psychiatry performing fewer appointments. Other uses of the video-conferencing platforms have also expanded. Since the start of the tele-education sessions in 2008, 876 professionals have participated in courses, with as many as ten sites connected simultaneously. Between 1998 and 2011, there were 102,753 teleradiology episodes of service.

5.1.1 Exploratory interviews

In November 2010, I conducted twelve semi-structured exploratory interviews with physicians and primary care managers who were, or had been, associated with the programme (an interview schedule/guide was used; a copy – in Portuguese – is provided in Appendix A). The participants included seven hospital consultants (dermatology, cardiology, psychiatry, physical and rehabilitation medicine, neurology and gastroenterology, general surgery), two primary care managers, and three GPs. Much of what came out of those

interviews mirrors the evidence presented in the previous chapter. See, for example, participants' responses to what they considered to be the advantages of teleconsultations (Table 5-1). It is noteworthy that the majority of the benefits, as perceived by physicians and managers, relates to non-clinical outcomes and accrues to patients and their families rather than the health care system.

Table 5-1 Advantages of teleconsultation as reported in exploratory interviews

What are the advantages of teleconsultation?	Agreement (N = 12)
Reduces need for transportation and associated costs	10
Reduces time lost by patients and those accompanying patients	9
Reduces unnecessary visits to the hospital	7
Contributes to continual medical education	7
Increases geographical access	5
Increases the service capacity and productivity	4
Reduces number of referrals	4
Reduces the probability that patients' conditions get worse	1

Teleconsultations were deemed to be quicker than face-to-face appointments, hence the increase in capacity and productivity. The presence of GPs was considered an advantage, not only because of continued education, but also because the GP had a superior understanding of the patients' needs and their clinical history, compared to that of the specialist. Indeed, this greater understanding of patients' histories was said to explain why teleconsultations were quicker. The majority of participants were emphatic that teleconsultation could not be a substitute for face-to-face appointments, but that it could prevent having certain patients go to the hospital. Only two participants considered teleconsultation a potential substitute. This, despite the fact that most considered teleconsultations could reach (i.e., effectively substitute for) as much as 60 percent of all specialist consultations (depending on the specialty). This growth potential was considered desirable.

Subsequent face-to-face appointments were needed following, at most, 20 percent of teleconsultations. Reasons for subsequent face-to-face appointments included technical difficulties, uncertainty in establishing a diagnosis, need for face-to-face procedures (such as examinations or surgery), suspicion of severe condition and resistance to treatment. On the subject of waiting times, three participants considered that patients waiting for specialist consultations might resort more to their primary care physician while they wait, however no quantitative evidence was provided. The majority of participants stated their workload had increased following the introduction of teleconsultation, and that there were no financial incentives to using teleconsultation at the physician level. Five participants were unable to identify any disadvantages associated with teleconsultation and most would encourage other physicians to use it.

There were a number of instances where it was clear that teleconsultations had an impact on utilisation. One participant stated “*[area A] was a forgotten region (...) [after the introduction of teleconsultations] very quickly, we started having a large number of patients for small as well as big surgeries, it is one of our biggest providers*’. A number of participants considered that teleconsultations might lead to a reduction in referrals, as a consequence of continued education and learning. The presence of GPs in teleconsultations increases their experience in a specific specialty, potentially reducing the number of patients they refer in the future. As one participant put it ‘*six, seven months after [the introduction of teleconsultations] we reduced the number of referrals by 50 percent, because we looked and knew, we did the diagnosis*’.

5.1.2 Survey

The exploratory interviews were useful towards understanding the impact of teleconsultations but were limited in two important ways. First, they provided very few quantitative data, and the few data provided was anecdotal. The second limitation is that only providers were interviewed, and so no information on patients was available. To deal

with these limitations, I conducted a survey of 100 teleconsultation patients and 100 face-to-face outpatients who had appointments in the second half of 2011 with consultants from the *Hospital do Espírito Santo de Évora* (a copy of the survey instrument – in Portuguese – is provided in Appendix B). The results indicate teleconsultation had a positive impact on patient experience (Cravo Oliveira et al., 2013), expenses and general quality of service (a number of selected results can be seen in Table 5-2). In a sample of just 100 patients, 2,841 km of travelling were avoided and 134 hours (almost 6 days) gained (not including waiting time).

The average distance travelled was 6 km for a teleconsultation compared to 47 km for a face-to-face appointment. The average time taken to attend an appointment (i.e., return travelling time plus physical waiting time plus appointment duration) was 93 minutes for a teleconsultation compared to 190 minutes for a face-to-face appointment. The average waiting time (i.e., the time between the day the patient is referred for an outpatient appointment and the day the appointment takes place) for a teleconsultation was 26 days compared to 74 days for a face-to-face consultation. Fewer patients in the teleconsultation pathway had subsequent appointments than in the face-to-face pathway. Yet, there was no information on whether subsequent appointments following teleconsultations were face-to-face or teleconsultations.

A number of patients are entitled to NHS-paid transportation to outpatient appointments by ambulance. While 8 percent of face-to-face survey respondents travelled in an ambulance, only 2 percent of teleconsultation patients did so. Most significantly, 39 percent of teleconsultation patients walked to their appointment, avoiding costs and emissions but also potentially promoting their own health (Jarrett et al., 2012). The average cost for teleconsultation patients in the sample was estimated at €9.31 compared to €25.32 for face-to-face patients, a reduction of 63 percent (costs included co-payment, travelling and time-off-work costs for both patients and those accompanying, estimated from survey

data on average income and means of transportation). In general, for patients living in remote locations (e.g., Odemira) savings could be considerably higher.

To estimate the impact of teleconsultation on greenhouse gas (GHG) emissions, survey results were combined with data from the *Hospital do Espírito Santo de Évora* on the referring and receiving health care units for all 20,824 teleconsultations performed in Alentejo from 2004 to 2011 (Cravo Oliveira et al., 2013). If teleconsultations had not been available, patients would have travelled an estimated 2,313,819 km more with an extra 455 tonnes of carbon dioxide equivalent (tCO₂e, a universal unit of measurement that allows the global warming potential of different GHGs to be compared) being emitted in the process (22 kg CO₂e per patient). The use of teleconsultation resulted in a 95 percent reduction in total direct (emissions at the point of use of a vehicle) and indirect (emissions from production and distribution of fuels to their point of use) GHG emissions associated with patient travel to appointments.

The sensitivity analysis showed that in the best-case scenario (lower quartile survey distances for teleconsultation), teleconsultation patients would cover about 34,000 km; in the worst-case (upper quartile) total distance would reach as high as 159,000 km. The reduction in total direct and indirect GHG emissions from the use of teleconsultation would be 448 tCO₂e in the worst-case scenario (a reduction of 94 percent from the 477 tCO₂e emitted to attend face-to-face appointments), and 472 tCO₂e in the most favourable scenario (a reduction of 99 percent). Using certified emission reduction credits from carbon trade schemes, it was estimated the 455 tCO₂e potentially avoided through the use of teleconsultation in Alentejo could be worth between \$4,500 and \$6,800 at current prices, and €22,700 at the level analysts believe is needed to achieve Kyoto protocol reductions (Cravo Oliveira et al., 2013). It is important to note that reducing GHG emissions is not the main rationale for teleconsultations and they are rarely even considered as evidence of economic impact. Yet, as previously suggested (Pencheon, 2009), these reductions should be considered when assessing teleconsultations as that their impact is clearly not negligible.

Table 5-2 Selected results from survey of 200 patients in Alentejo

Variable	Teleconsultation	Face-to-face
Gender		
Male	40	37
Female	58	63
<i>Total</i>	<i>100</i>	<i>100</i>
Medical specialty of appointment		
Dermatology	69	21
Neurology	13	7
Physical and rehabilitation medicine	12	21
Surgery	6	51
<i>Total</i>	<i>100</i>	<i>100</i>
Income		
< €7,000	62	40
€7,000 - €15,000	30	42
€15,001 - €30,000	5	6
€30,001 - €60,000	1	2
<i>Total</i>	<i>98</i>	<i>99</i>
Level of education		
No formal qualifications	10	8
Basic education	44	53
Secondary education	32	29
Higher education	6	8
<i>Total</i>	<i>92</i>	<i>98</i>
General health status		
Very good	3	2
Good	30	16
Neither good nor bad	49	50
Bad	17	28
Very bad	1	4
<i>Total</i>	<i>100</i>	<i>100</i>

Table 5-2 Selected results from survey of 200 patients in Alentejo (continued)

Variable	Teleconsultation	Face-to-face
Had subsequent appointment		
Yes	47	63
No	52	36
Total	99	99
Means of transportation		
Car	57	76
Bus	2	9
Taxi	0	2
Ambulance	2	8
Walk	39	5
Total	100	100
Satisfaction		
Very satisfied	42	26
Somewhat satisfied	54	68
Indifferent	2	3
Somewhat unsatisfied	2	3
Total	100	100
Patient came accompanied		
Yes	46	67
No	54	33
Total	100	100
Occupation of companion		
Active (employed)	16	40
Inactive (retired, unemployed, student)	30	27
Total	46	67

5.1.3 Teleconsultation in other regions of Portugal

The use of telemedicine in other regions of Portugal is extremely limited compared with Alentejo. In 1998, the same year the Alentejo Telemedicine Programme was created, the paediatric hospital in Coimbra started a telecardiology consultation in association with

hospitals in Leiria and Oporto (Castela et al., 2005). Since then, other hospitals (Santa Maria da Feira, Lisboa, and Covilhã in Portugal, Mindelo in Cape Verde, Luanda in Angola, and Madrid in Spain) have joined the service and between 1998 and 2006, there were 4,741 teleconsultations (Associação Portuguesa para a Promoção e Desenvolvimento da Sociedade da Informação and Associação para o Desenvolvimento da Telemedicina, 2007). Based on the few available studies, other projects are either small or in early stages of implementation (Burnay et al., 2013, Costa et al., 2013).

In July 2012, the Ministry for Health set up a working group to develop a national telemedicine program drawing on the experience of the Alentejo Telemedicine Programme. Among the priorities identified by the group was to advance the use of teleconsultation in other regions of Portugal, and ensure that lawmakers and payers develop legislation and policies that promote the use of telemedicine (Grupo de trabalho para a telemedicina, 2013).

5.2 Evaluating teleconsultation in Alentejo

The review of evidence on teleconsultations presented in Chapter 4 is somewhat puzzling. There is an astounding wealth of research into the many effects of teleconsultations and yet not one relevant economic evaluation can be found by the Centre for Reviews and Dissemination (CRD). Despite the CRD's prominent role in informing such institutions as the UK NICE, a lack of formal support for the intervention has not stopped hundreds of locations around the world (including the UK) promoting the use of teleconsultation. If anything, practitioner support only seems to be growing (Doarn and Merrell, 2013). As this chapter shows, the same applies to Alentejo. There is abundant data on over 10 years of teleconsultation. Yet, very limited use of the data has been made, and there have been no evaluations of economic impact. Despite this, the Portuguese Ministry for Health has created a working group to promote a national telemedicine programme, building on the Alentejo experience.

It seems unreasonable that so many primary and secondary studies, so many data, from so many different places, exploring so many different effects, have produced so little relevant material for economic evaluations. As I argued in Chapter 3, the problem may lie with previous approaches to evaluation, not with lack of evidence or data. A starting point would be to acknowledge that teleconsultations involve multiple interdependent components, are associated with significant behavioural changes, lead to many diverse effects within and outside the health care system, and are often tailored to specific contexts and settings, at different levels of the health care system. In essence, teleconsultations exhibit all the defining characteristics of a complex health intervention. It is possible that previous efforts to evaluate the economic impact of teleconsultation have failed because they have controlled for, rather than explicitly address, these characteristics.

Evaluating teleconsultation in Alentejo needs to be about whether it works as well as how. First, there are numerous examples in the literature, and in Alentejo, of how teleconsultation affects utilisation in non-trivial ways. It is unrealistic to assume that all these can be put into a black box, looking only at the total episodes of service at the end of the evaluation period. Opening these black boxes of utilisation, and identifying the important causal mechanisms at play, is both desirable and possible. Second, instead of relying on a single measure of outcome, one should assess as many outcomes and effects as possible, such as avoided waiting and travelling, and the number of appointments provided. Third, one should make room for different scales of implementation and time-frames. With much of the data coming from a handful of practices, it is important to understand how the costs and consequences change with scale, time and contextual tailoring. Which brings me to the fourth and final point: context and flexibility should be made explicit. For example, in some practices in Alentejo, there is a coordinating GP who performs all of the teleconsultations in that practice; in others, each GP is present in his/her patients' teleconsultations. This apparently small detail has implications for continued GP education and learning, and consequently referrals rates.

Given recent policy developments in the region, and in Portugal in general, promoting a wider use of the technology, it is especially timely to explore the costs and consequences of implementing teleconsultation at a larger scale. In essence, the evaluation should have policy and theoretical implications, informing decision makers in Alentejo of the impact of teleconsultation, and also contributing to previous research on how and why teleconsultation works. Furthermore, it should signal areas which warrant further investigation and data collection.

In summary, the evaluation should: 1) address the impact of behaviour and preferences on outcomes, through explicit modelling of causal mechanisms; 2) assess multiple costs and consequences, presented in a disaggregated manner that allows subsequent valuation; 3) incorporate different perspectives; 4) examine the impact of different scales and time-frames; 5) explicitly account for contextual features; and 6) inform decision makers of the likely effects of different policy levers.

5.2.1 Studies and techniques

Simulation modelling, informed by previous theoretical and empirical research on individual decision making and behaviours, can be used to conduct an evaluation of economic impact of teleconsultation in Alentejo that meets all six requirements set above. By answering the questions in Table 5-3, the choice of simulation technique to use, given those six requirements, is straightforward: system dynamics. Answers to some of the questions are unclear at the outset (e.g., the importance of randomness and the scope). Notwithstanding, other answers are clear and the result is quite definitive.

To meet the six requirements above, it is also important to be unequivocal in defining the object of evaluation. As previously stated, the use of teleconsultation in different specialties leads to markedly different costs and consequences. Indeed, it is highly questionable whether telecardiology and teledermatology should be grouped under the term teleconsultation and evaluated as one. As such, I focus on a specific medical specialty. I

chose dermatology for a number of reasons: 1) most importantly, it accounts for the large majority of teleconsultations since the beginning of the programme, thus providing ample data; 2) dermatological conditions tend to develop independently of other illnesses, allowing co-morbidities to be disregarded and impact on other parts of the health care system to be limited (Chen, 2001); 3) supply of dermatology physicians has been relatively constant for the last decade, effectively eliminating a piece of the puzzle.

Table 5-3 The right simulation technique to evaluate teleconsultations in Alentejo

	Answer	Technique
<i>Can behaviour be aggregated?</i>	Yes	SD
<i>Is variability/randomness important?</i>	Unsure	SD, DES & ABM
<i>Do changes to the system happen over time?</i>	Yes	SD
<i>Is the time-frame of interest long?</i>	Yes	SD
<i>Is the number of entities large?</i>	Yes	SD
<i>Is feedback important?</i>	Yes	SD
<i>Is the scope more operational rather than strategic?</i>	Unsure	SD, DES & ABM
<i>Is understanding more important than optimisation?</i>	Yes	SD

SD – system dynamics; DES – discrete event simulations; ABM – agent-based modelling

To parameterise the model and run simulations, considerable amounts of data are needed. Some of the parameters will come from straightforward statistical analyses (e.g., simple counts or averages, such as the ones presented in this chapter) while others will require more sophisticated econometric and statistical analyses. While the model builds on a large body of literature (both on decision making in general and on teleconsultation specifically) and regional primary data, at least two areas warrant further study, both due to their potential for determining final outcomes, and due to limitations of previous research: the impact of waiting times on utilisation, and the determinants of GPs' decision referrals. Based on previous studies of these phenomena, more sophisticated econometric methods must be used. In the following chapters, I describe how I undertook these two studies, providing detailed discussions of the methods used.

Chapter 6

WAITING TIMES AND UTILISATION

Waiting times can influence rates of utilisation of health care services. As illustrated in the literature review in Chapter 2, waiting times can affect the patient's decision to seek care (Scott, 2000), the GP's decision to refer (Goddard and Tavakoli, 1998), and the patient's decision to adhere to or complete physician recommendations (Zuckerman et al., 2011a, Zuckerman et al., 2011b). Teleconsultations have been associated with significant reductions in waiting times for outpatient appointments and important changes in patterns of utilisation, as discussed in Chapters 4 and 5. It is thus important to explore how waiting times and utilisation are related, especially as utilisation is strongly associated with costs and consequences.

In this chapter, I build on previous work investigating the impact of waiting times on demand. Besides contributing to the literature on waiting times and demand, this study has implications for the evaluation of teleconsultation in Alentejo. Both are discussed at the end of the chapter.

6.1 Background

Waiting times can influence rates of utilisation of health care services, as illustrated in the literature review in Chapter 2. Waiting lists, and associated waiting times, are considered a key policy issue, not only because they are frequently associated with patient dissatisfaction, but also because they are affected by policy interventions (Siciliani and Hurst, 2005). In their review of elective surgery waiting times in twelve countries of the OECD, Siciliani and Hurst (2005) suggest that increases in waiting times may be caused by advances in surgical technology. Because many procedures have a lower unit cost and can be carried out as day cases, there have been dramatic increases in demand for surgical procedures in all OECD countries. Much work has gone into quantifying the actual impact of waiting lists on demand (Martin and Smith, 1999, Martin and Smith, 2003, Windmeijer et al., 2005, Gravelle et al., 2002, Goddard and Tavakoli, 1998).

Waiting lists are typically found in health care systems where there is tax-financed insurance and a budget on expenditure (Cullis et al., 2000). In such systems (e.g., Sweden, UK, New Zealand, Portugal, etc.), health care services are either free at the point of delivery or heavily subsidised. Because financial barriers to access are low, there is the potential for moral hazard: ex ante, reducing one's efforts to prevent illness, or ex post, increasing one's use of health care services when ill (Zweifel and Manning, 2000). In practice, however, budget constraints prevent capacity from infinitely matching demand. Waiting lists come into being when demand exceeds capacity. Since more severe clinical presentations require urgent care, waiting lists do not operate on a *first come first served* basis (Siciliani and Hurst, 2005). Rather, they tend to include different streams of patients moving at different speeds (and consequently with different waiting times) based on perceived clinical urgency. Movement between streams is possible if a patient's condition deteriorates.

Since people do not actually wait physically in a line, there is no concept of wasted time from being on a waiting list, and thus no opportunity cost in the form of wasted time

(Martin and Smith, 1999). Naturally, there may be other costs to waiting, such as not being able to work, being in pain or discomfort, etc. The issue is one of optimal timing of treatment (Propper, 1995): the utility patients derive from whatever medical good or service they are on the list for declines the longer people wait for it (Lindsay and Feigenbaum, 1984). It is this declining utility that makes waiting lists a rationing device. Previous researchers have sought to quantify the impact of waiting time on demand using a number of approaches. Both cross-sectional (Martin and Smith, 1999) and panel data (Martin and Smith, 2003) have been used, at the national (Gravelle et al., 2002), regional and hospital level (Windmeijer et al., 2005), and for inpatient (Siciliani and Hurst, 2005) and outpatient care (Windmeijer et al., 2005). All these studies have one thing in common: the use of elasticities to represent the impact of waiting times on quantities demanded and supplied. The elasticity of demand for, or supply of, a service (e.g., elective surgery, outpatient appointments, etc.) with respect to some quality or characteristic of that service (e.g., price, waiting time, distance, etc.) is defined as the percentage change in quantity demanded brought about by a one percent change in the quality or characteristic. The elasticity of demand (E_D) with respect to waiting time (W_t), for example, is:

$$E_D = (\Delta Q/Q) / (\Delta W_t/W_t) \quad (6-1)$$

where Q is the quantity demanded (number of entrants to the list). The elasticity of demand with respect to waiting time, for example, captures different phenomena: patients deciding to go to private providers because waiting times are too long in public services, patients getting better with time even without formal care, patients getting worse with time, etc. Waiting lists are naturally a consequence of both inflows (i.e., patients being added to the list) and outflows (i.e., patients exiting the list). While patients may leave the list at any time (González-Busto and Garcia, 1999), this is not usually known until the date of their exit, and so waiting lists often contain *ghosts*: patients who are still in the waiting list even though they have given up on the service (Windmeijer et al., 2005). Inflows basically represent

demand while outflows represent supply and non-completion. At any given point in time, the waiting time is determined by the number of patients in the list, plus the number of patients flowing in (i.e., demand), divided by the number of patients flowing out, both through consumption of the service (i.e., supply) and non-completion (i.e., the ghosts). As such, previous studies have explored both inflows and outflows, although the non-completion outflow has received significantly less attention.

Besides using different sources of data, previous studies have used different measures of demand and waiting time to determine elasticities (the four studies mentioned in this paragraph are well-known examples, illustrative of a larger body of literature on the topic). For example, Gravelle et al. (2002) explored routine surgical admissions at the level of English Health Authorities over 24 quarters between 1987 and 1993. They estimated an elasticity of demand with respect to the proportion of patients waiting more than three months, lagged one quarter, of -0.21. Martin and Smith (1999) investigated elective surgical episodes from a cross-section of 4460 synthetic wards (i.e., an indicator of area of residence) for 1991-1992. They found an elasticity of demand with respect to the standardized waiting time (i.e., a measure of waiting time that takes into account both area and national characteristics such as population size, and age and sex profile) of -0.2063. In a later study, Martin and Smith (2003) extended their cross-section of 4460 synthetic wards to a panel of 5499 electoral wards between 1992 and 1998. As before, they explored elective surgical episodes across a number of specialties, finding an elasticity of demand of -0.09 across all specialties (elasticities ranged from -0.24 for general surgery to 0.38 for urology). Finally, Windmeijer et al. (2005) investigated rates of first outpatient appointments in a single Scottish hospital serving an area with around 375,000 inhabitants. They used a panel of 61 primary care practices observed over 19 quarters between 1997 and 2001, and estimated an elasticity of demand with respect to the mean realised waiting duration, lagged two quarters, of between -0.3117 (estimated using ordinary least squares) and -0.3903 (estimated using generalised method of moments).

A key issue in all these studies is the use of proxy measures of demand. Gravelle et al. (2002) use surgical admissions, which do not necessarily lead to entrance in the list, and the other three use actual episodes (i.e., exits from the list) rather than entries (Martin and Smith, 1999, Martin and Smith, 2003, Windmeijer et al., 2005). As Martin and Smith (2003) state, “*a study of waiting times should in principle analyse entrants to and exits from the waiting list, and their responses to expected waiting times*” (page 370), but invariably the availability and quality of data on entries is poor. As such, these researchers assume that the waiting list is in equilibrium – the number of entrants, at any given time period, equals the number of exits – and that exits can be used in place of entries (Martin and Smith, 2003). This assumption may be too strong in certain settings (e.g., Alentejo). Previous studies of non-completion have calculated rates of non-completion and have explored the determinants behind non-completion (usually using odds-ratios and logistic regression), one of them being waiting times (Zuckerman et al., 2011a, Zuckerman et al., 2011b, Forrest, 2011, Forrest et al., 2007). None of these studies have estimated the quantitative impact of waiting time on non-completion at an aggregate level (e.g., hospital or practice level).

Using data on actual referrals, rather than proxy measures, and on missed appointments, between July 2001 and August 2010, from the *Hospital do Espírito Santo de Évora*, the most specialised hospital in Alentejo, I estimate elasticities for demand and non-completion with respect to waiting time.

6.1 A model of referrals

Various economic models of demand have been proposed in the literature. Demand is said to depend on the clinical need for care, the costs of getting on the list (e.g., travel costs), the perceived waiting time, and the supply of other health services (Gravelle et al., 2002, Martin and Smith, 2003):

$$referrals = f \left\{ \begin{array}{l} waiting\ time(-), need(+), costs(-), \\ provision\ of\ private\ services(-), \\ provision\ of\ other\ public\ services(?) \end{array} \right\} \quad (6-2)$$

The expected direction of each effect is indicated in parentheses above. Waiting time, entry costs, and access to private care, are all expected to have a negative impact on demand for NHS dermatology outpatient appointments, while clinical need should have a positive effect. The impact of access to other public services is not clear, as they can act both as a substitute and a complement (Gravelle et al., 2002). The measure of referrals used was total referrals per practice per quarter per 1,000 inhabitants. The measure of waiting time used was the total waiting time – defined as the time elapsed between the date of referral and the date of the appointment – lagged one quarter (other lags were tested as well but are not reported here).

Clinical need poses difficulties. Martin and Smith (2003) use a general index of acute care needs which measures a small area's relative need for NHS expenditure adjusted for sex and age profile. It is based on a number of socio-economic factors (e.g., the proportion of economically active unemployed) and health statistics (e.g., standardised mortality ratio). Gravelle et al. (2002) use similar data from the Office for National Statistics 1991 Census of the Population. Windmeijer et al. (2005) also use illness ratios, but also age and sex variables. In Portugal there is no publicly available data on illness ratios or any index of clinical needs, at least not at the level of primary care practices. As such, a combination of factors from all three studies was selected according to data availability. These included age and sex profile, standardised mortality ratio, proportion of pensioners and average value of pensions, and proportion of beneficiaries of *rendimento social de inserção* (a social inclusion benefit). Ideally, one would use data that can inform the need for dermatology appointments specifically. In practice, it is unclear what these data would be, and whether they would be available at the level of GP practices.

Costs of entering the waiting list for outpatient appointments were measured in terms of travelling distance (in km) and time (in minutes). Access to other health services, both

private and public, is also expected to affect demand. A measure of provision of private care would ideally include both geographical accessibility and costs, yet no data on the latter is available. As for geographical accessibility, addresses of private providers with dermatology services in Alentejo were provided by the *Entidade Reguladora da Saúde* (the Portuguese Health Regulation Authority) and used to determine the closest private provider for each practice. As for measures of provision of other public health services, these included number of GPs in practice, the number of pharmacies in the municipality, and the travel distance and time to the hospital in Beja. Although the hospital in Beja did not at any point in the period under study provide dermatology appointments, a number of dermatology conditions can be treated by general surgeons and so there is the possibility that some patients (especially those closest to Beja) might seek to get an appointment in the hospital in Beja. Details on how distances and times were estimated are provided in Appendix A.

6.2 A model of non-completion

Previous literature on referral non-completion has used individual-level data to predict the probability of non-completion given individual characteristics. Data at the level of the individual is the best for understanding why patients miss appointments. Because activity data rarely provides enough detail on individual characteristics, such studies usually complement data collection with surveys (Forrest et al., 2007). Such studies provide excellent opportunities to explore the impact of patient and provider socio-economic and psychological factors on the probability of non-completion (Zuckerman et al., 2011b). At the aggregate level, such individual characteristics are likely to be less important, since individual idiosyncrasies are diluted in average population effects (naturally, the level of aggregation determines the type of explanatory variables that need to be included). To my knowledge, there have been no studies of the impact of waiting time on referral non-completion at the aggregate level so no economic models exist. I assume that the same

factors that affect the decision to enter the list, also affect the decision to exit before the appointment takes place:

$$non - completion = f \left\{ \begin{array}{l} waiting\ time(+), need(?), costs(+), \\ provision\ of\ private\ services(+), \\ provision\ of\ other\ public\ services(?) \end{array} \right\} \quad (6-3)$$

The expected effects of waiting time, costs and private care are now positive (e.g., the longer the waiting time the bigger the number of patients missing appointments). As for need and provision of other public services, the expected effects are less clear. As with referrals, other public services (such as primary care service) can both complement and substitute for secondary outpatient appointments. As for clinical need, a sicker population may require care more urgently. If the urgency stream mechanism in waiting lists is efficient, then such a population is more likely to have quicker appointments and thus higher completion rates. On the other hand, patients might seek alternatives sooner and thus non-completion will be higher. The ultimate effect is ambiguous.

The paucity of behavioural models of non-completion makes it difficult to determine which model to adopt here (and a more detailed study of the non-completion phenomenon is beyond the scope of this thesis). The measure of non-completion used was the total number of missed appointments in a given quarter for a given practice per 1,000 inhabitants. The measure of waiting time used was the total average waiting time. No lags were used as, behaviourally, there is no clear rationale for lagging the waiting time (patients do not miss appointments based on the previous quarter's waiting time; in fact, with waiting times of 200 days, after a quarter they may already have decided to not complete). Measures of costs, need and provision of other services were the same as for referrals.

6.3 The data set

The principal data set is based on a database of outpatient activity from the *Hospital do Espírito Santo de Évora*, the most specialised hospital in Alentejo. The database contains

information on outpatient appointments since July 2001, when the information system that collects and aggregates the data was first introduced. The data set in this study includes all records of outpatient referrals for first appointments, completed first appointments and missed first appointments for the period between July 2001 and August 2012, for practices in health sub-regions *Alentejo Central I*, *Alentejo Central II* and *Baixo Alentejo*. There were a total of 9,126 records of referrals which led to completed appointments, and 3,474 records of referrals which were not completed. There were a number of errors and issues with the data that had to be analysed one by one. This process of cleaning the data is described in Appendix A. The final data set includes 12,600 records of referrals, including both completed and missed appointments for the period between July 2001 and August 2012.

The data set includes dates for a number of events or stages between the referral and the actual appointment. Namely, the date of the referral, the date the referral request is sent for triage, the date the result of the triage is known, the date the appointment is booked, and the date the appointment takes place. Table 6-1 shows the number of completed and non-completed appointments and respective total waiting time (for simplicity, other components of waiting times are not reported) between July 2001 and August 2012. The total waiting time is the time elapsed between the date of referral and the date of the appointment.

As Table 6-1 shows, there are very few referrals for 2001-2002. This is understandable, since this period includes data for only a few months. Furthermore, data on appointments is only available for referrals from July 2001, so appointments for which referrals were made before July 2001 are censored. As average waiting times are calculated directly from the data, this means that waiting times for the initial periods are likely to underestimate the true waiting times. Table 6-1 also illustrates the significant difference between waiting times for completed referrals and non-completed referrals. As expected, non-completion seems to be associated with longer waiting times. Again, because of censoring in the initial periods, numbers of non-completed referrals are probably underestimated.

Even despite initial censoring, the data set allows the use of referrals, rather than appointments, as the dependent variable. This is in contrast with Windmeijer et al. (2005) who use appointments as the dependent variable because they only observe referrals when they lead to appointments, and so at the end of the study period they have few referrals. Furthermore, given their study period is 1997-2001, eliminating the final quarters where censoring occurs would significantly reduce the data available for estimation. On the other hand, in this study censoring exists in initial periods. After calculating the waiting times for each quarter, it was clear that eliminating the two quarters of 2001 and the first quarter of 2002 would adequately deal with censoring (the differences in completed referrals and waiting times from 2003-2004 to 2004-2005 are due to changes in supply, as discussed later in Section 8.2, not issues with data).

Table 6-1 Average waiting times for dermatology appointments, 2001-2013

Period	Referrals	Completed		Non-completed		Δwait*
	<i>Number</i>	<i>Number</i>	<i>Avg. wait*</i>	<i>Number</i>	<i>Avg. wait*</i>	
<i>2001-2002</i>	257	257	70	-	-	-
<i>2002-2003</i>	1,419	1,124	97	295	103	6
<i>2003-2004</i>	1,417	1,016	99	401	112	13
<i>2004-2005</i>	1,006	644	157	362	180	23
<i>2005-2006</i>	968	518	261	450	301	40
<i>2006-2007</i>	1,039	553	309	486	366	57
<i>2007-2008</i>	1,317	815	276	502	353	77
<i>2008-2009</i>	847	617	169	230	162	-7
<i>2009-2010</i>	1,060	806	207	254	259	52
<i>2010-2011</i>	936	772	216	164	259	43
<i>2011-2012</i>	1,322	1,150	183	172	232	49
<i>2012-2013</i>	1,012	854	161	158	200	39
<i>2001-2013</i>	<i>12,600</i>	<i>9,126</i>	<i>180</i>	<i>3,474</i>	<i>244</i>	<i>64</i>

* Waiting times in days.

Each referral record is associated with a primary care practice. The period between April 2002 and June 2012 was divided into quarters (first quarter of each year starts on 1st of January and finishes on 31st of March, and so on; there are 41 quarters in total) and referrals aggregated at the practice level. There are 27 primary care practices in the data set. Each can be associated with a municipality. Data on the 27 corresponding municipalities was collected from publicly available databases of Statistics Portugal (*Instituto Nacional de Estatística*, the Portuguese equivalent to England's Office for National Statistics). Table 6-2 presents a description of the variables collected and summary statistics. Distances and travelling times to the *Hospital do Espírito Santo de Évora*, *Hospital José Joaquim Fernandes de Beja* and closest private clinics for each practice were estimated using Google Maps™ (a routine to estimate the distances was developed in Matlab™, details provided in Appendix A). Unfortunately, unlike in the UK, in Portugal there is no publicly available information on illness ratios or index of clinical needs, at least not at the level of municipalities (or more to the point, primary care practices). As such, simple mortality ratios were collected. The data collected through Statistics Portugal are yearly data.

Finally, as is often the case with studies of health care utilisation, there is a significant number of zeros in the data set, both for referrals and non-completed appointments (see Figure 6-1). For example, Windmeijer et al. (2005) choose quarter as the time period because "*using monthly data would result in too many zero outpatient visit rates*" (page 985). Not only is the use of quarters in this study insufficient to deal with zeros, as Figure 6-1 illustrates, there is also considerable overdispersion. This is evident in the means and standard deviations of referrals and non-completed referrals in Table 6-2 (e.g., the variance for referrals is almost 10 times the value of the mean). These statistical characteristics have implications for estimating models of referrals and non-completion, which are discussed below.

Table 6-2 Variable description and summary statistics of practice/municipality data

Acronym	Description	Mean	SD
REFT	Referrals per quarter	11.7173	31.9854
NCT	Non-completed referrals per quarter	3.1030	8.6050
DIST*	Distance to hospital in Évora (in km)	72.8486	36.7627
TIME*	Travelling time to hospital in Évora (in min)	69.9380	29.1283
DISTB*	Distance to hospital in Beja (in km)	77.9196	35.4776
TIMEB*	Travelling time to hospital in Beja (in min)	74.3415	31.8429
DISTP*	Distance to closest private clinic (in km)	37.2060	20.1548
TIMEP*	Travelling time to closest private clinic (in min)	40.6606	18.6290
GPS	Number of GPs	8.3415	8.9733
FRAGE014	Fraction of population aged 14 or younger	12.7823	1.5925
FRAGE1524	Fraction of population aged 15 to 24	10.9366	1.0942
FRAGE2564	Fraction of population aged 25 to 64	51.2509	2.3114
FRAGE6574	Fraction of population aged 65 to 74	12.5130	1.6094
FRAGE75P	Fraction of population aged 75 or older	12.5173	2.2518
POP	Total population	10,993.52	10,739.81
FRFEMALE	Fraction of population that is female	0.5047	0.0180
MORT	Mortality rate (per 1000 inhabitants)	15.5236	3.4475
MORT50P	Mortality rate for over 50 year olds (per 1000)	14.8776	3.3969
PENSIONS	Average monthly value of pensions (in EUR)	3,586.78	434.14
PENSIONERS	Pensioners (per 1000)	455.37	59.88
SS	Beneficiaries of social inclusion (per 1000)	63.0870	28.5154
PHARMA	Number of pharmacies	5.3785	4.2127

* Distances and travelling times estimated using Google Maps™ (see Appendix A)

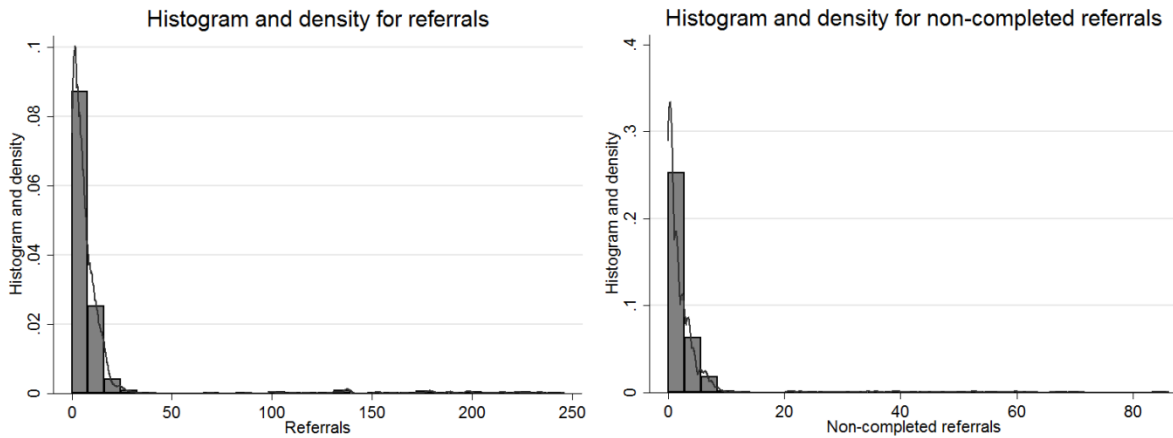


Figure 6-1 Histograms and density plots for referrals per quarter per practice

6.4 Estimation

As Figure 6-1 illustrates referrals and non-completed appointments are not normally distributed. As such, they do not meet the assumptions of parametric statistical tests (e.g., ANOVA, t-test, or linear regression) and assuming normality can give misleading results. A frequently used solution is to transform the data. Many variables in biology and ecology are log-normally distributed so that a log transformation will result in a normally distributed variable, besides potentially helping with outliers. Naturally, any transformation of the data will affect the final estimates obtained (Bland and Altman, 1996a). The first set of problems has to do with back-transformation of the results following estimation. The geometric mean (i.e., the back-transformed mean) will not be the same as the arithmetic mean (i.e., the actual sample mean), and back-transformation of some statistics (e.g., standard deviations and negative differences) is simply not possible (Bland and Altman, 1996b). The second problem is how to deal with zeros. The logarithm of zero is not defined so a second transformation is needed, usually adding an arbitrary value (such as 1) to all observations (O'Hara and Kotze, 2010). Even with this second transformation, if the number of zeros is large, then log-transforming the data will leave a large part of the sample with the same value, making it effectively impossible to achieve a normal distribution. Figure 6-2 shows a

number of transformations of the total referrals in the data set. It clearly shows that transforming the data will not lead to normality.

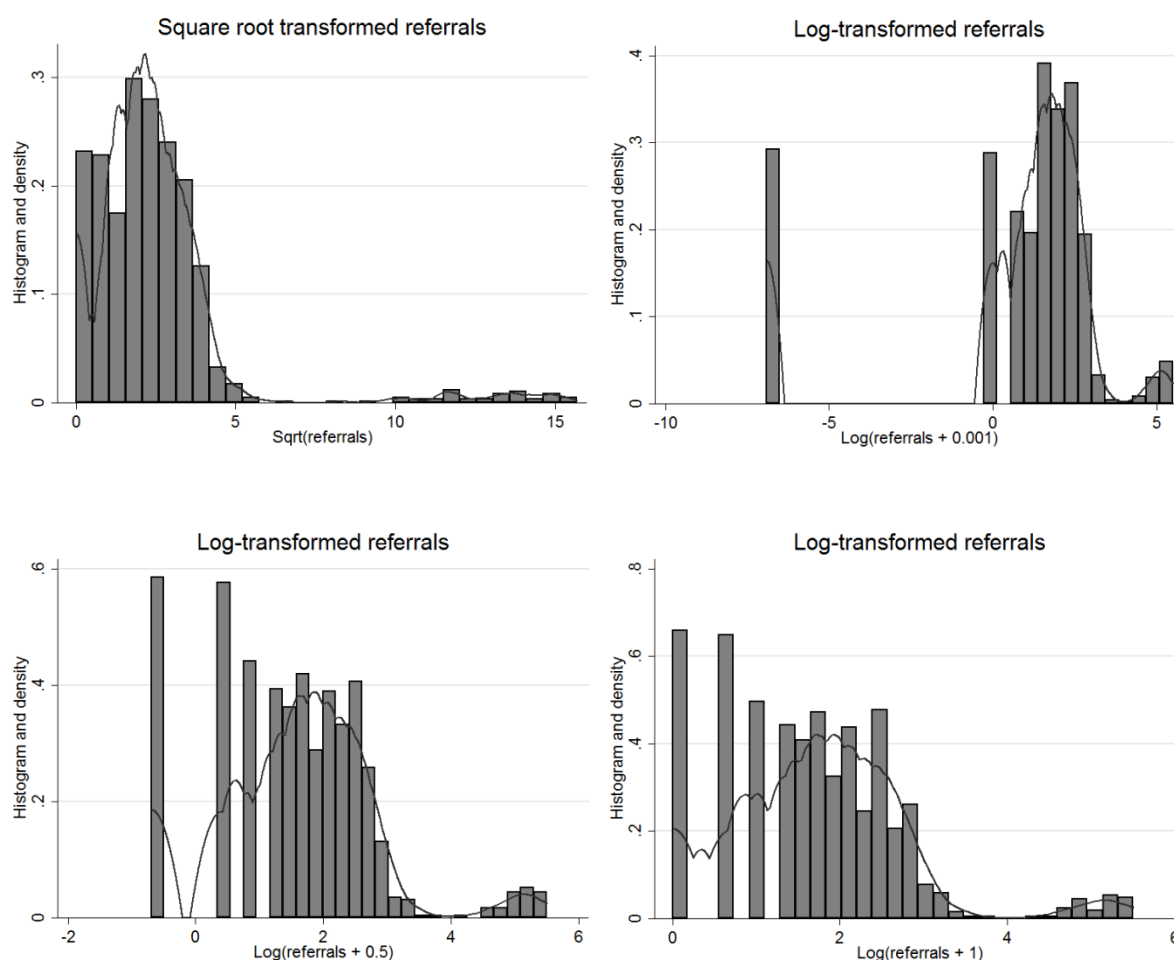


Figure 6-2 Histograms and density plots for transformations of total referrals

Deciding how to deal with the zeros will fundamentally affect the choice of estimation approach. One solution is to build the data set so that the zeros are eliminated. For example, Windmeijer et al. (2005) use quarters rather than months so as to reduce the number of zeros in the data set. Even then, as with the data set in this study, there may still be a non-negligible number of zeros in the data set. It then becomes important to understand the nature of the zeros: whether they are genuine or true zeros or whether they are sampling artefacts. For example, an individual income of zero may mean one of three things: 1) a non-observable response or missing value; 2) an income of zero for an individual in the labour

market; and 3) an indication that the individual is not in the labour market (non-participation). The latter two are true zeros.

First, the data set in this study includes all referrals and appointments between 2001 and 2013. Because there is no sampling (the data set is the actual population), zeros are genuine or true zeros. Second, the distinction between a qualitative and a quantitative zero does not apply to referrals or non-completion. Zeros do not denote a two-step process, whereby a practice first determines whether to refer anyone, and then how many people to refer. As such, two-part models – such as hurdle and Tobit – are not appropriate. The solution is to use generalised linear models, more specifically regression models of count data.

Generalised linear models are extensions to the standard linear regression model that incorporate various types of responses – such as counts and binary responses – thus obviating the need for data transformations that achieve linearity in the predictors and normality of the distribution (McCulloch, 2000). They are characterised by three components (Hilbe, 1994): 1) a random response component Y which is exponentially distributed; 2) a linear or systematic component that relates the linear predictor η with a set of explanatory variables X ; and 3) a link function $g(\cdot)$ describing how the mean expected response $E(Y)$ is related to the linear predictor η . Generalised linear models build on the linear regression model by allowing the use of distributions other than the normal. The most commonly used for count data are the Poisson and negative binomial distributions.

6.4.1 Poisson and negative binomial regression models

The Poisson distribution for a count variable y is:

$$P(y) = \frac{e^{-\mu} \mu^y}{y!} \quad (6-4)$$

A characteristic of the Poisson distribution is that the first two moments (i.e., the mean and the variance) are equal:

$$E[Y] = V[Y] = \mu \quad (6-5)$$

The usual Poisson regression model specifies for observation i :

$$E[y_i|x_i] = \mu_i = \exp(x_i'\beta') = \exp(\beta_1 + \beta_2\beta_{2i} + \dots + \beta_k x_{ki}), \quad i = 1, \dots, n \quad (6-6)$$

where x_{ki} are independent explanatory variables (chosen as in a linear regression model) and β_k the associated parameters. Because of (6-5) and (6-6), the Poisson model is intrinsically heteroskedastic (Cameron and Trivedi, 2007). Given the distribution function (6-4), the regression equation (6-6), and assuming that observations are independent, then maximum likelihood can be used to find the set of parameters β' that maximise:

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^N y_i \ln(\mu_i) - \mu_i - \ln(y_i!) = \sum_{i=1}^N x_i (y_i - \mu_i) = 0 \quad (6-7)$$

The interpretation of the coefficients β' is different from the standard linear model due to the exponentiation. For the Poisson model (and any other model with exponential conditional mean),

$$\frac{\partial E[y_i|x_i]}{\partial x_{ji}} = \beta_j \exp(x_i'\beta') \quad (6-8)$$

for the j^{th} regressor. If the regressor is log-transformed then β_j is an elasticity (Cameron and Trivedi, 2007). The Poisson model is restrictive due to its assumption of equidispersion (i.e., the mean equals the variance). As mentioned above, that assumption does not hold for referrals and non-completed appointments, in this data set. In both cases, the variance is many times the value of the mean, a feature called overdispersion. This can lead to incorrect standard errors in Poisson regression estimates. One solution is to introduce latent heterogeneity in the conditional mean of the Poisson model (Greene, 2008):

$$Var[y_i|x_i, \varepsilon_i] = \exp(x_i'\beta' + \varepsilon_i) = \exp(x_i'\beta') \exp(\varepsilon_i) = \mu_i h_i \quad (6-9)$$

Where h_i has a one parameter gamma distribution, $G(1, \alpha)$, with mean 1 and variance α . By integrating h_i out of the joint distribution, the negative binomial distribution is obtained:

$$f(y|\mu, \alpha) = \frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu+\alpha^{-1}}\right)^y, \alpha > 0 \quad (6-10)$$

The first two moments of the negative binomial distribution are:

$$E[y|\mu, \alpha] = \mu \quad (6-11)$$

$$V[y|\mu, \alpha] = \mu(1 + \alpha\mu) \quad (6-12)$$

It is clear that when α is zero, the negative binomial distribution becomes a Poisson distribution. Maximum likelihood can be used to estimate the parameters β' and α (Greene, 2008). Due to the quadratic term in (6-12), the negative binomial distribution in (6-10) is usually referred to as the NB2 distribution. Both Poisson and negative binomial are appropriate when zero counts come from the same process as positive counts, which is the case in the data in this study.

6.4.2 Panel models for count data

Since the data set in this study is a panel of practices over time, the Poisson and negative binomial models must be extended to account for issues like practice-specific effects, endogeneity and dynamics. Standard linear regression methods for short panels can be extended to count models. A basic linear regression panel model can be defined as:

$$y_{it} = \alpha + x'_{it}\beta' + z'_i\gamma' + \mu_i + \eta_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T \quad (6-13)$$

where y_{it} is the dependent variable of interest in practice i in quarter t ; x'_{it} is a vector of time-varying explanatory variables and β' the associated coefficients; z'_i is a vector of time invariant independent variables and γ' the associated coefficients; μ_i represents unobserved time invariant practice effects; and η_{it} represents time-varying idiosyncratic random errors assumed to be independent and identically distributed (IID) across practices and time, with

mean 0 and variance σ^2 . If one assumes μ_i to represent N fixed unknown parameters (one for each practice), then the model is a fixed effects model, and can be estimated simply by specifying $N - 1$ practice dummy variables. Alternatively, one may consider that μ_i are draws from a distribution with mean μ and variance σ^2 , leading to a random effects model. In the fixed effects model, the error term reduces to η_{it} . In the random effects, μ_i and η_{it} are assumed to be uncorrelated. The fixed effects model has two disadvantages: 1) it increases the number of parameters being estimated (as many as the number of practices minus one); and 2) it renders γ' – the parameters of time invariant z'_i variables – unestimable, as they are captured in the practice dummies (i.e., they are perfectly collinear). The random effects model is not without flaws: 1) it assumes that μ_i and x'_{it} are uncorrelated, which if not true can lead to biased estimators. Much of this can be extended to count data panel models.

In general, individual effects (i.e., practice-specific effects in this case) are included in count models with multiplicative effects (for simplicity, I use the Poisson model in the following equations):

$$E[y_{it}|\alpha_i, x_{it}] = \alpha_i \times \exp(x'_{it}\beta') = \exp(\ln(\alpha_i) + x'_{it}\beta'), \quad i = 1, \dots, n, \quad t = 1, \dots, T \quad (6-14)$$

The intercept is merged into α_i so that the regressors x'_{it} no longer include an intercept (Cameron and Trivedi, 2013). The model can then be expressed as

$$\mu_{it} \equiv \exp(\ln(\alpha_i) + x'_{it}\beta') = \exp(\delta_i + x'_{it}\beta') \quad (6-15)$$

In the random effects count model, the individual effect α_i (or δ_i) is assumed to be distributed independently of the regressors x'_{it} . The joint density for the i^{th} observation, conditional on the regressors is (Cameron and Trivedi, 2013):

$$f(y_i|X_i) = \int_0^\infty [\prod_{t=1}^T f(y_{it}|\alpha_i, x'_{it})] g(\alpha_i|\eta) d\alpha_i \quad (6-16)$$

where $g(\alpha_i, \eta)$ is the specified density of α_i and $f[y_{it}|\alpha_i, x'_{it}]$ is the density for the i^{th} observation conditional on both α_i and the regressors. The Poisson random effects model is

obtained by assuming y_{it} is Poisson distributed, and that α_i is gamma distributed. Integrating out α_i , there is a closed form solution to (6-16). The negative binomial random effects model can be obtained by assuming y_{it} is negative binomial NB2 IID and that α_i is lognormal, equivalent to assuming that δ_i is normally distributed. In this case, there is no closed form solution to (6-16), and Gaussian quadrature or simulated maximum likelihood have to be used.

In the fixed effects model, the individual effect α_i (or δ_i) is potentially correlated with the regressors x'_{it} , and hence must be eliminated before estimating β' . A quasi-differencing approach, similar to mean-differencing used in linear fixed effects models, can be used under the strong assumption that regressors x'_{it} are strictly exogenous, after conditioning for α_i . Given this assumption, the coefficients β' can be estimated by the method of moments estimator (Cameron and Trivedi, 2013). An alternative is to use dummy variables for practices, but this can be computationally inconvenient. A third approach is to use quasi-differencing to eliminate the individual effect, in a way similar to first-differences in linear regression. However, in order to do so, one must assume weak exogeneity of regressors so that conditioning is on past and current values of regressors:

$$E[y_{it}|X_i^{(t)}] = E[y_{it}|x_{it}, \dots, x_{i1}] = \alpha_i \lambda_{it} \quad (6-17)$$

Unlike linear and Poisson fixed effects models, the negative binomial 1 (NB1) fixed effects model (proposed by Hausman et al. (1984)) allows the estimation of coefficients of time-invariant regressors. An alternative to NB1 fixed effects model is to estimate an NB2 model with practice dummy variables, although this can lead to inconsistent estimations of β' due to the incidental parameters problem. The choice of fixed effects or random effects is important, since random effects estimators are inconsistent when the individual effect α_i is correlated with the regressors x'_{it} .

6.4.2.1 Dynamic panel models for count data

The practice-level effect α_i induces dependence over time in y_{it} (i.e., unobserved heterogeneity). Another way to do so is to make y_{it} depend directly on lagged values of y_{it} (i.e., true state dependence). With panel data, it is possible to account for both dynamics and practice-specific effects. In such a model, y_{it} is a function of y_{it-1} and the regressors x'_{it} . For the random effects dynamic model, the log-likelihood will depend on the initial condition y_{i0} , which will be correlated with the random effect α_i . As such, estimates will not be consistent for short panels, but inconsistency will decline as T increases and eventually become asymptotically irrelevant as $T \rightarrow \infty$ (Cameron and Trivedi, 2005). One approach is to use the conditionally correlated random effects model, which assumes the individual effect α_i is correlated with x'_{it} through the group means $\bar{x}'_i$ (similar to the approach proposed by Mundlak (1978) for linear regression):

$$\alpha_i = \exp(\bar{x}'_i \lambda + \varepsilon_i) \quad (6-18)$$

and ε_i represents unobserved heterogeneity that is uncorrelated with the regressors (and thus holding the assumption inherent to a random effects model). For a dynamic random effects count model, the individual effect can be modelled to include the initial condition:

$$\alpha_i = \exp(\delta_0 y_{i0} + \bar{x}'_i \lambda + \varepsilon_i) \quad (6-19)$$

As for dynamic fixed effects models, strict exogeneity can no longer be assumed since it rules out predetermined regressors (a regressor is predetermined when it is correlated with past shocks, but not with current and future shocks). However, making the assumption of weak exogeneity, as done before in first-differencing, one can include lagged values of the dependent variable as regressors:

$$E[y_{it} | X_i^{(t)}, Y_i^{(t-1)}] = E[y_{it} | x_{it}, \dots, x_{i1}, y_{i,t-1}, \dots, y_{i1}] = \alpha_i \lambda_{it} \quad (6-20)$$

The Poisson fixed effects estimator is not consistent if regressors are predetermined. Instead, generalised methods of moments estimation (GMM) is used to eliminate α_i under the assumption (6-20). Both Chamberlain (1992) and Wooldridge (1997) have proposed transformations $q_{it}(\theta)$ to eliminate α_i . For either one, it can be shown that, given the assumption of weak exogeneity:

$$E[q_{it}(\theta)|z_{it}] = 0 \quad (6-21)$$

where z_{it} can be drawn from $x_{i,t-1}, x_{i,t-2}, \dots$ and, if lags up to $y_{i,t-p}$ (frequently $p = 1$) appear as regressors, z_{it} can also be drawn from $y_{i,t-p-1}, y_{i,t-p-2}, \dots$ (Cameron and Trivedi, 2013). In two-step GMM estimation, model adequacy can be tested using an over-identifying restrictions test (when, as often is the case, there are more instruments z_{it} than regressors).

Finally, GMM methods can also be adapted to estimate fixed effects models with endogenous regressors. In the context of this study, waiting time is a function of referrals, which in turn are a function of waiting time. This loop will cause the standard fixed effects estimator to be inconsistent. As in the linear case, panel GMM can be used to deal with endogenous regressors.

6.4.3 Models estimated

Given the discussion above, the following models were estimated, both for referrals and non-completed appointments: a static NB2 random effects (NB2-RE); a static NB1 fixed effects (NB1-FE); and a dynamic GMM Poisson fixed effects (Poisson-GMM). Different models were estimated to test the impact of different assumptions on estimators, namely the importance of unobserved heterogeneity and state dependence.

As Windmeijer et al. (2005), practices were assumed to be waiting time takers. In other words, no practice could affect the waiting time through its referral or non-completion decisions, and so there is no simultaneous equation bias. I also assumed, as Gravelle et al. (2002), that expectations are myopic: in period t , patients and GPs base their expectations

of waiting time on what they observed in period $t - 1$. Dependent variables in the model of referrals included (the prefix “L” indicates the variable is log-transformed, and the suffix “_t” indicates the variable is lagged t quarters): LWT_1 (the lagged value of the waiting time, log-transformed), LTIME, LTIMEB, LTIMEP, LGPS, FRFEMALE, LMORT50P, FRAGE014, FRAGE1524, FRAGE6574, FRAGE75P, LPENSIONS, LPENSIONERS, LSS, LPHARMA, REG1 (a dummy variable indicating practices belonging to Alentejo health sub-region *Alentejo Central I*), and REG2 (a dummy variable indicating practices belonging to Alentejo health sub-region *Alentejo Central II*, base category is region *Baixo Alentejo*). Descriptive statistics for many of these variables are shown in Table 6-2. The dependent variables in the non-completion model included all of the above except that, instead of the lagged value of the waiting time, the present waiting time LWT was used.

The instrument sets used for the dynamic GMM Poisson models comprised the explanatory variables listed above, together with the lagged values of the dependent variables (REFT_1 and NCT_1) and the waiting time (LWT_1 and LWT_2 for the referral model, LWT_1 for the non-completion model). The validity of the instrument sets was tested using Hansen’s J statistic (Baum et al., 2003). Models were estimated using Stata 12.

6.5 Results

The results of static panel negative binomial regression models of referrals and non-completion rates are shown in Table 6-3 and Table 6-5 respectively (results for dynamic GMM models of referrals and non-completion rates are shown in Table 6-4 and Table 6-6 respectively).

Table 6-3 Static panel negative binomial regression models of REFT

Variables	NB2-RE		NB1-FE	
	<i>Coefficient</i>	<i>Standard error</i>	<i>Coefficient</i>	<i>Standard error</i>
CONST	13.88	(24.00)	14.97	(26.37)
LWT_1	-0.340***	(0.0800)	-0.330***	(0.0860)
LTIME	-1.381	(2.630)	-2.123	(6.980)
LTIMEB	-0.871	(2.890)	-1.406	(4.918)
LTIMEP	-0.182	(2.794)	0.168	(4.208)
LGPS	-0.822**	(0.401)	-0.869*	(0.445)
FRFEMALE	18.73	(20.71)	20.23	(21.82)
LMORT50P	0.220	(0.688)	0.204	(0.641)
FRAGE014	-0.0917	(0.167)	-0.0650	(0.175)
FRAGE1524	-0.353**	(0.161)	-0.353**	(0.170)
FRAGE6574	-0.00522	(0.127)	0.0264	(0.131)
FRAGE75P	-0.0752	(0.138)	-0.0513	(0.153)
LPENSIONS	-5.921	(3.793)	-5.596	(3.727)
LPENSIONERS	-1.535	(4.351)	-2.386	(4.388)
LSS	3.796**	(1.681)	3.952**	(1.656)
LPHARMA	-0.472	(1.202)	-0.648	(1.596)
REG1	0.466	(1.646)	0.415	(3.352)
REG2	0.538	(1.676)	0.121	(3.218)
<i>Observations</i>	<i>1080</i>		<i>1080</i>	
<i>Log likelihood</i>	<i>-2807.7</i>		<i>-2628.3</i>	
<i>AIC</i>	<i>5655.5</i>		<i>5292.6</i>	
<i>BIC</i>	<i>5755.2</i>		<i>5382.3</i>	

***significant at 1% level; **significant at 5% level; * significant at 10%. Dependent variable is total referrals per practice per quarter per 1,000 inhabitants. Standard errors were estimated using bootstrapping (50 repetitions, seed = 123). Prefix *L* indicates variable is log-transformed.

In table Table 6-3, the dependent variable is the total number of referrals per practice per quarter per 1,000 inhabitants. As illustrated by the log likelihood, the Akaike information criteria (AIC) and the Bayesian information criteria (BIC), the NB1-FE model provides the best fit. Estimates of the elasticity of referrals with respect to the waiting time, lagged 1 quarter, are -0.33 and -0.34 (because the waiting time variable is log-transformed, the

coefficients can be interpreted directly as elasticities). Both are statistically significant. Other statistically significant predictors of referrals include the number of GPs in the practice, the fraction of the population which is between 15 and 24 years old (the base category for age is 25 to 64), and the proportion of social inclusion beneficiaries. The coefficients suggest that (*ceteris paribus*): the more GPs there are in a given practice the smaller the number of referrals to secondary outpatient dermatology appointments; patients between 15 and 24 are less likely to enter the waiting list than those aged 25 to 64; and that the bigger the proportion of social inclusion beneficiaries in the population the bigger the number of referrals.

The results of the dynamic Poisson-GMM model are presented in Table 6-4. Also shown are the results of extensions of the NB2-RE and NB1-FE models which incorporate lagged referrals. It is important to explore dynamics because REFT is highly and significantly correlated with its lagged value. As previously discussed, panel data allows the simultaneous analysis of unobserved heterogeneity (through the inclusion of a practice-level effect α_i) and true state dependence (through the inclusion of lagged values of y_{it}). As Windmeijer et al. (2005) points out, there are two explanations for persistent referral counts: practice-specific fixed effects which cause some practices to always have higher referrals (i.e., unobserved heterogeneity), or an adjustment towards a long run equilibrium as GPs and patients collect information on, and react to, changing waiting times (i.e., true state dependence). As the results in Table 6-4 illustrate, the coefficient on the lagged referral count is small and similar to that in the NB1-FE model. This would suggest that practice-specific fixed effects provide the dominant explanation for persistence in the referral count. As confirmed by Hansen's J statistic, the instrument set used in the Poisson-GMM model is valid (at a 10% level).

Table 6-4 Dynamic panel count models of REFT

Variables	NB2-RE	NB1-FE	Poisson-GMM
REFT_1	0.00359 (0.0193)	0.00358 (0.0183)	0.00654*** (0.00200)
LWT_1	-0.266*** (0.0667)	-0.261*** (0.0821)	-0.226*** (0.0524)
<i>Observations</i>	1080	1080	1053
<i>GP practices</i>	27	27	27
<i>Time period</i>	2002q2-2012q4	2002q2-2012q4	2002q3-2012q4
<i>Log likelihood</i>	-2798.9	-2620.2	-
<i>Hansen's J statistic</i>	-	-	2.8469, <i>p</i> -value=0.0916

***significant at 1% level; **significant at 5% level; * significant at 10%. Dependent variable is total referrals per practice per quarter per 1,000 inhabitants. Negative binomial models use bootstrapping (50 repetitions, seed = 123), Poisson-GMM uses cluster-robust estimation. Prefix *L* indicates variable is log-transformed. Other independent variables included as in Table 6-3.

The results of the static negative binomial regression models of non-completion are shown in Table 6-5. The dependent variable is the total number of non-completed referrals (i.e., missed appointments) per practice per quarter per 1,000 inhabitants. As with the model for referrals, the log likelihood, AIC and BIC all indicate that NB1-FE provides the best fit. Estimates of the elasticity of non-completion with respect to the waiting time are 0.610 and 0.633 (as with referrals, the log-transformation means the coefficients can be interpreted as elasticities). Both are statistically significant and positive, as expected. The only other significant variable is the average value of pensions, and only in the NB2-RE model.

Table 6-5 Static panel negative binomial regression models of NCT

Variables	NB2-RE		NB1-FE	
	<i>Coefficient</i>	<i>Standard error</i>	<i>Coefficient</i>	<i>Standard error</i>
CONST	2.737	(27.55)	-16.68	(38.34)
LWT	0.610***	(0.131)	0.633***	(0.131)
LTIME	0.550	(2.010)	2.946	(9.596)
LTIMEB	1.182	(2.091)	2.213	(10.13)
LTIMEP	-0.119	(1.672)	1.506	(7.446)
LGPS	0.211	(0.513)	0.375	(0.530)
FRFEMALE	5.306	(20.50)	21.31	(29.08)
LMORT50P	0.0398	(0.875)	0.00230	(0.765)
FRAGE014	0.118	(0.175)	0.121	(0.268)
FRAGE1524	-0.0934	(0.175)	-0.0629	(0.228)
FRAGE6574	0.0312	(0.131)	0.0224	(0.189)
FRAGE75P	-0.139	(0.118)	-0.225	(0.229)
LPENSIONS	-7.942*	(4.479)	-5.851	(5.047)
LPENSIONERS	2.899	(4.607)	1.083	(5.424)
LSS	0.812	(1.717)	1.173	(1.863)
LPHARMA	-0.234	(0.775)	0.0449	(1.482)
REG1	-0.210	(1.191)	-0.615	(5.389)
REG2	0.770	(1.007)	1.157	(4.966)
<i>Observations</i>	<i>1107</i>		<i>1107</i>	
<i>Log likelihood</i>	<i>-1791.8</i>		<i>-1649.7</i>	
<i>AIC</i>	<i>3623.7</i>		<i>3335.5</i>	
<i>BIC</i>	<i>3723.9</i>		<i>3425.6</i>	

***significant at 1% level; **significant at 5% level; * significant at 10%. Dependent variable is total non-completed referrals per practice per quarter per 1,000 inhabitants. Standard errors were estimated using bootstrapping (50 repetitions, seed = 123). Prefix *L* indicates variable is log-transformed.

As with referrals, the non-completion rate is highly and significantly correlated with its lagged value. The results of the dynamic NB2-RE, NB1-FE and Poisson-GMM models shown in Table 6-6 suggest that practice-specific fixed effects provide the dominant explanation for persistence in the non-completion count. As confirmed by the highly insignificant Hansen's *J* statistic, the instrument set used in the Poisson-GMM model is valid.

Table 6-6 Dynamic panel count models of NCT

Variables	NB2-RE	NB1-FE	Poisson-GMM
NCT_1	0.000593 (0.0401)	0.0000534 (0.0334)	0.00540 (0.00381)
LWT	0.581*** (0.105)	0.602*** (0.145)	0.442*** (0.0938)
<i>Observations</i>	1080	1080	1080
<i>GP practices</i>	27	27	27
<i>Time period</i>	2002q2-2012q4	2002q2-2012q4	2002q2-2012q4
<i>Log likelihood</i>	-1745.5	-1603.8	-
<i>Hansen's J statistic</i>	-	-	0.0119, <i>p</i> -value=0.9130

***significant at 1% level; **significant at 5% level; * significant at 10%. Dependent variable is total non-completed referrals per practice per quarter per 1,000 inhabitants. Negative binomial models use bootstrapping (50 repetitions, seed = 123), Poisson-GMM uses cluster-robust estimation. Other independent variables included as in Table 6-5.

6.5.1 Model prediction

It has been shown that the elasticity of demand with respect to waiting time can vary as a function of the initial waiting time (van Ackere and Smith, 1999). Building on that finding, I estimated the elasticities of referrals and non-completion as a function of the waiting time, using population-averaged negative binomial models, equivalent to those estimated in the previous section but assuming effects are not practice-specific. The results are illustrated in Figure 6-3 and Figure 6-4 (95% confidence intervals are shown as well). The elasticity of referrals becomes more negative as the waiting time increases, up to a point where it plateaus. The elasticity of non-completion becomes more positive as waiting time increases, initially slowly but very rapidly after 365 days. The range of values that waiting time takes in Figure 6-3 and Figure 6-4 goes up to two years, which is deemed realistic as 13 percent of the patients in this data set had waiting times longer than a year.

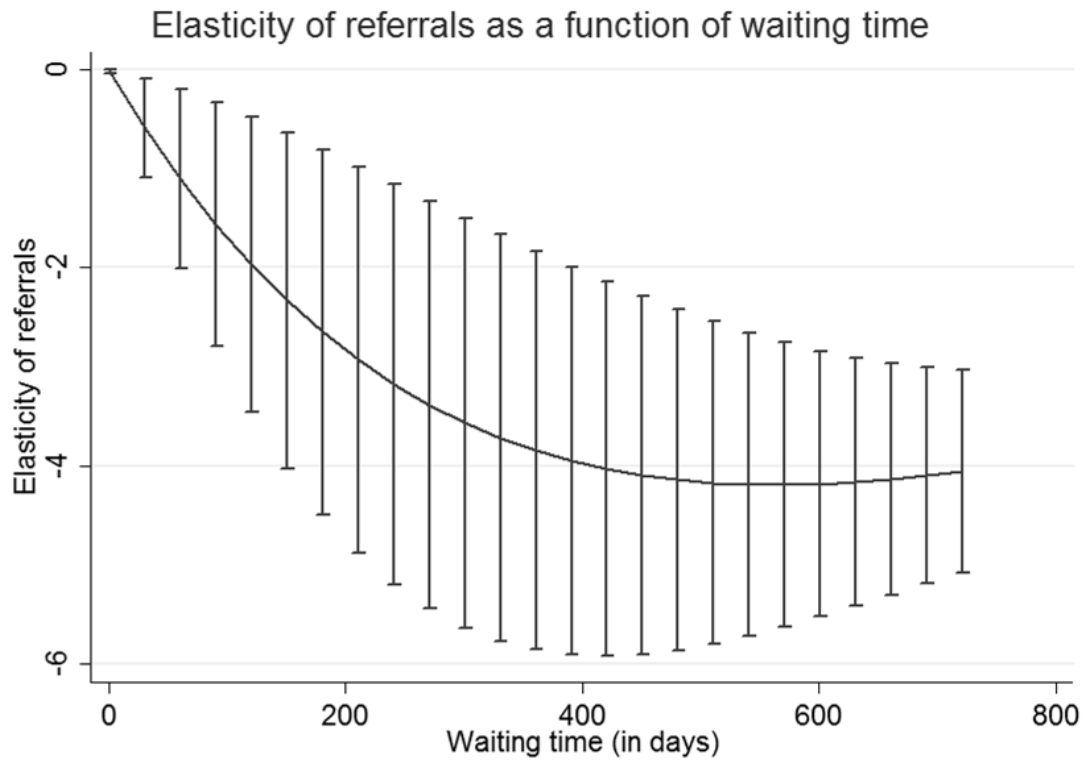


Figure 6-3 Elasticity of referrals with respect to waiting time as a function of waiting time (estimated using a population-averaged negative binomial model)

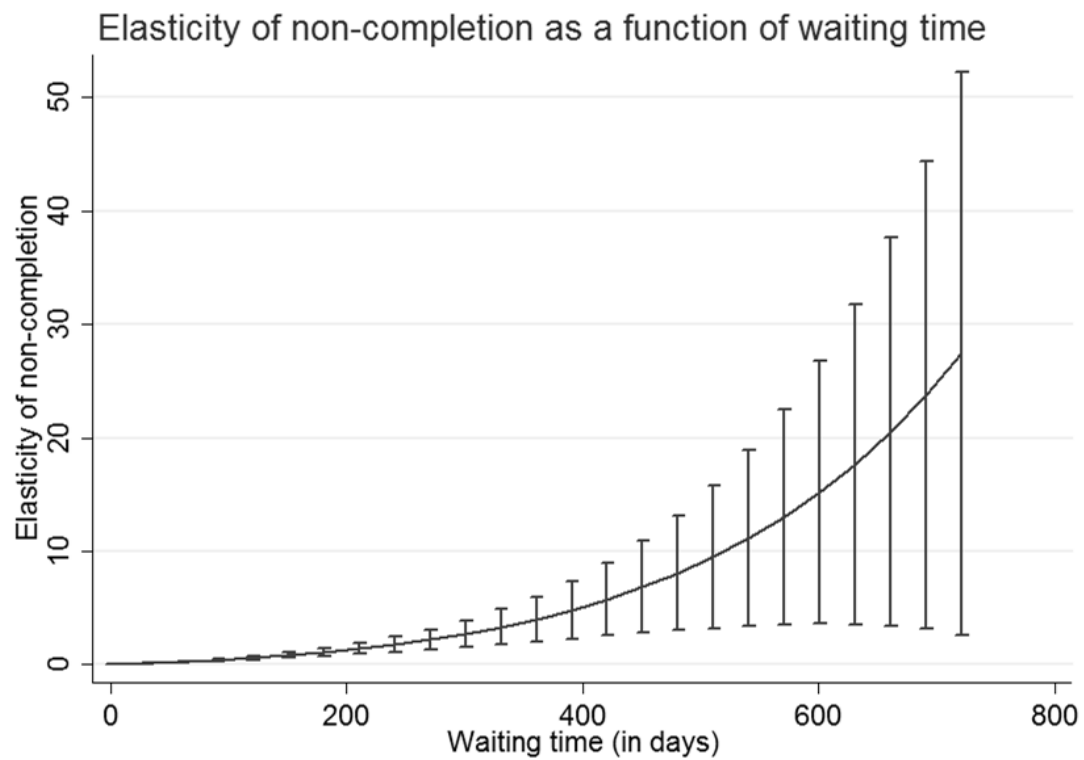


Figure 6-4 Elasticity of non-completion with respect to waiting time as a function of waiting time (estimated using a population-averaged negative binomial model)

6.6 Discussion

There have been significant developments in the application of count models to panel data, especially with regards to dealing with unobserved heterogeneity and state dependence (Trivedi, 2010). Given these advances it is unreasonable to use data transformations to achieve normality and apply the linear regression model (O'Hara and Kotze, 2010, Osgood, 2000). Furthermore, transforming the data may not be an option, as is the case here.

The elasticity of referrals with respect to waiting time was found to be -0.33, in line with previously reported elasticities using linear models (Windmeijer et al., 2005, Martin and Smith, 2003). A closer look at how the elasticity varies with waiting time showed that it was very close to zero for short waiting times and decreased steadily as waiting time increased, eventually plateauing around -4, a value comparable to that estimated by van Ackere and Smith (1999). The results further suggested that the higher the number of GPs in the practice the lower the referral rates, that patients between 15 and 24 were less likely to enter the waiting list than those aged 25 to 64, and that the bigger the proportion of social inclusion beneficiaries in the population the bigger the number of referrals. The elasticity of non-completion with respect to waiting time was statistically significant and found to be 0.633. There is no previously reported elasticity to compare it with.

Elasticities estimated through count models should be interpreted with care, since dependent variables are concentrated on a few values and do not vary continuously (Cameron and Trivedi, 2005). Notwithstanding, the impact of waiting time was statistically significant and had the expected effect in both models (referrals and non-completion). The negative binomial fixed effects model provided the best fit (as measured by the log likelihood, AIC and BIC). Results from dynamic Poisson-GMM models indicated that persistence in the referral and non-completion counts was more likely to be explained by unobserved heterogeneity rather than true state dependence.

6.6.1 Limitations and future work

The fact that other independent variables had low explanatory power may be related to the way the data were collected. Practice-level characteristics such as number of GPs, population by age groups, average value of pensions, number of pensioners per 1000 inhabitants, mortality rates, beneficiaries of social inclusion, and number of pharmacies, were all taken from Statistics Portugal, which collects the data on an annual basis. At the quarterly level, variation is thus limited. Furthermore, distance and time measures do not change at all, remaining constant for the study period. A second possibility is that the covariates included here are not the most adequate for explaining variation in demand for dermatology outpatient appointments. Previous studies have included measures of need such as an acute sector index (Martin and Smith, 1999) and standardised illness and mortality ratios (Windmeijer et al., 2005). It could be that these measures are inappropriate in the specific context of dermatology. It would be interesting to explore this in more detail in the future, by using some measure of dermatological need. This would require a better understanding of what affects the need for dermatology outpatient appointments.

6.7 Conclusion

To my knowledge, this is the first study using actual referrals, rather than appointments, to estimate an elasticity of demand with respect to waiting time. It is also the first study to report an elasticity of non-completion with respect to waiting time. The fact that the elasticity of demand is of the same order of magnitude of elasticities estimated in previous studies is an indication that panel models of count data can be used in this context. While there are no previous studies to compare the elasticity of non-completion with, the estimates are positive, as expected. Exploring both elasticities in more detail, the behaviours shown in Figure 6-3 and Figure 6-4 appear reasonable. More specifically, the elasticity of demand exhibits a relationship to waiting time in line with that estimated by van Ackere and Smith (1999). The elasticity of non-completion increases sharply as waiting time increases, indicating that

increases in waiting time at longer initial waiting times lead to higher non-completion rates than at lower initial waiting times, a finding which was expected. In other words, at longer waiting times, more people will seek alternative care, get better, get worse, etc., than at lower waiting times. These findings add to the face validity of the models, even given limitations with data.

6.7.1 Implications for the evaluation of teleconsultation in Alentejo

This study confirms findings from previous reports that waiting times affect the number of referrals to secondary care and the number of patients completing these referrals (i.e., actually attending appointments). As mentioned in the last two chapters, teleconsultations can lead to considerable reductions in waiting times. Furthermore, elasticities of referrals and non-completion change as a function of waiting times, as seen in Figure 6-3 and Figure 6-4. Given these results, changes in waiting times following the introduction of teleconsultations should lead to changes in referrals and non-completion. It is both unrealistic to assume that behaviour will remain static and unnecessary to enclose it inside a black box. This study of the impact of waiting time on referrals and non-completion provides quantitative estimates of two relevant causal mechanisms that should not be neglected by evaluations of teleconsultation. In Chapter 8, these estimates are integrated into a simulation model and their implications explored and tested.

Chapter 7

THE DETERMINANTS OF GENERAL PRACTITIONERS' REFERRALS

As discussed in Chapter 2, the decision to refer a patient to specialist care is a shared one, involving GPs and patients and families. While determined to a great extent by clinical need, decisions to refer are also influenced by numerous non-clinical factors. Research into the determinants of referrals has been prolific. However, despite evidence presented in Chapters 4 and 5 that GPs who participate in teleconsultation have lower referral rates, previous studies of the phenomenon have relied chiefly on interviews and expert opinion. Furthermore, it is unclear how the determinants investigated in previous studies affect GPs and patients separately. As GPs get involved in specialist care, through their participation in teleconsultations, it is important to explore how their decisions to refer patients evolve.

In this chapter, I undertake a study of GPs' referral decisions using stated preference data collected through a DCE. The study makes contributions to the literatures on the determinants of referrals, and the evaluation of teleconsultation in general, and to the assessment of teleconsultation in Alentejo. I discuss these at the end of the chapter.

7.1 Background

In Portugal, as in many other health care systems throughout the world, patients seeking an appointment with a specialist must first be seen by a primary care physician, or GP, who assesses their motivation for a specialist appointment. In such systems, GPs act as gatekeepers. Gatekeeping has been associated with several benefits (Scott, 2000). Secondary care is expensive compared to primary care. As agents for their patients, GPs are better prepared to judge when specialist care is needed and when adequate care can be provided in the primary setting, avoiding higher costs (Newton et al., 1991). Furthermore, because of GPs' aggregate experience from many patients and referrals, they are also potentially better at advising patients on the quality of secondary providers (although in certain health systems in which referral networks are mandated, such as in Portugal, this is likely less important). Finally, GPs establish long term relationships with their patients and see them more often than a specialist would, potentially leading to a better understanding of their patients' needs and preferences (Scott, 2000). Through their role as agents for their patients, and their power to refer them to specialist services, GPs are not only suppliers but they are also part of the demand for health care.

Since the early 1960s it has been evident there is high variation in the rates of referrals of different GPs (Newton et al., 1991). Referrals have been the focus of much debate and analysis since then (Wilkin and Smith, 1987, O'Donnell, 2000, Mehrotra et al., 2011). Initially it was believed this variation merely reflected differences in case mix and clinical need. The fact that variation persisted after controlling for clinical determinants raised concerns over the quality of care being provided, and researchers began exploring the impact of non-clinical factors. Understanding what affects the referral decision is important for various reasons. First and foremost, it is an objective of health care systems that patients receive the same standard of care, irrespective of location, social status and other factors. Large variation in referral rates, which persists after adjusting for clinical factors, may indicate this

is not happening. Second, referrals trigger entry into more expensive secondary care. Large differences in rates of referral, after adjusting for clinical need, mean that patients with similar complaints can receive treatment at substantially different levels of cost. Finally, as patients demand an increasingly active role in managing their care, it is important to understand how much variation in referrals can be attributed to patient preferences and expectations.

As shown in the literature review in Chapter 2, research around referrals has been prolific. It has evolved from merely reporting referral rates for different GPs, practices or specialties (Noone et al., 1989) to exploring the reasons behind referral decisions (Newton et al., 1991) and determining whether referrals were appropriate and timely (Fertig et al., 1993). In the context of evaluating a complex health intervention (and specifically teleconsultations), the primary focus is on the reasons behind referral decisions. Figure 7-1 summarises the determinants identified in the literature and previously discussed in more detail in Section 2.2.2. They are grouped into four categories: patient characteristics, GP characteristics, practice characteristics, and secondary care factors. Given the number and diversity of factors identified, it seems unrealistic to simply put them into black boxes and disregard them. As the literature on teleconsultations shows, the intervention affects many of the determinants in Figure 7-1, namely: waiting time for and distance to secondary care, GP uncertainty, type and degree of specialisation, patient expectations, and quality of communication between GP and specialist. It is thus important to explore whether these determinants are indeed important, and how important they are. Previous studies have taken considerable steps further in increasing our understanding of GPs' referral decisions, but there are limitations.

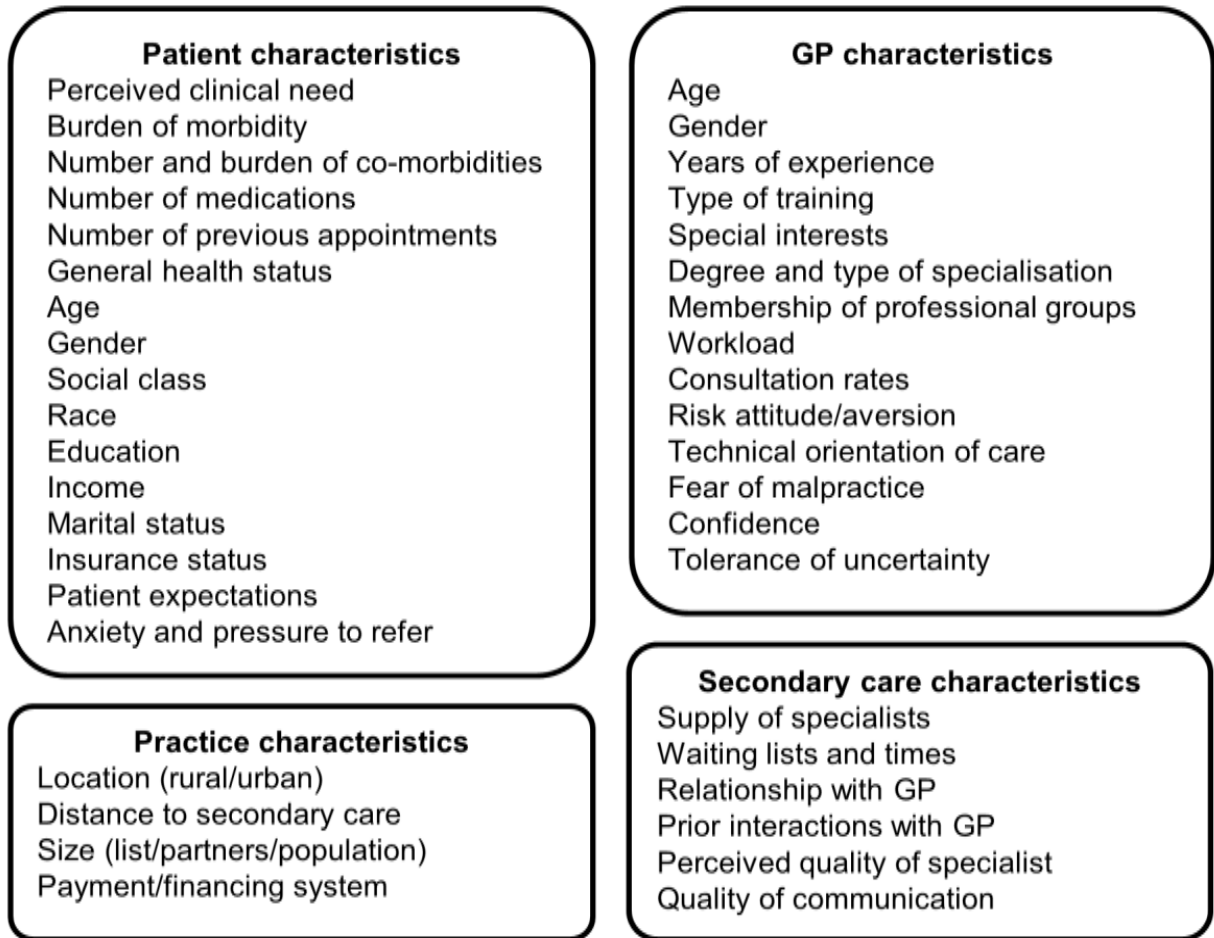


Figure 7-1 Determinants of GPs' referrals identified in the literature

The quality and usefulness of previous research are highly variable, with different studies reporting opposite effects for the same factor (e.g., patient or GP gender). One reason for this may be the choice of method. Some studies use simple pairwise correlations to assess effect, while others model the effect of all factors simultaneously, resulting in very different findings (Carlsen et al. (2008) provide a good example). Another reason might be that a certain factor interacts with other variables, or affects different GPs in different ways (i.e., there is GP heterogeneity). In her review of 91 referral studies, O'Donnell (2000) identified a number of limitations: small number of referrals, data on short period of time, difficulty in controlling for random variation rather than variation attributable to preferences, and problems collecting additional data (e.g., on age and gender). These limitations are typical of studies using revealed preference data. In this chapter, I take a different approach

to the study of the determinants of GPs' referrals, exploring not revealed preference data but stated preference data using a DCE.

7.2 Methods

A DCE is an attribute-based survey method based on the assumption that a good or service can be described by a set of characteristics or attributes, and the extent to which individuals value that good or service is determined by the nature and levels of the characteristics (Ryan and Gerard, 2003). The approach has been used extensively to identify patient preferences regarding health care services. The use of DCEs in eliciting physicians' preferences has been significantly more limited, and has focused mostly on treatment choices (Mantovani et al., 2005) and job preferences (Ubach et al., 2003, Lagarde and Blaauw, 2009). To my knowledge, there has been no application of DCEs in the study of GPs' referral decisions.

As previously discussed, stated preference choice methods can be useful when revealed preference data is not available or difficult to collect. This is usually the case with referral data. Unless an experimental study is conducted, data are normally collected for managerial and audit purposes, not for assessing the reasons for – or the appropriateness of – referrals. Although stated choice data may suffer from self-report bias, DCEs afford a number of advantages. First, all participants are presented with the same group of clinical complaints. Differences in their decisions will thus be due to differences in their judgements, not in the information provided about the patients. Second, collecting data through a DCE can be significantly cheaper than collecting revealed choice data. Third, stated preference methods can be used to disentangle the effects of different attributes and covariates, something that is not trivial with revealed preference data. For example, in a revealed data set it is difficult to explore the impact of not having insurance, low income or poor general health individually as all these factors tend to be highly correlated. Fourth, data from DCEs can be analysed using a variety of econometric models, testing various assumptions concerning how factors influence the decision to refer. Fifth, DCEs provide information on

non-demanders, which is not usually available in revealed preference data. Finally, and perhaps most significantly, DCEs can be used to disaggregate patients and GPs' preferences, something that is usually not possible with revealed preference data found in observational studies.

7.2.1 Development of attributes and levels

In the literature review in Chapter 2, I discussed a significant number of determinants of referral decisions identified in the literature. Figure 7-1 shows there is a comprehensive list of potentially relevant attributes for inclusion in the DCE. First and foremost amongst the reasons for referring a patient to a specialist appointment is the GP's assessment of clinical need. Four dermatological cases, which are essentially levels of the attribute *need*, were selected from a database of actual GP referrals and developed with the help of a senior dermatologist at the *Hospital do Espírito Santo de Évora*. Each of the four clinical presentations included in the study was described in terms of size, shape, colour and evolution (a common practice in dermatology). The language and terms used were based on actual referrals made by GPs in Alentejo. The textual descriptions were supplemented by photographs of the lesions.

Dermatology was selected given the interest in evaluating teledermatology. However, there are other reasons. First, a significant percentage of all referrals to outpatient appointments are for dermatology, and variation in referral rates has been shown to exist (Chen, 2001, Starfield et al., 2002). Second, as previously stated in Section 5.2.1, dermatological conditions tend to develop independently of other illnesses, allowing the impact of co-morbidities as a potential confounder to be disregarded (Chen, 2001). Third, patient descriptions and images can be presented easily and accurately in paper form with relatively little loss of information. The patient cases were selected to cover the spectrum of clinical need or urgency, from conditions which can be commonly treated in primary care (keratosis) to conditions which require immediate referral (melanoma). The selected

conditions were seborrhoeic keratosis, cutaneous malignant melanoma, chronic plaque psoriasis and melanocytic naevi.

Seborrhoeic keratoses are benign growths on the skin, commonly harmless and do not become malignant (The British Association of Dermatologists, 2008b). Even so, they can itch, become inflated, catch on clothing, and be considered unpleasant to look at. They are usually diagnosed and managed in primary care, although referral to a dermatologist or surgeon with a special interest in skin lesions, may be justified for removal or when melanoma is suspected.

Cutaneous malignant melanoma is a cancer of the pigment cells of the skin, most often brought about by exposure to too much ultraviolet light in sunlight (The British Association of Dermatologists, 2008a). GPs are expected to identify and refer patients with melanoma to a specialist. Because treatment is most effective at the early stages of the disease, timely referral is essential.

Chronic plaque psoriasis is a common skin problem consisting of sharply demarcated dull-red scaly plaques located particularly on extensor prominences and the scalp (National Institute for Health and Care Excellence, 2012). GPs are expected to manage patients presenting with psoriasis in the primary setting, with referral being justified when there is diagnostic uncertainty, when the condition is severe or extensive, cannot be controlled with topical therapy or when it is having a major impact on a person's physical, psychological or social well-being (National Institute for Health and Care Excellence, 2012).

Finally, melanocytic naevi, commonly referred to as moles, are localised and benign accumulations of melanocytes (Primary Care Dermatology Society, 2012). They can be removed if there is suspicion of melanoma, if they cause discomfort (by, for example, catching on clothing), or for cosmetic reasons. GPs are expected to monitor and advise patients on self-examination of naevus so as to make sure there are no changes in size,

shape and colour. If there are changes, if there is diagnostic uncertainty or a family history of melanoma, then a referral is justified (Primary Care Dermatology Society, 2012).

Given the information provided for each of the above cases in the questionnaire, I hypothesised that respondents would more likely refer patients presenting with melanoma and naevi, while they would more likely not refer patients presenting with keratosis or psoriasis. Although there is margin for different interpretations, this is considered a fair representation of what happens in real practice.

As the decision to refer is not solely based on perceived clinical need, three other attributes were included: the distance from the GP practice to the specialist appointment, the average waiting time for a normal priority dermatology appointment (defined as the average time, in days, between the date the patient is referred for a normal priority appointment and the date she is seen by the dermatologist), and whether or not the patient and/or family are pressuring for a referral. These attributes and their levels were selected so as to inform the evaluation of teleconsultations in Alentejo. Other factors such as patient age, gender and social class were excluded given previously mixed evidence on their impact.

On the first page of the questionnaire, participants were asked to provide information on a number of personal and practice characteristics: age, gender, years of experience, practice size (number of patients), health sub-region, distance to closest public referral hospital (in Portugal, GPs must refer patients to specific hospitals, depending on the specialty; for example in Alentejo, the referral hospital for dermatology is in Évora), distance to closest private dermatology practice, whether or not the GP practiced (or had practiced) telemedicine, whether or not the GP had a special interest in dermatology, and an assessment of his/her patients' general health status. The responses provided further information to characterise the referral decision.

A questionnaire was developed and tested with four GPs from two practices, one in the Lisbon area, and another in the Alentejo area. Respondents took approximately 8

minutes, on average, to complete the questionnaire (range 3-12 minutes). Table 7-1 provides brief descriptions of the attributes and levels included in the questionnaire. Attribute levels were chosen in the context of the Portuguese NHS. For example, the levels for waiting time reflect Ministry for Health targets.

Table 7-1 Descriptions of attributes and levels

Attribute	Levels / Descriptions
<i>Clinical need*</i>	Seborrheic keratosis (baseline) Malignant melanoma Psoriasis Melanocytic nevus
<i>Waiting time for normal priority appointment</i>	30 days 60 days 150 days 365 days
<i>Distance from GP practice to consultation</i>	Consultation takes place in GP practice (0 km) 30 km 60 km 100 km
<i>Pressure from patient and/or family to refer</i>	Yes No (baseline)

* Each clinical case was described in terms of lesion size, type, shape and symmetry, colour and pigmentation, surface features and distribution over body (according to dermatological good practice). For each case, a picture was provided.

7.2.2 Experimental design

The complete factorial design for the 3 attributes with 4 levels and 1 attribute with 2 levels contains $4^3 \times 2^1 = 128$ scenarios. I used SASTM macros⁴ to block the 128 scenarios into 8 versions of 16 scenarios each. All versions included all levels of all attributes and maintained

⁴ A linear arrangement of all factors (levels of attributes in Table 7-1) was produced using the %MktEx macro. The %MktBlock macro was then used to add a blocking factor to the linear arrangement, a factor that is orthogonal to all of the attributes of all of the alternatives. This was confirmed using the %MktEval macro which shows correlations between all factors. For a more detailed discussion of these macros refer to Kuhfeld (2009).

orthogonality (i.e., attributes were statistically independent). Furthermore, all levels of each attribute appeared with equal frequency across versions (i.e., the design was level balanced). Respondents were randomly assigned to version. In order to test for heterogeneity of preferences across respondents one of the 128 scenarios was repeated across all 8 versions. This meant that 7 versions had 17 scenarios rather than only 16. For each choice scenario, respondents were given three options: not to refer; refer for a normal priority appointment; or refer to an urgent priority appointment. Participants were asked to choose only one of the options in each scenario.

7.2.3 Recruitment and data collection

I distributed the questionnaires in two events of the Portuguese Association of General Practice in Aveiro and Évora, in the beginning of 2013. The two events involved around 600 participants (accurate numbers are not available). The questionnaires were intended to be self-completed and were distributed in the delegates' welcome packages. Copies of parts of the questionnaire (in Portuguese), including an example of a choice set, are provided in Appendix C.

7.3 Econometric analysis

McFadden's random utility theory supplies the basis for the econometric analysis of response data (McFadden, 1980):

$$U_{ni} = \beta_n X_n + \gamma_n Z_i + \varepsilon_{ni} \quad (7-1)$$

where Z represents the observable (i.e., measurable) vector of characteristics (e.g., age, gender, etc.) of GP i ; X denotes features (i.e., attributes such as clinical need, waiting time, location, patient pressure) of choice n ; and ε represents the stochastic elements that are specific to, and known only by, the GP and not by the observer (i.e., ε accounts for the researcher's inability to completely and accurately observe GP i 's behaviour).

7.3.1 Model specification

In each of the choice sets, in each version, respondents must choose one of three mutually exclusive alternatives: not to refer, refer to a normal priority appointment, or refer to an urgent priority consultation. There are a number of models that can be used to analyse the data. One option is to define a binary dependent variable and apply a binary logit or probit model (see for example (Hall et al., 2002)). Another option is to apply a multinomial logit and analyse all three response options concurrently. The multinomial logit model assumes independence from irrelevant alternatives, so that excluding one of the options (e.g., refer to an urgent priority consultation) should not affect the estimates for the other two options. If that is not the case, then a nested logit model can be used to nest options that are close or somehow similar (e.g., refer to a normal priority appointment and to an urgent one).

I estimated all three options: binary logit models, a multinomial logit and nested logit models. For the nested logit model, I created two decision levels: to refer or not, and the choice of priority given a decision to refer (two options: normal or urgent). The inclusive values or logsum parameters (also called dissimilarity parameters) were very close to 1 for the nested model, suggesting very low correlations among pairs in the refer nest (normal and urgent priorities), and suggesting that the nested logit approach is not appropriate. With regards to the binary and multinomial logit models, these produced similar results, both qualitatively and quantitatively (i.e., coefficients of explanatory variables had same sign and the values were in the same order of magnitude). For simplicity of presentation and discussion, only the binary models are included here.

I estimated models for the decision to refer or not to refer and for the referral priority. In the first case, the binary choice being modelled is the decision to refer (denoted $Y = 1$) or not to refer ($Y = 0$) the patient. I infer from $Y = 1$ that the utility the GP derives from referring the patient (U_1) is greater than the utility he or she derives from not referring the patient (U_0). I assume y_i^* is a latent (unobserved) variable representing the difference in

utility between the referral options being compared for the i^{th} GP. I do not observe the difference in utility but rather the choice made (y_i). Therefore:

$y_i = 1$ implies $y_i^* > 0$ and 0 else, with

$$y_i^* = (\beta_n X_n + \gamma_n Z_i + \varepsilon_{ni}) - (\beta_m X_m + \gamma_m Z_i + \varepsilon_{mi}) \quad (7-2)$$

where n and m are the choices available (e.g., to refer or not to refer). Because each GP responds to more than one choice set, errors are not independent and so equation (7-2) is rewritten to include an unobserved, individual-specific term:

$$y_i^* = (\beta_n X_n + \gamma_n Z_i + \varepsilon_{ni} + \mu_i) - (\beta_m X_m + \gamma_m Z_i + \varepsilon_{mi} + \mu_i) \quad (7-3)$$

Terms that are common to the utility of both referring and not referring (e.g., μ_i) are dropped from equation (7-3). However, including a constant term (α) and a GP-specific error term (μ_i) can provide relevant information (e.g., to test for misspecification due to unobservable attributes). The model to be estimated for each pairwise choice k is then:

$$y_{ki}^* = \alpha + \beta' X + \gamma' Z_i + v_{ki} + \mu_i \quad (7-4)$$

where y_{ki}^* is the probability of GP i choosing option k ; α is a constant term capturing the difference in the average utilities of the choices being compared; X is a matrix of attributes and their levels describing the referral conditions and β' the associated coefficients; Z is a matrix of GP and practice characteristics and γ' the associated coefficients; v_{ki} is the error term capturing unmeasured variation across different GPs and μ_i is the random error specific to each GP, constant across that GP's choices.

The same reasoning applies to the priority decision, so that the choice being modelled is the decision to refer the patient to an urgent appointment (denoted $Y = 1$) or to a normal priority appointment ($Y = 0$). Data were analysed using Stata 12. In both cases, logit models were estimated. As mentioned, separate models were estimated for the decision to refer or not to refer and for the referral priority. In the first case, I aggregated all urgent and

normal priority referrals, while in the second case I eliminated all non-referrals from the sample. In each case, I estimated two models: (1) only attributes and (2) attributes and covariates.

7.4 Results

A total of 45 questionnaires were returned (it is not possible to determine the response rate since there is no exact number for participants who received questionnaires). The 45 questionnaires represent a total of 759 observations. One respondent answered only 4 questions (out of 17) and was thus excluded. Six other respondents did not complete the whole questionnaire with two answering 11, one answering 12, one answering 14 and another answering 16 questions. These were included, giving a total of 721 usable observations. The 44 questionnaires were spread across the 8 versions in the following way: three versions had 6 respondents and the remaining versions had 2, 4, 5, 7 and 8 respondents. There were also fifteen questionnaires with missing values for the explanatory variables (namely age, list size and distance to closest private dermatology clinic).

The characteristics of all 44 respondents and practices are shown in Table 7-2 (continuous variables) and Table 7-3 (discrete variables). The majority of respondents were young women. Statistics for the average age and gender of the GP population show the sample is not representative. In 2012, 40 percent of portuguese GPs were female (Instituto Nacional de Estatística I.P., 2014). In 2007, 71 percent of GPs were older than 50 and only 9 percent younger than 35 (Natário and Amaral, 2011). Differences between the sample and the actual population with regards to a certain variable (e.g., age or gender) can affect findings by biasing the coefficient for that variable (and other functions of the variable, e.g., quadratic forms) and/or by biasing other explanatory variables that may be related to that variable. Previous evidence on the importance of age and gender as predictors of referral behaviour is mixed (O'Donnell, 2000), and discussions of how these variables relate to other predictors is virtually non-existent.

Table 7-2 Characteristics of respondents and practices: descriptive statistics

Characteristic	N	Min	Max	Mean	σ	NR*
Age	42	26	59	36.33	10.86	2
Years of experience	44	0.08	31	8.89	10.10	-
Distance to referral hospital (km)	44	0	60	15.53	18.75	-
List size (number of patients)	43	3,600	180,000	22,092	26,276	1
Distance to private dermatologist (km)	30	0	60	6.00	13.43	14

* NR – did not respond

Across all 721 choice situations, the average referral rate (defined as the number of choice situations in which participants chose to refer, independent of priority, divided by the total number of choice situations) was 72.0 percent. The average rate across different versions was 70.7 percent, suggesting the experimental design was indeed level balanced. Among the participants, the lowest rate was 47.1 percent and the highest was 100 percent (there were two respondents who chose to refer every patient). The average referral rate for melanoma was 96.2 percent (a total of five GPs did not refer), 97.6 percent for melanocytic nevus, 28.2 percent for seborrheic keratosis and 60.5 percent for psoriasis. The latter two dermatological cases exhibited the largest variation, with some GPs choosing to refer everyone and others choosing to refer no one.

Table 7-4 shows the number of referral choices for each clinical presentation, demonstrating the importance of the attribute in deciding between the three outcomes. The decisions for melanoma and nevus seem to be more about referral priority, while for psoriasis and keratosis the question seems to be around whether or not to refer. As for the question repeated in all versions (attributes: nevus, 30 days wait, appointment takes place in GP practice, patient and family are pressuring for a referral), 2 GPs did not refer, 12 referred to a normal priority appointment and 25 to an urgent consultation.

Table 7-3 Characteristics of respondents and practices: categorical variables

Characteristic	<i>n</i>	%
Sex		
Male	7	16%
Female	37	84%
<i>Total</i>	<i>44</i>	<i>100%</i>
Health sub-region (ARS)		
North (<i>ARS Norte</i>)	14	32%
Centre (<i>ARS Centro</i>)	11	25%
Alentejo (<i>ARS Alentejo</i>)	12	27%
Algarve (<i>ARS Algarve</i>)	2	5%
Other	1	2%
Missing	4	9%
<i>Total</i>	<i>44</i>	<i>100%</i>
Practices or has practiced telemedicine		
Yes	11	25%
No	33	76%
<i>Total</i>	<i>44</i>	<i>100%</i>
Special interest in dermatology		
Yes	29	66%
No	15	33%
<i>Total</i>	<i>44</i>	<i>100%</i>
Average general health status of patients		
Very good	-	-
Good	1	2%
Neither good nor bad	20	46%
Bad	22	50%
Very bad	1	2%
<i>Total</i>	<i>44</i>	<i>100%</i>

It is also interesting to explore whether respondents traded-off non-clinical attributes (i.e., made different choices for the same clinical presentation when levels of non-clinical attributes changed). A total of 26 participants referred the same patient with different

priorities, while 23 GPs changed their decision to refer altogether depending on non-clinical attributes. In other words, 23 respondents could refer or not refer exactly the same clinical presentation given differences in waiting time, distance or pressure to refer. Furthermore, as many as 13 respondents chose all three options at least once for the same clinical case. Only 8 GPs did not trade-off non-clinical attributes (i.e., made always the same choice regardless of levels of non-clinical attributes).

Table 7-4 Referral decisions by clinical presentation

Choice	Melanoma	Keratosis	Nevus	Psoriasis	Total
Do not refer	8	122	4	68	202
Refer with normal priority	46	43	52	94	235
Refer to urgent appointment	156	5	113	10	284
<i>Total</i>	<i>210</i>	<i>170</i>	<i>169</i>	<i>172</i>	<i>721</i>

7.4.1 Estimation results

Separate models were estimated for the decision to refer or not to refer and for the referral priority. For the referral decision, I created a dummy variable indicating whether a referral (either for a normal priority or an urgent appointment) was chosen by the GP, taking a value of 1 if a referral had been made. For the priority decision, I eliminated all non-referrals from the sample and defined the baseline as referring for a normal priority consultation. In each case, two models were estimated: (1) only attributes and (2) attributes and covariates. Table 7-5 and Table 7-6 present the results for the referral and priority decisions respectively.

Table 7-5 Results of random effects logit models of decision to refer / not refer

Variables	Model 1a		Model 1b	
	<i>Coefficient</i>	<i>Standard error</i>	<i>Coefficient</i>	<i>Standard error</i>
CONST	-1.062***	(0.371)	-13.12***	(4.503)
NEVUS	5.008***	(0.474)	5.901***	(0.737)
MELANOMA	5.549***	(0.598)	5.409***	(0.680)
PSORIASIS	1.746***	(0.279)	1.761***	(0.358)
WAITING TIME	0.000184	(0.000938)	0.0000828	(0.00118)
DISTANCE	-0.00403	(0.00325)	-0.00286	(0.00411)
PRESSURE	0.0671	(0.246)	0.219	(0.316)
AGE			0.606**	(0.245)
MALE			-0.186	(0.650)
DIST_HOSPITAL			0.0400**	(0.0162)
DIST_PRIVATE			-0.0293	(0.0214)
LIST SIZE			0.0000163	(0.0000227)
TELEMEDICINE			-1.911***	(0.603)
SPECIAL_INT			-1.149**	(0.526)
HS_POOR			0.423	(0.481)
HS_GOOD			-1.698	(1.377)
AGE_SQUARED			-0.00680**	(0.00293)
<i>Observations</i>	<i>721</i>		<i>473</i>	
<i>AIC</i>	<i>514.1</i>		<i>327.2</i>	
<i>Log likelihood</i>	<i>-249.1</i>		<i>-145.6</i>	
<i>Chi-squared</i>	<i>152.5***</i>		<i>95.37***</i>	
<i>Hosmer & Lemeshow Chi2</i>	<i>12.22</i>		<i>13.84*</i>	
<i>% pred. correctly</i>	<i>82.25%</i>		<i>86.47%</i>	
<i>Area under ROC</i>	<i>0.8763</i>		<i>0.9215</i>	

***significant at 1% level; **significant at 5% level; * significant at 10%. Dependent variable is a dummy indicating whether a referral was made. Base categories for dummy variables: Need – Keratosis; Pressure – No; Telemedicine – No; Special interest – No; Health status – Neither good nor bad

Table 7-6 Results of logit models of choice of priority

Variables	Model 2a		Model 2b	
	<i>Coefficient</i>	<i>Standard error</i>	<i>Coefficient</i>	<i>Standard error</i>
CONST	-3.172***	(0.641)	4.010	(3.583)
NEVUS	4.117***	(0.604)	4.958***	(1.113)
MELANOMA	3.348***	(0.589)	3.963***	(1.040)
PSORIASIS	-0.367	(0.653)	0.206	(0.922)
WAITING TIME	0.00504***	(0.00108)	0.00446**	(0.00226)
DISTANCE	0.00507	(0.00330)	0.00663*	(0.00350)
PRESSURE	-0.450*	(0.250)	-0.628*	(0.344)
AGE			-0.471***	(0.180)
MALE			0.879**	(0.434)
DIST_HOSPITAL			0.0185	(0.0144)
DIST_PRIVATE			-0.0175	(0.0126)
LIST SIZE			0.0000245	(0.0000159)
TELEMEDICINE			0.306	(0.532)
SPECIAL_INT			0.529	(0.419)
HS_POOR			-0.319	(0.469)
HS_GOOD			0.669	(0.421)
AGE_SQUARED			0.00584***	(0.00217)
<i>Observations</i>	<i>519</i>		<i>342</i>	
<i>AIC</i>	<i>488.0</i>		<i>296.1</i>	
<i>Log likelihood</i>	<i>-236.0</i>		<i>-132.0</i>	
<i>Chi-squared</i>	<i>114.9***</i>		<i>81.85***</i>	
<i>Hosmer & Lemeshow Chi2</i>	<i>9.29</i>		<i>8.38</i>	
<i>% pred. correctly</i>	<i>77.65%</i>		<i>82.46%</i>	
<i>Area under ROC</i>	<i>0.8350</i>		<i>0.9000</i>	

***significant at 1% level; **significant at 5% level; * significant at 10%. Dependent variable is a dummy indicating whether a referral for an urgent appointment was made (base category is referral to normal priority appointment). Base categories for dummy variables: Need – Keratosis; Pressure – No; Telemedicine – No; Special interest – No; Health status – Neither good nor bad

Concerning the decision to refer or not to refer the patient, both the percentage of correctly predicted responses and the area under the receiver operating curve (ROC) indicate the inclusion of covariates and age-squared provides a better fit. The Hosmer &

Lemeshow test statistic also confirms goodness of fit (chi-squared not significant). The signs of the coefficients in the two models are in line with the literature, providing evidence of theoretical validity (although the sign for the attribute waiting time is positive, it is very close to zero and not statistically significant, indicating that waiting time has little or no impact on the GP's decision to refer or not to refer a patient, *ceteris paribus*). Significant effects and signs of coefficients from model 1a (only attributes) persist in model 1b (attributes and covariates).

The need variables (melanoma, nevus and psoriasis) are all highly significant and positive, indicating that GPs are more likely to refer patients presenting with these conditions compared to keratosis (the base case). Non-clinical attributes (waiting time, distance and pressure) are not significant. Physicians from practices farther away from the referral hospital are more likely to refer. Age affects the likelihood of a referral being made in a non-linear way ($0.61(\text{age}) - 0.0068(\text{age})^2$). As physicians get older the probability of a referral increases reaching a maximum around 45, after which they begin to refer fewer patients, all other things being equal. Those GPs having practiced or currently practicing telemedicine are less likely to refer. This is also the case for GPs with a special interest in dermatology.

Regarding the choice of priority, we found evidence of random effects in model 2a ($Rho = 0.21$, $p < 0.000$) but not in model 2b ($Rho = 0.03$, $p < 0.293$). Table 7-6 presents random effects logit estimators for model 2a and pooled logit estimators for model 2b. The percentages of correctly predicted responses, the area under the ROC, and the Hosmer & Lemeshow test statistic all indicate that including covariates and age-squared leads to a superior fit over model 2a. The fact that model 2b no longer has evidence of random effects suggests that the extra explanatory variables are adequately capturing the heterogeneity among GPs. This is confirmed by a Lagrange multiplier test for omitted variables, which indicates the model is well specified (I added the linear predicted value squared to the model which was shown not to be significant: $Chi\text{-squared} = 0.61$, $p < 0.4355$). Significant effects from model 2a persist in model 2b. Although there is no previous literature on which to base

the theoretical validation of the model the results are in line with expectations. Namely, GPs are more likely to refer patients presenting with melanoma or nevus for urgent appointments when compared to patients presenting with keratosis. There is no significant difference in referral priorities between keratosis and psoriasis. As for non-clinical attributes, the longer the waiting time for a normal appointment, the more likely it is that GPs refer to an urgent consultation. Age affects the likelihood of a referral to an urgent appointment being made (compared to a referral to a normal priority consultation) in a non-linear way $(-0.47(\text{age}) + 0.0058(\text{age})^2)$. The likelihood of making a referral for an urgent appointment decreases up to an age of 40, after which it starts to increase. Being a male GP is associated with a higher probability of making an urgent referral.

7.4.2 Model prediction

Marginal effects can be used to illustrate how different attributes and covariates affect the probabilities of being referred and the referral priority. Because a number of explanatory variables are categorical, marginal effects at representative values are used, rather than average marginal effects or marginal effects at the means. The rationale is that, in reality, GPs either practice (or have practiced) telemedicine or they do not (or have not); the sample mean value (0.25 for the 44 respondents) has no practical meaning and thus should not be used to determine marginal effects. This applies to other factor variables such as, for example, being pressured by patients and/or families, having a special interest in dermatology or being male. Using marginal effects at representative values, one can determine the effect of significant predictors on choices made in real life situations.

The values of categorical variables were selected so as to be representative of the samples used in estimation (models 1b and 2b). For example, since 69 percent of GPs in the sample used in model 1b had a special interest in dermatology, I estimated marginal changes in the probability of referral for GPs with a special interest in dermatology (the same

reasoning applies to other categorical variables). Continuous variables were held at their sample means. The base cases used are described in Table 7-7.

Table 7-7 Base cases used to determine marginal effects

Attributes/covariates	Level/value*
Model 1b	
Waiting time	143.74 days
Distance	45.62 km
Pressure	Yes
Age	36.58 years
Male	No
Distance hospital	17.69 km
Distance private	5.97 km
List size	19,666 patients
Telemedicine	No
Special interest	Yes
Health status	Neither good nor bad
Model 2b	
Waiting time	141.65 days
Distance	44.88 km
Pressure	Yes
Age	37.25 years
Male	No
Distance hospital	19.17 km
Distance private	6.41 km
List size	19,609 patients
Telemedicine	No
Special interest	Yes
Health status	Neither good nor bad

*Values for discrete variables chosen so as to represent sample modes while values for continuous variables are sample means.

Figure 7-2 illustrates the impact of distance to the referral hospital on the probability of referring (compared to not referring). It shows that distance affects the probability of referring

patients presenting with keratosis or psoriasis much more than patients presenting with melanoma or nevus. This is comforting, since it means that patients with melanoma or nevus will almost always get referred, irrespective of where the GP practice is located. On the other hand, for conditions for which uncertainty is higher, the difference in probabilities can be almost twofold, confirming the high variability in referral decisions for these conditions.

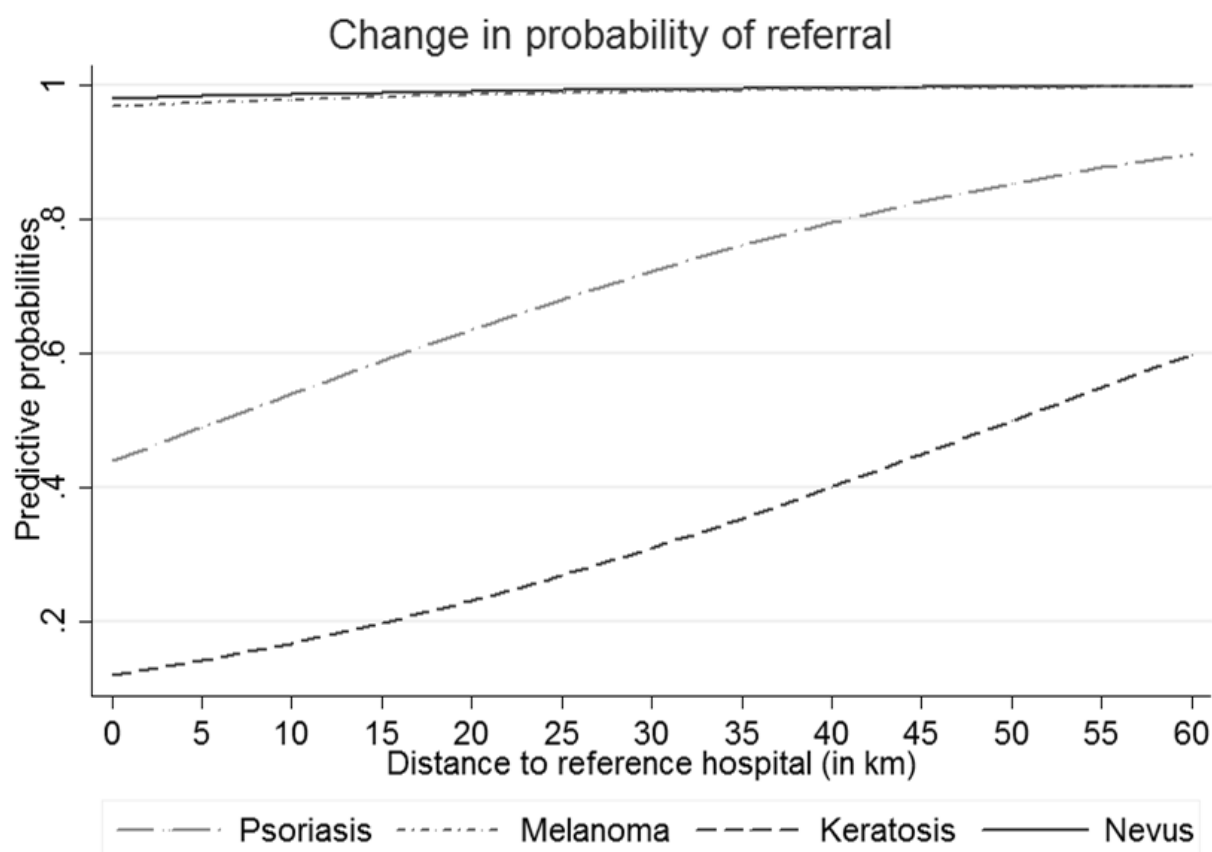


Figure 7-2 Impact of distance to referral hospital on probability of referral

Figure 7-3 shows the impact of waiting time for a normal priority appointment on the choice of priority. It illustrates that waiting time has again a different effect depending on whether patients present with more severe conditions (melanoma and nevus) or less severe ones (keratosis and psoriasis). In both cases, the longer the waiting time for a normal priority appointment, the more likely GPs are to refer to an urgent appointment. Figure 7-3 also shows that GPs are more sensitive to waiting time when patients present with melanoma or

nevus, while for those presenting with keratosis and psoriasis GPs only start to react at around 150 days (~4 months).

Marginal effects were also estimated for statistically significant discrete explanatory variables. These are presented in Table 7-8 for each clinical presentation. GPs who have practised or are currently practising telemedicine, as well as those who have a special interest in dermatology, are significantly less likely to refer patients presenting with keratosis or psoriasis. They are also less likely to refer patients presenting with melanoma or nevus, not a reassuring result, although it is important to note these marginal changes were not statistically significant.

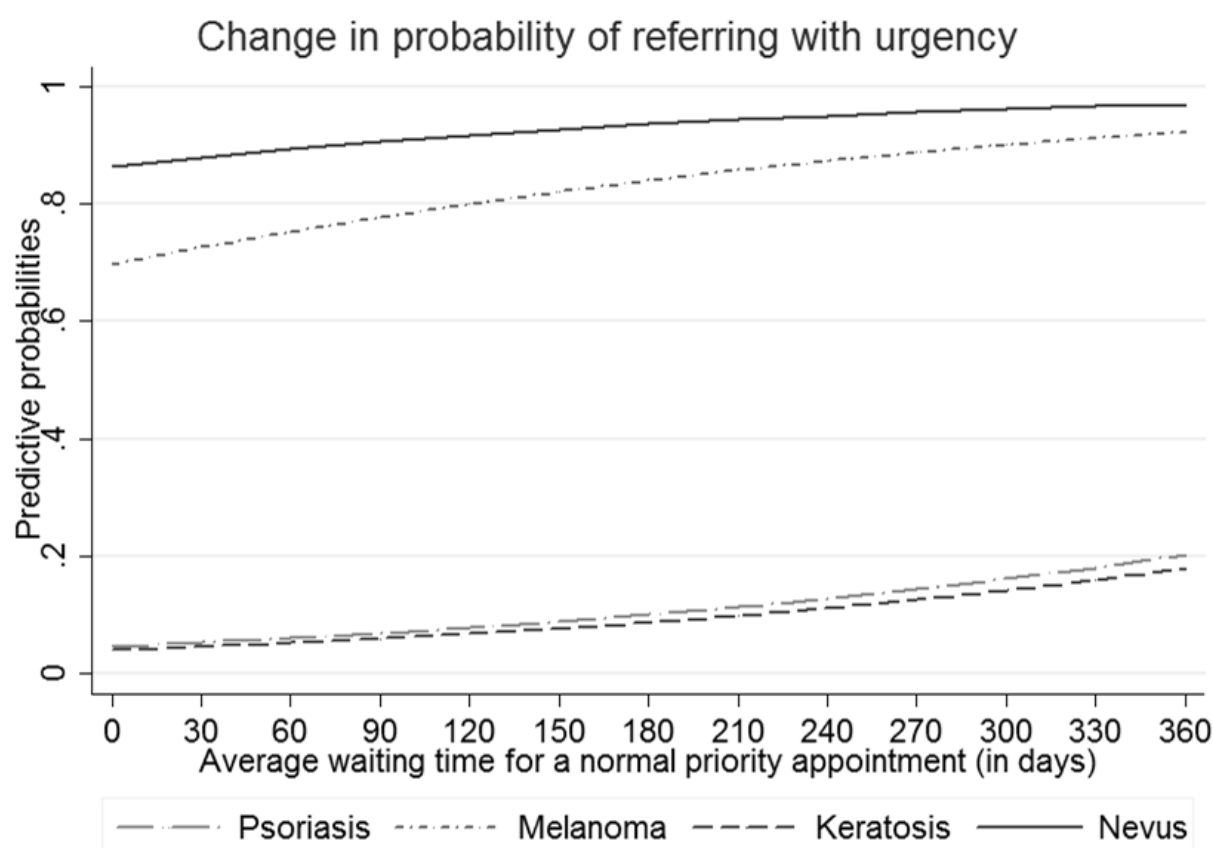


Figure 7-3 Impact of average waiting time for a normal priority appointment on the probability of referring to an urgent consultation

Table 7-8 Marginal changes in probabilities of referral for different conditions

	Melanoma	Keratosis	Nevus	Psoriasis
<i>Decision: Refer vs not refer</i>				
Telemedicine	-8.35%	-17.58%**	-5.35%	-42.34%***
Special interest	-1.09%	-24.85%*	-0.68%	-21.99%**

*** significant at 1% level; ** significant at 5% level; * significant at 10%.

Base case for dependent variable: Refer – No.

7.5 Discussion

This study shows that it is possible to use stated choice methods to explore the determinants of GPs' referral decisions. The results suggest that models of both the decision to refer and the choice of priority are theoretically valid. Both models fitted the data well, and in both cases clinical and non-clinical variables were important. As expected, the different clinical presentations (intended as a gradient of clinical need) have the biggest impact on GPs' decisions, already visible in Table 7-4, even before running regressions. It is important to note that the diagnoses were made by the respondents based on the descriptions provided in the vignettes, and that all participants were given the same clinical information. There was considerable variation at different levels: both across GPs and clinical presentation categories. Patients with the same clinical presentation could either be referred to a normal priority appointment, to an urgent consultation or not at all, depending on the GP assessing the case. Furthermore, almost 30 percent ($n = 13$) of GPs chose all three outcomes for effectively the same clinical condition (only non-clinical attributes changed). Only 8 GPs (18 percent) chose always the same option for a specific clinical presentation.

The regression models confirm that non-clinical attributes are important. Patients presenting with melanoma and nevus were almost always referred, with non-clinical attributes affecting only the choice of priority. For patients presenting with keratosis and psoriasis, non-clinical attributes affected both the decision to refer and the choice of priority.

These are important findings since they demonstrate that future studies of referral decisions should investigate the impact of non-clinical attributes given specific clinical presentations.

For the conditions included in the study, patient access indicators, such as the average waiting time for a normal appointment and the distance to the appointment, were not significant predictors of GPs' decisions to refer. Contrary to findings from previous studies, GPs did not seem to have clear preferences for patient access characteristics. One explanation is that these factors tend to be stable over time. Waiting time changes slowly and distance not at all, unless the GP changes practices, or when the specialist is either physically present in the practice or virtually, as in telemedicine. Thus, respondents might find it difficult to interpret hypothetical differences in the choice sets as a fair representation of reality. An alternative explanation is that GPs simply do not take into account patient access indicators, when deciding whether or not to refer. Although this sounds unlikely, given previous evidence, it is important to note that the literature has focused on revealed preference data, which aggregate physician and patient preferences. This experiment elicits only GPs' stated preferences, and as such does not necessarily contradict previous research (i.e., there is the possibility that access matters to patients and not GPs).

Interestingly, the coefficient for the distance to the referral hospital (essentially a measure of patient access) is significant and positive. In reality, this distance would be the same as the distance to the appointment, which was found to be statistically insignificant. The fact that they have different effects would reinforce the notion that GPs either did not interpret the distances in the choice scenarios correctly, or simply did not take them into account. As for why the coefficient is positive, an argument can be made that GPs from practices farther away from the referral hospital are more likely to be isolated from peers, have fewer patients, and be overall more in need of support. The fact that distance was especially important for keratosis and psoriasis (as opposed to the more severe nevus and melanoma) suggests that for less serious diagnoses the probability of referral increases with distance. Distance to the referral hospital can also be interpreted as a proxy for rurality, in

which case more remote GPs will be more likely to refer patients. Again, it should be emphasised that this DCE elicits GPs' – not patients' – preferences. Even though GPs act as agents for their patients, distance is effectively a cost only for patients (GPs do not share in the costs incurred by a patient referred to an appointment 100 km away). The results may indicate that distance has opposing effects for patients and GPs. Due to the way the questionnaire was developed (to safeguard anonymity, it is not possible to identify practices in the sample) and the small sample size, it is not possible to investigate this matter further. Yet, this is an interesting finding which motivates further research.

The results also show that GPs' age influences decisions to refer patients – as well as the choice of priority – in a non-linear way. As GPs grow older they are more likely to refer patients, *ceteris paribus*, up to a certain age (around 45 years old), after which the reverse is true. The effect is most important for keratosis and psoriasis. Findings from previous research concerning the impact of GP age are mixed, although it has been suggested that the more experienced the GP the lower his/her referral rate. Older GPs are more experienced, while younger GPs tend to be more risk averse and uncertain, thus referring more often (O'Donnell, 2000). This study's results suggest the relationship is more complex. There are two possible explanations for this. The first explanation is that this relationship is a direct result of the age distribution in our sample (see Figure 7-4). As previously mentioned, the majority of respondents are young and for the interval around 38 to 52 years old, there are very few respondents. If the GPs in this age interval in our sample are, for some reason, more inclined to refer, then this would explain the quadratic effect of age. A second, more speculative, explanation has to do with the relationship between memory and experience. As GPs leave medical school their memory of dermatological symptoms and signs should be good. As time passes, and GPs see a multitude of conditions, most of them not dermatological, it is natural that memory will fade. However, with time there is also a build-up of experience. The quadratic effect of age could represent opposing and simultaneous

dynamic effects of memory and experience. Larger sample sizes are needed to investigate this further.

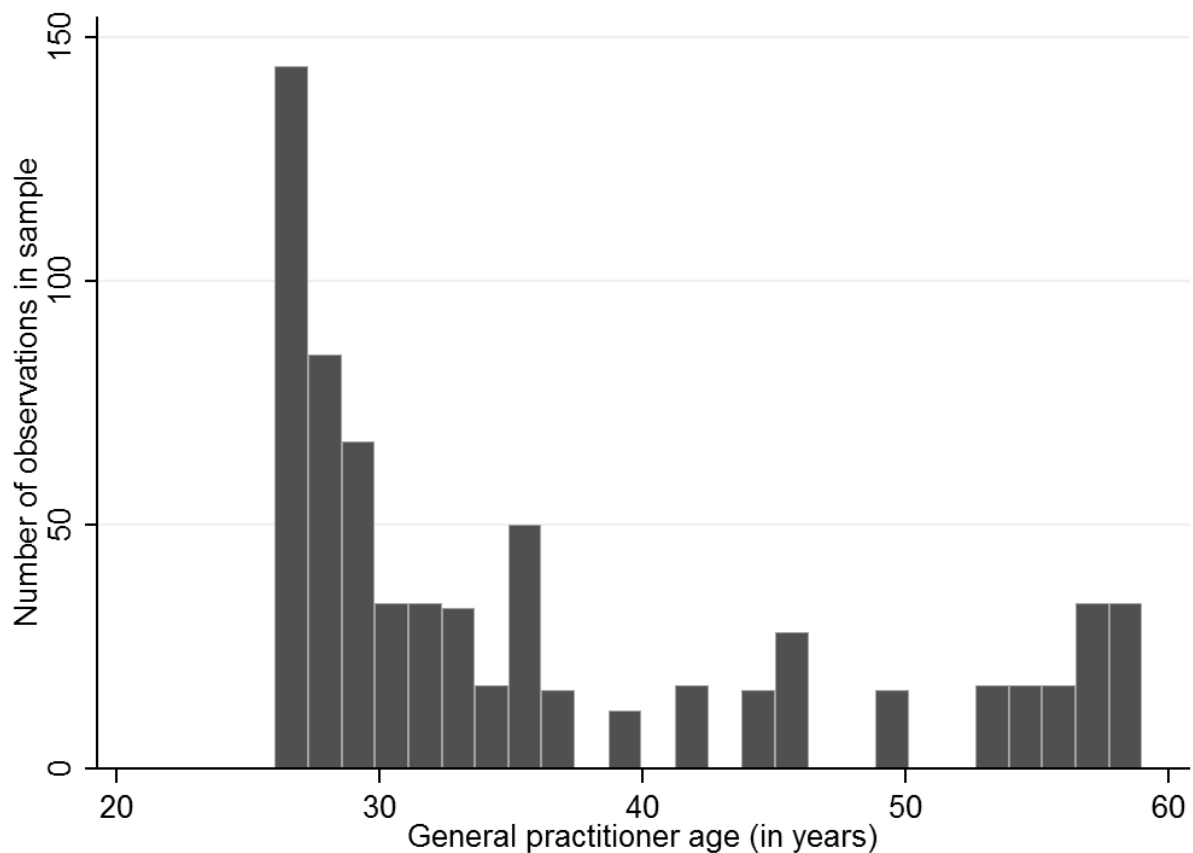


Figure 7-4 Distribution of GP age in sample

The two final significant predictors of GPs' decisions to refer are having practised or currently practicing telemedicine, and having a special interest in dermatology. Both these characteristics are associated with a reduced likelihood of a referral being made. They are especially important for keratosis and psoriasis, less so for melanoma and nevus. These results give strength to the theoretical validity of the DCE, as they are in accordance with previously published research into the effect of telemedicine on referral rates and physician learning (Jarvis-Selinger et al., 2008, Loane et al., 2001, Taylor, 2005). As for having a special interest in dermatology, the mechanism may involve uncertainty and risk. GPs with a special interest in dermatology are more likely to interpret signs and symptoms and to know how to treat them. Finally, although pressure from patients and/or family was not statistically

significant, previous studies on the impact of patient pressure concern referrals for headache (Morgan et al., 2007, Ridsdale et al., 2007). It may be that patient pressure matters less for dermatological conditions.

Regarding the choice of priority, significant variables included clinical presentation (e.g., whether patients presented with melanoma or nevus), average waiting time for a normal priority appointment and GP age and gender. As with the decision to refer, age affects the likelihood of a referral to an urgent appointment (compared to a referral to a normal priority consultation) in a non-linear way. However, the relationship is now U-shaped, suggesting that younger and more experienced GPs are more likely to refer to an urgent appointment than GPs between 35 and 55. Again, this relationship could be both a result of the GPs in our sample and of a memory-experience balance (see discussion on the impact of GP age above). Given that only 7 GPs in our sample were male, the result that male GPs are more likely to refer to urgent appointments than female GPs may simply reflect a biased sample (i.e., male GPs in the sample have, for some reason, a preference for referrals to urgent appointments). Finally, although waiting time does not seem to affect the decision to refer, it is a significant predictor of the choice of priority. As waiting time increases, GPs are more likely to refer patients to urgent appointments, as opposed to normal priority consultations. This means that as waiting time increases, patients with less severe conditions start to get referred as urgent. While one might make the argument that even less severe conditions might get worse with time, thus justifying an urgent referral, the fact is that if all GPs refer patients presenting with psoriasis and keratosis for urgent appointments when waiting time increases, there will be no differentiation between those and truly urgent conditions (such as melanoma and nevus). Although the reasoning is understandable, the end-result may be counterproductive.

7.5.1 Limitations and future work

There are a few limitations. First, and most importantly, the sample is small and, with respect to age and gender, not representative of the general GP population in Portugal. The sample size is comparable to that in Hall et al. (2002), although sample sizes can reach significantly larger numbers in other studies (e.g., (Gerard et al., 2008)). In the future, it is essential to investigate whether the preferences of the 45 respondents in this study are comparable to those of the general GP population in Portugal. While the findings are still informative of a younger predominantly female GP population, they may not be generalizable to the larger population of Portuguese GPs.

In the future, it would also be important to understand the influence of sample age and gender on the coefficients of other predictors. Besides age and gender, statistically significant explanatory variables in models 1b and 2b (Tables 7-5 and 7-6) included clinical presentations, distance to referral hospital, practicing or having practiced telemedicine, having a special interest in dermatology, and waiting time. Using interactions it is possible to explore whether the coefficient for a given predictor (e.g., distance to referral hospital) depends on the age and gender of the GP. Presently, given previous studies, there is no clear expectation that age and gender affect any of the other variables found to be statistically significant in this study.

A second limitation, typical of any stated preference model, is the external validity of findings. As previously mentioned, there should be a *“healthy scepticism about relying on what consumers say they will do compared with observing what they actually do”* (Page 20, Louviere et al. (2000)). One way forward would be to combine this study with revealed preference data on actual referrals made by the respondents. A third limitation is that issues of internal validity were not explored. Namely, there were no tests of respondent rationality.

Even given these limitations, the findings seem theoretically valid (either in line with previous literature or with expectations). In the future, more complex designs could be

explored which would include more response options (e.g., 'start therapy and reassess later' or 'wait and reassess later'), and more clinical and non-clinical attributes and/or levels. These more complex designs necessitate higher numbers of respondents. Bigger samples also allow for the estimation of more sophisticated models, with more interactions (e.g., psychological factors such as tolerance of uncertainty are more relevant when uncertainty is present) and more parameters (e.g., different GPs may have different preferences for a specific attribute, in other words there may be respondent heterogeneity).

7.6 Conclusion

Significant predictors of the decision to refer or not to refer included clinical presentation, distance from GP practice to referral hospital, GP age, practicing or having practiced telemedicine and having a special interest in dermatology. Regarding the choice of priority, significant explanatory variables included clinical presentation, the average waiting time for a normal priority appointment, and GP age and gender. The results are in line with findings from previous research that non-clinical attributes affect GPs' referral decisions. There are two policy implications. First, improving communication and collaboration between primary and secondary physicians (through, for example, the use of telemedicine or potentially equivalent initiatives such as joint consultations) can lead to reduced referrals of less serious clinical presentations. Second, the results indicate that GPs are more likely to refer patients for urgent appointments when waiting times increase. This has the potential to cause problems since it defeats the point of referral priority: to ensure that specialists give precedence to severe clinical presentations in lieu of less urgent ones. Such behaviour has the potential to disrupt the triage process and hinder the GP – specialist relationship. Policy and decision makers should ensure that longer waiting times are not persistent, or at the least, that guidelines are available for GPs to refer only those cases which require urgent attention.

This study illustrates how stated choice data and DCEs can be used to identify the impact of multiple factors on the referral decisions of GPs. The results can be used to predict the likelihood of a referral being made, and the priority chosen, based on changes in determinants. Even given a small sample size, the results show high variation in referral rates and decisions. To my knowledge, this is the first study to use stated choice methods to explore the referral decisions of GPs. It is also one of the first quantifications of the impact of experience with teleconsultations on referral rates. It is an important step forward in understanding the impact of continued medical education on referral rates, and moving away from expert opinion as the sole source of data. In the future, it would be desirable to explore the notion of shared decision making, and how it relates to referrals. Although the DCE in this study focused on referral decisions made by Portuguese GPs, the approach can be used to study physicians' preferences regarding health care services in other contexts and applications, not just patient preferences as in the past.

7.6.1 Implications for the evaluation of teleconsultation in Alentejo

The findings from this study have implications for the evaluation of teleconsultations. First, waiting times seem to affect the likelihood of GPs referring to an urgent appointment, compared to a normal priority consultation, but not the decision to refer or not to refer. Taking these results as valid, findings of revealed preference studies suggesting waiting times have a negative impact on referrals, reflect patient preferences rather than GPs' preferences, which is interesting. In the context of teleconsultation, this would mean that reductions in waiting times will not cause GPs to change their referral rates.

Second, GPs who participate in teleconsultations have statistically significant lower referral rates than those GPs who do not practice teleconsultation, especially for keratosis and psoriasis. As such, an evaluation of dermatology teleconsultations should differentiate between urgent conditions (such as melanoma and nevus) and less pressing conditions (such as keratosis and psoriasis). The reductions in probability of referral estimated in this

study for keratosis and psoriasis can be used to determine the impact of introducing teleconsultations on the number of referrals. This, combined with the findings from the previous chapter, makes it possible to explore two mechanisms that were previously contained in black boxes in previous evaluations of teleconsultations.

Third, for patients, distance is a measure of access, which explains why more remote locations have lower rates of utilisation. On the other hand, for GPs distance might be a measure of isolation from peers associated with a tendency to rely on the expertise of consultants. In that case, the reduction in distance due to teleconsultation will have a similar effect to that of continued GP education: improving communication between GPs and specialists and reducing isolation from peers.

Chapter 8

THE SIMULATION MODEL

In the previous chapters of Part 2, I have: introduced the complex health intervention under analysis; reviewed the evidence on the intervention's costs and consequences; described the intervention in a specific setting using interviews, surveys and documentation; and conducted two studies of preferences and decision making with implications for evaluation studies. In this chapter, I build on the wealth of information and data from previous work, to develop a simulation model which takes into account the wider system and context in which those delivering and receiving the intervention operate. As discussed in Chapter 3, the costs and consequences of complex health interventions emerge from the behaviours and decisions of those involved (e.g., patients, families, health care workers, etc.). Identifying the important causal mechanisms at play, and their likely impact from a holistic systems viewpoint, is essential to determining how and why the intervention works, and whether it could work elsewhere. This is the first of two objectives of this chapter: to assess how explicitly accounting for individual behaviours and the wider context informs the evaluation of teleconsultation in Alentejo. In doing so, I address the following questions:

- 1) What are the multiple costs and consequences of teleconsultation in Alentejo?

- 2) What are the costs and consequences from different perspectives? Namely, what is the impact for patients and families as well as the Portuguese NHS?
- 3) What is the impact of scale and time on the results of the evaluation? Specifically, can positive effects from previous studies at practice-level be extrapolated to the broader regional level, and how do costs and consequences evolve over time?
- 4) What is the impact of contextual differences in the use of the intervention? For example, how do differences in the intervention at the practice-level affect outcomes?
- 5) What are the policy levers and their likely effects? Particularly, what are the effects of current policies, and what would be the likely impact of changes to those policies? For example, should the use of teleconsultation be promoted at a broader level?
- 6) What are the generalizable findings? Are causal mechanisms present in other settings, and are they likely to lead to the same general outcomes?
- 7) What are the areas which warrant further research and/or data collection? Does a holistic systems perspective lead to the identification of gaps and/or inconsistencies in previous theoretical and empirical research? What is the sensitivity of results to data limitations, and what are the mechanisms with high impact on results for which further data collection is justified?

The second objective is to assess the advantages and limitations of the approach proposed in Chapter 3, based on its application, in Part 2 of this thesis, to a specific intervention and setting. These are discussed briefly at the end of the chapter and in greater detail in the final discussion and concluding remarks of the thesis, in Chapter 9. In the remainder of this chapter, I introduce the simulation model describing how causal mechanisms identified in previous chapters were operationalised and parameterised. Given the complexity of the problem, the model is itself relatively complex. As such, in the following sections, I present aggregated and simplified versions of the underlying simulation model,

omitting a number of variables and relationships for readability. Comprehensive documentation for the full model can be consulted in Appendix B.

It is good practice that the boundaries of the modelling effort be determined at the outset and based on the objectives of the study. As discussed in Chapter 5, any evaluation of a complex intervention must identify very precisely the intervention. I focus on real-time teledermatology (dermatology teleconsultations). Second, I explore impact at two levels: the level of the individual practice and the regional level. The simulation model can be run at both levels by changing certain parameters. Simulations at the practice-level allow me to answer questions such as: what is the improvement in waiting time following the implementation of teledermatology, and what are the savings for patients and families from avoided travelling to the hospital? Simulations at the regional level allow me to answer more complex questions: what are the consequences for face-to-face practices when one practice in the region implements teledermatology, and what are the consequences for that teledermatology practice when more practices join in?

Third, the simulation model embodies what is known from previous theoretical and empirical research into the general preferences of patients and carers and the use of teleconsultation. To keep speculation and guesswork to a minimum, I have focused solely on those mechanisms and effects which have been previously mentioned and studied. The model does not explicitly address private care, self-care or visits to the A&E department, although these are implicit in some of the causal mechanisms modelled (e.g., in the elasticities of demand and non-completion). Treatments and examinations were also considered outside the scope as I found no evidence in the review in Chapter 4, or in the exploratory interviews in Chapter 5, that there were significant differences between teleconsultations and face-to-face appointments, in that respect. Future work may focus on challenging these boundaries.

Finally, having established the goals, scope and boundaries of the modelling exercise, the choice of modelling technique should now be more clearly justified. In Table 5-3, I used a number of criteria to identify system dynamics as the modelling technique most likely to yield insights in a comprehensive evaluation of teleconsultation in Alentejo. The fact that behaviour could be aggregated, changes to the system were mostly continuous, interest was on the longer term, the number of entities was large, feedback was likely to be an issue, and understanding was more important than optimisation, all pointed to system dynamics as the more appropriate method. This does not mean that other techniques could not be useful. For example, discrete event simulation would be useful if I was interested in how the participation of GPs affects scheduling at the practice-level; and agent-based simulation would be appropriate for exploring in-depth how continued GP education and learning given different patterns of participation in teleconsultation might affect referral rates. Given my interest in the broader questions above, I am confident system dynamics is the appropriate technique.

8.1 Background

The demand for medical care in general, and dermatology outpatient appointments specifically, is determined by numerous factors: need, measures of access such as waiting time and distance, availability of substitutes, behaviour of gatekeepers, etc. In their model of waiting lists for elective surgery in the UK NHS, Ann van Ackere and Peter Smith posit that demand is motivated by changes in the waiting time through a feedback loop (Smith and van Ackere, 2002, van Ackere and Smith, 1999). They show that, depending on the initial waiting time, an increase in supply can lead to an increase in demand, leading to new equilibrium conditions (Smith and van Ackere, 2002). A 10 percent increase in supply initially causes the waiting list, and waiting time, to decrease, which in turn causes demand to increase, as determined by the elasticity of demand with respect to waiting time. A year after the increase in supply, demand has risen by more than 10 percent, and both the waiting list and waiting

time are now longer. Another year later, the system seems to have converged to a new equilibrium, with demand permanently higher, and more people being treated, but waiting times just marginally shorter.

Although there are bound to be many more factors concurrently affecting supply and demand, their model shows in a very simple manner that, given a set of assumptions taken from previous theoretical and empirical studies, an increase in supply will lead to a permanent increase in demand, with only marginal improvement in waiting time. Despite the fact that more people are being treated, from a policy perspective the result is counter-intuitive: increasing supply does almost nothing to reduce waiting time and, one year after the increase, waiting time is actually higher. van Ackere and Smith (1999) illustrate how a simple model can make explicit and plain the consequences of well-established assumptions regarding how a system works. In the following sections, I describe a system dynamics model of teleconsultation in Alentejo, building on van Ackere and Smith's work.

8.2 A model of teleconsultation in Alentejo

In essence, the study of teleconsultation in Alentejo is the study of patients moving through the outpatient dermatology care pathway. The movement of patients can be depicted in system dynamics as in Figure 8-1. The routes are depicted by pipes (double lines) through which patients move. Patients flow from sources of people seeking care to sinks of people who either feel better or gave up (both shown as clouds). These clouds represent the boundaries of the model, in the sense that whatever happens after patients flow into these clouds is not considered. Decisions and actions made before patients seek care, as well as those made after patients flow out of the system, are not addressed. Patients' progress is regulated by rate of flow variables on the pipes (shown as valves). The rates of flow are often dependent on further variables, which may move in the same direction or in an opposite direction to the flow they affect (the direction is indicated by a plus or minus sign in the arrow connecting the variables; a plus sign means the variables move in the same

direction). Patients flow into and out of different states, depicted by boxes, where they wait for further progress through the system. These boxes are stocks of patients, essentially accumulations representing the difference between patients entering and exiting the lists.

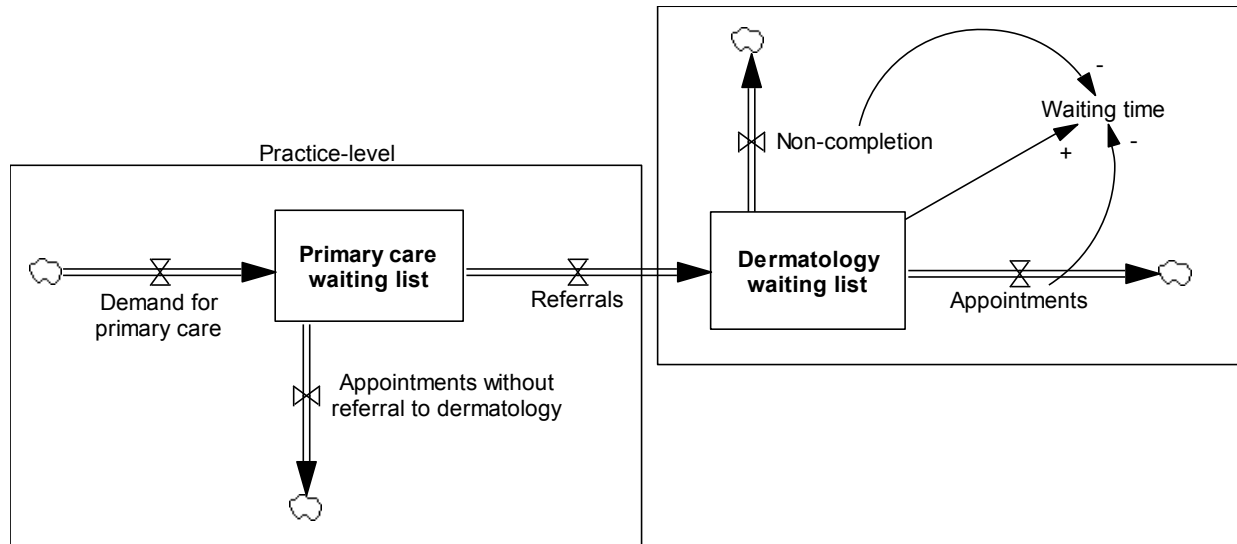


Figure 8-1 Simplified depiction of patient flows in dermatology outpatient care

At the model's core are the twin concepts of a waiting list and a waiting time. In 2002 and 2003 there were three dermatologists in *Hospital do Espírito Santo de Évora*, one of whom left after 2004. From 2004 to today, the remaining two dermatologists have been the only full-time dermatologists in the hospital. Notwithstanding a number of internists that came for short periods of time (usually a couple of months) and thus increased dermatology outpatient capacity momentarily, the supply of dermatologists in central and lower Alentejo can be considered fixed for the period under analysis (2002-2013). This allows me to focus more thoroughly on the demand side, treating supply as an exogenous variable.

Figure 8-1 shows how patients flow and accumulate in the ambulatory dermatology care pathway. Once they decide to visit their GP, they enter into the primary care waiting list. When they are seen by the GP, they can either be referred to a dermatology appointment or not, in which case they exit the system. A patient who exits the system through this cloud can naturally re-enter through the demand for primary care flow. It is important to note here a trait of system dynamics: aggregation. There is no way to identify or track individual patients

as they flow through the various stages, except to model a population of one individual. Rather, the focus is on the average patient or group of patients. As van Ackere and Smith (1999), I assume a homogeneous set of patients, which are representative of the average individual in Alentejo. It is theoretically possible to introduce heterogeneity, in terms of age, gender, or other characteristics. In Chapters 6 and 7, I found these characteristics to be poor predictors of referrals so that assuming homogeneity is considered reasonable.

If the patient gets referred to a dermatology appointment, she joins a second waiting list. There are two ways to exit this list: by seeing a dermatologist or by exiting before the appointment. Little's law can be used to determine the average waiting time spent in the waiting list (Little and Graves, 2008). It states that under steady state conditions (i.e., when the number of entrants roughly equals the number of exits), the number of patients on the list equals the number of entrants (or indeed the number of exits) multiplied by the average waiting time. Since the first two are generally known, the average waiting time can be calculated. Note that there are an infinite number of distributions of individual waiting times with the same aggregate average waiting time (Sterman, 2000). In other words, it is not possible to determine how long each individual patient in the list waits. As such, I assume patients queue under a *first come first served* rule, so that the time each patient spends in the waiting list is exactly the average waiting time estimated using Little's law.

My main focus is on the dermatology waiting list, how it evolves, and how it affects, and is affected by, other elements in the system. Figure 8-1 is a useful starting point, but it does not explain how the rates of flow and stocks interact. In the previous chapters, I identified two ways in which waiting time affects utilisation: through referrals and non-completion. The model in Figure 8-1 can be extended to incorporate two feedback loops representing these two effects of waiting time. The resulting model can be seen in Figure 8-2, depicting the dermatology outpatient pathway prior to the introduction of teleconsultation.

It is assumed that referrals are determined by a rate, which is determined endogenously as a function of the waiting time. I adopt van Ackere and Smith's formulation of adaptive expectations (van Ackere and Smith, 1999). Patients and physicians have an expectation of waiting times based on previous waiting times. Because they become aware of the actual waiting time through an imperfect process subject to delay, they adapt their expectations only after some time elapses (the time to perceive the waiting list). Their perception of the waiting time then affects the rate of referrals through the influence of the elasticity of referrals with respect to waiting time (itself a function of the perceived waiting time). The referral rate represents the outcome of a complex shared decision making process involving the preferences of both patients and physicians. A formula for the elasticity of demand with respect to waiting time was presented in Chapter 6 (equation 6-1). The same formula can be used to determine referrals:

$$dQ = E_D(f(W_t))Q \frac{df(W_t)}{f(W_t)} \quad (8-1)$$

where Q is the number of referrals per 1000 inhabitants, $E_D(f(W_t))$ is the elasticity of referrals with respect to waiting time, as a function of perceived waiting time $f(W_t)$, and $df(W_t)$ is the change in perceived waiting time. As van Ackere and Smith (1999), I model the perceived waiting time as an exponential smoothed average of past waiting times, with the reciprocal (multiplicative inverse) of time to perceive as the smoothing factor. As indicated by the letter B in loop B1, this is a *balancing* loop: increases in waiting time will cause the referral rate to drop and therefore waiting time to decrease, all else being equal.

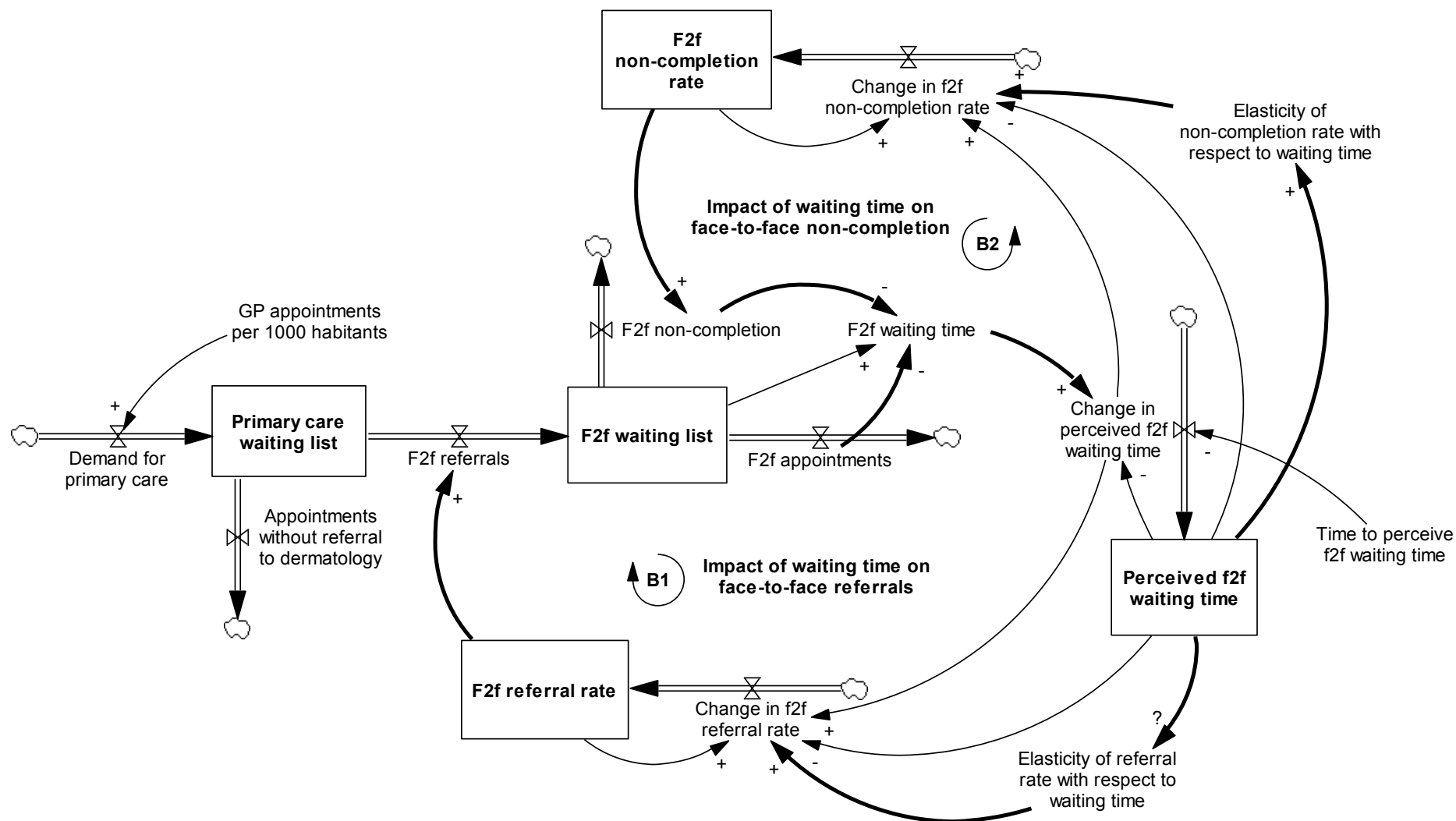


Figure 8-2 The impact of waiting time on patient flows before the introduction of teledermatology (F2f: face-to-face)

The non-completion loop B2 is also a balancing loop and mirrors the referral rate loop in every way except two: the elasticity and the rate are naturally different. While loop B1 captures the effect of waiting time on entry into the list, the non-completion loop represents the impact of waiting time on those patients already in the list. It accounts for a number of possibilities: getting better, getting worse, seeking alternatives, etc. Finally, it is important to note that both the referral rate and non-completion rate are average aggregate rates. There is no conception of heterogeneity.

Even before the introduction of teleconsultations, it is clear there are feedback effects at play. As illustrated by Smith and van Ackere (2002), changes in waiting times can lead to changes in referrals and supply which are not negligible. After the introduction of teledermatology, the picture becomes somewhat more complex (see Figure 8-3). To begin with, there is a new balancing loop B3: the impact of continued GP education and learning on the referral rate. As mentioned in Chapter 4, numerous studies have suggested that GPs who participate in teleconsultations experience a learning effect which increases their ability to better manage patients in primary care, thus reducing referrals to secondary care (Nordal et al., 2001, Jarvis-Selinger et al., 2008, Taylor, 2005). This effect was also mentioned in the exploratory interviews in Chapter 5.

In order to operationalize the impact of continued education on the referral rate, a few modelling decisions must be made. First, I assume that learning is an individual-level process that is not easily transferred from GP to GP. In other words, GPs who practice teledermatology cannot transfer their experience to GPs in their practice who do not. Second, there is a limit to how much referral rates can be reduced, reflecting the fact that some conditions (e.g., melanoma) will always require a referral. Third, there is a certain amount of experience needed to achieve the maximum reduction possible, and it is a function of how many teleconsultations have been performed (in accordance with findings from the exploratory interviews).

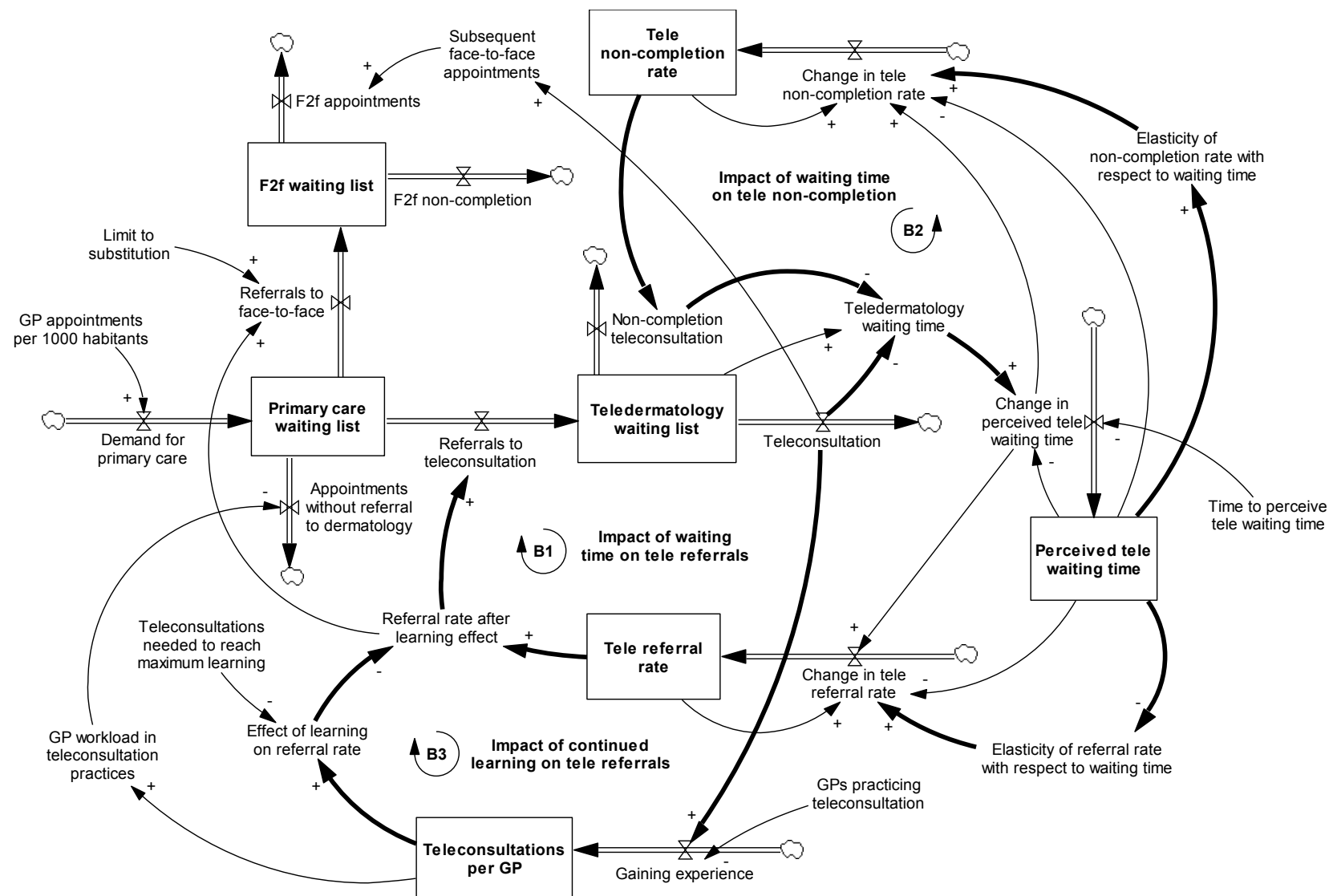


Figure 8-3 Dermatology outpatient care after the introduction of teledermatology (Tele: teleconsultation; F2f: face-to-face)

Finally, how the number of teleconsultations affects the referral rate is described by the curve *effect of learning on referral rate*. It is important to note that the reduction in referral rate is also at the individual-level. The referral rate depicted in Figure 8-3 is an aggregate rate which must account for the possibility that not all GPs in the practice participate in teleconsultations, and those that do not have higher referral rates.

The presence of GPs in teleconsultations may lead to more educated physicians, but it also leads to busier GPs. Assuming, as I do, that GP practices operate at capacity, then time spent in teledermatology must come at the expense of time spent on activities related to primary care. In the model, GPs that practice teleconsultation reduce their primary care consultations by the same amount of time. Below, I explain why this does not affect the flow of referrals to face-to-face dermatology or teledermatology, making use of actual data from Alentejo.

Another important aspect is that a number of patients from teleconsultation practices will still be referred directly to a face-to-face appointment (e.g., when a severe condition is suspected). The percentage of patients who still get referred to face-to-face appointments is governed by the exogenous variable *limit to substitution*. It is, very simply, the percentage of face-to-face referrals relative to all referrals from that practice. Furthermore, a percentage of teleconsultations will lead to subsequent face-to-face appointments. In accordance with the exploratory interviews, these patients do not get added to the face-to-face waiting list, but rather get booked directly for an immediate appointment, circumventing the waiting list. For simplicity, I assume that subsequent face-to-face appointments take place immediately after the teleconsultation, and that for each subsequent appointment one less appointment is available for waiting list patients. In the following sections, I discuss some of the causal mechanisms included in the simulation model in more detail.

8.2.1 Primary and secondary care capacities

Both primary and secondary care capacities are exogenous. These are represented in Figure 8-4 as *Face-to-face capacity*, *Primary care capacity*, and *Teledermatology capacity*. Before the introduction of teledermatology, teleconsultation capacity is null, and the variable *Actual face-to-face capacity* and *Actual primary care capacity* are equal to *Face-to-face capacity* and *Primary care capacity* respectively. After the introduction of teledermatology, part of face-to-face capacity is diverted to teleconsultations. How much capacity is diverted is a policy lever (I will show below how different assumptions concerning the allocation of face-to-face capacity to teledermatology can lead to significantly different outcomes, especially depending on the scale of use of teleconsultations in the region).

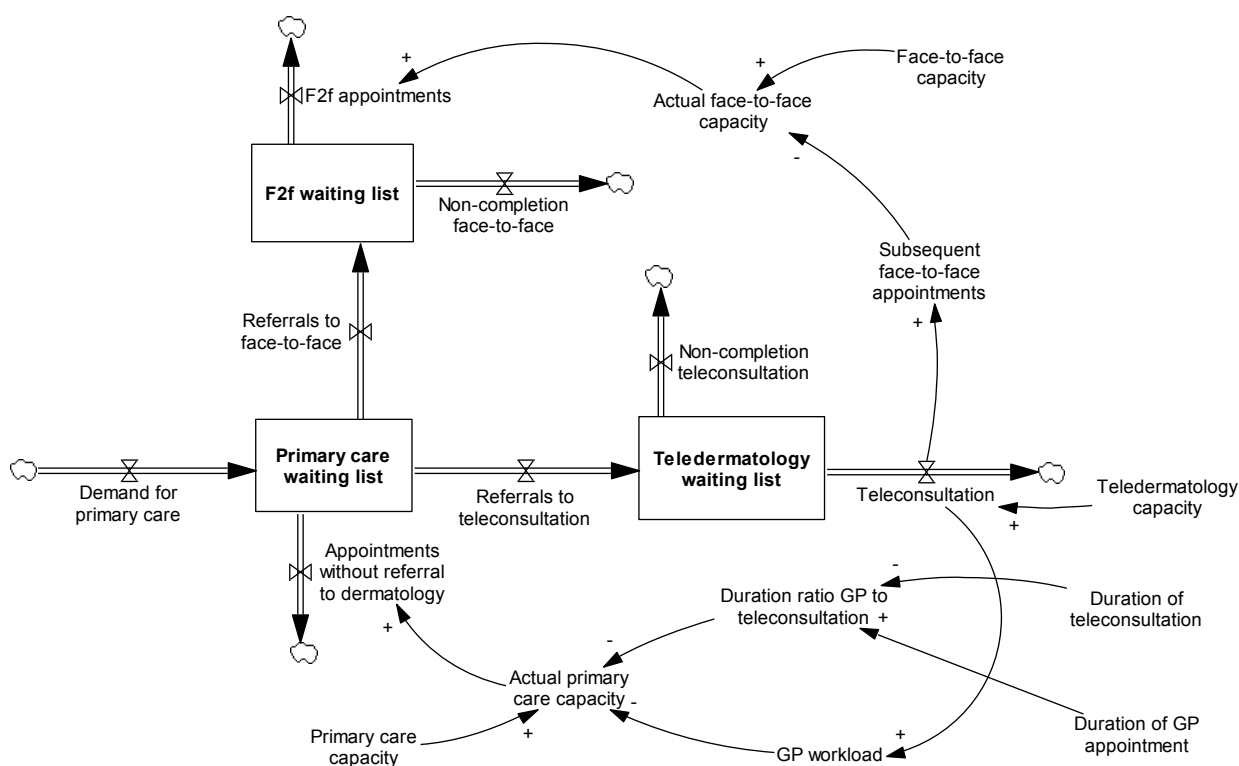


Figure 8-4 Primary and secondary care capacities after the introduction of teledermatology

As the number of teleconsultations increases, GP workload increases as well causing the *Actual primary care capacity* to become a fraction of the previously available *Primary care capacity*. Similarly, subsequent face-to-face appointments cause *Actual face-to-face capacity* to be smaller than *Face-to-face capacity*. The magnitude of these reductions is

determined by duration ratios, which take into account the fact that teleconsultations, face-to-face appointments and GP appointments all take different amounts of time.

8.2.2 Implementation of teleconsultation

It is unrealistic to assume that all referrals in a specific practice will simply shift to teleconsultations overnight. Implementation takes time, as both patients and physicians adapt to the new pathway. It is likely that the number of referrals to teleconsultation will increase slowly in the beginning, and eventually progress towards the previously mentioned limit to substitution. I represent this process as s-shaped growth towards the limit to substitution (see Figure 8-5).

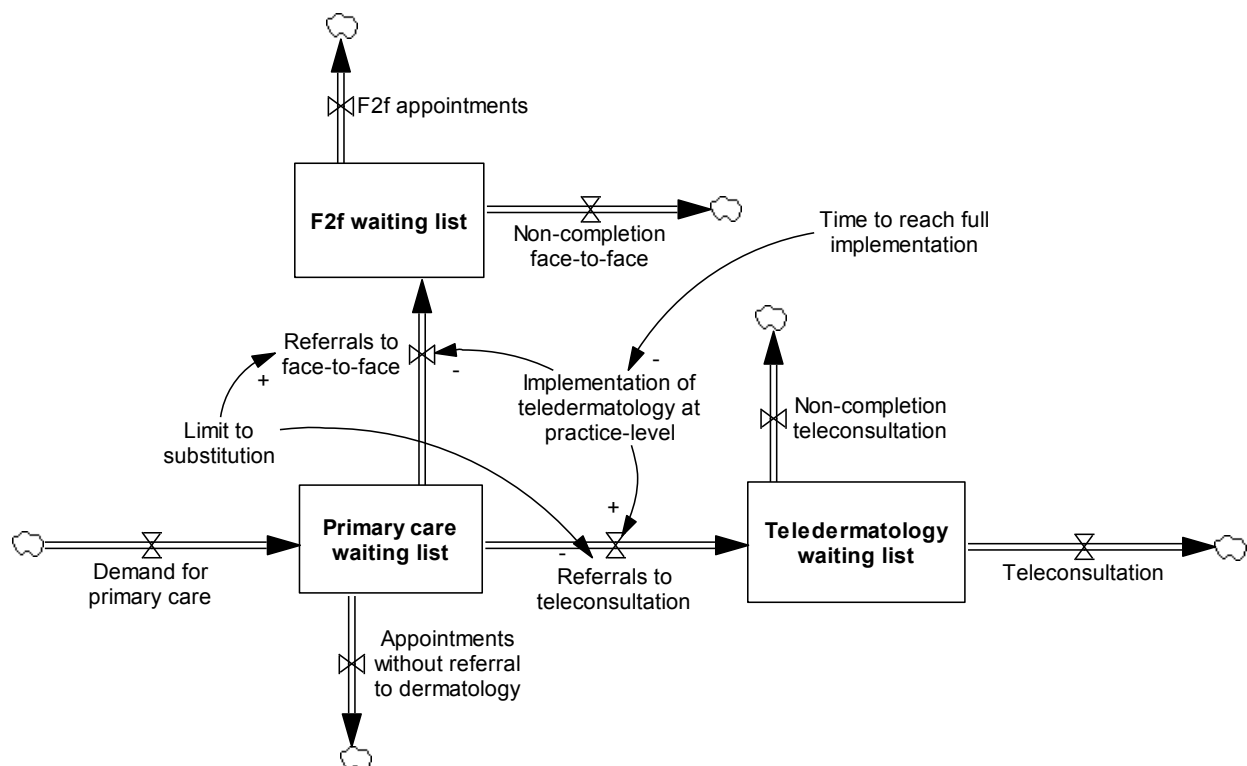


Figure 8-5 Implementation of teledermatology at practice-level and destination of referrals

The variable *Implementation of teledermatology at practice-level* is a lookup function of an s-curve (it is specified as a 3rd order smooth of a step function of value 1, with smoothing constant of 1). The parameter *Time to reach full implementation* can be used to test the sensitivity of the formulation to different implementation times (it essentially changes the

smoothing constant in the s-curve). I assume there are no negative productivity effects in the early stages of implementation (e.g., the duration of the teleconsultation is the same irrespective of the stage of implementation).

8.2.3 Continued GP education and referral rates

There are a number of important elements in the specification of the impact of continued GP education and learning on the referral rate (see Figure 8-6).

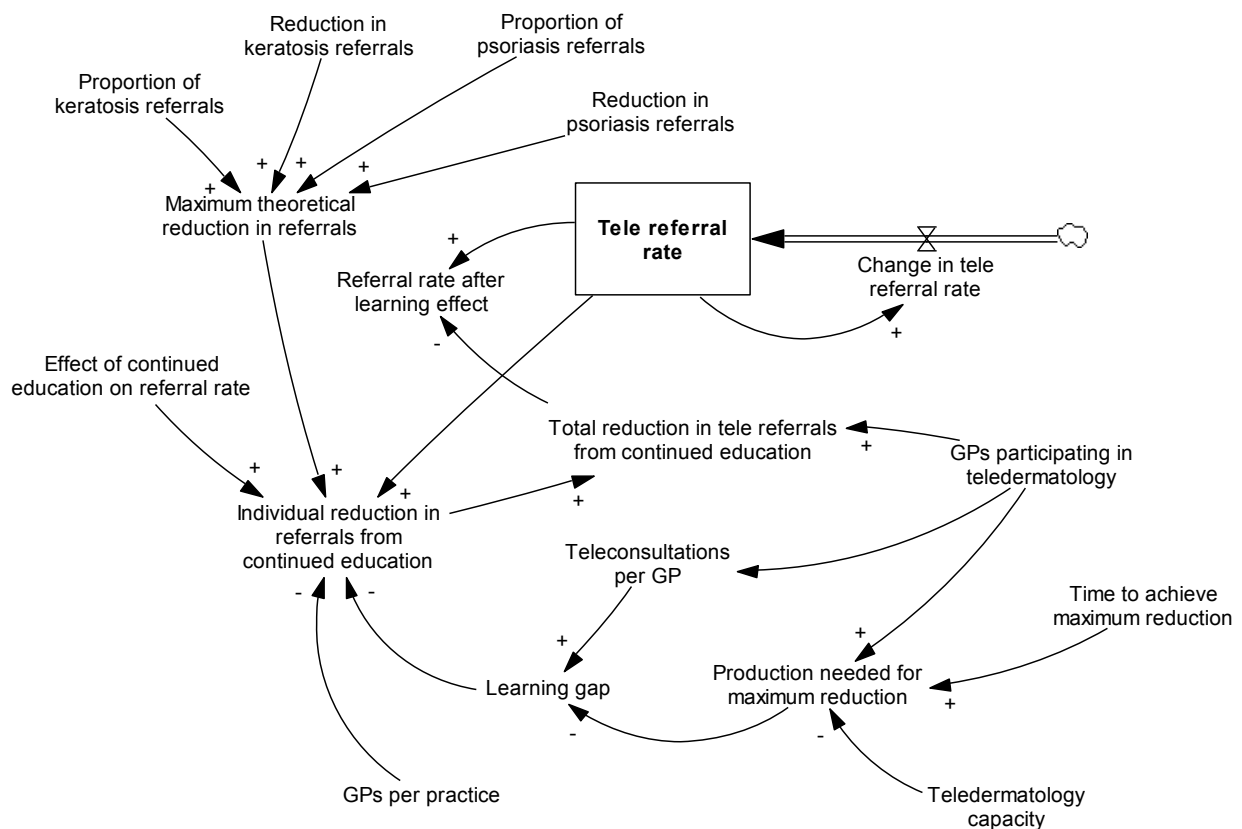


Figure 8-6 Impact of continued GP education and learning on referral rate

First, the *Maximum theoretical reduction in referrals* represents the fact that a number of clinical presentations will always need to be referred. Taking advantage of the empirical study in Chapter 6, I have modelled reductions in referrals for patients presenting with psoriasis and keratosis. The aggregate maximum reduction possible is a function of both the proportion of patients with these presentations that a GP sees and the reduction in probability of referring these patients given the GP participates in teleconsultations. It is

important to note that this may be an underestimation of the actual reduction in referrals, since it is likely that other conditions besides keratosis and psoriasis are likely to benefit from continued education.

Second, the reduction effect is at the level of the individual GP. As such, it is necessary to estimate the number of teleconsultations performed by each participating GP, so as to determine individual experience. Third, there is a certain amount of experience needed to achieve the maximum reduction possible, and it is a function of how many teleconsultations have been performed. The variable *Production needed for maximum reduction* (in patients) is simply the product of the *Time to achieve maximum reduction* (in months) and the *Teledermatology capacity* (in patients/month). The *Learning gap* (dmnl) is defined as the ratio between the actual experience (i.e., cumulative production) of each participating GP and the production needed to achieve the maximum reduction.

The *Effect of continued education on referral rate* is an s-shaped lookup function (as the implementation curve, it is specified as a 3rd order smooth of a step function of value 1, with smoothing constant of 0.4), that takes the *Learning gap* as an input and produces a number between 0 and 1 as the output (0 being no reduction and 1 being maximum reduction possible). The s-shape adequately illustrates the process through which learning occurs, initially slow, growing rapidly after a few months and eventually reaching a limit. The *Individual reduction in referrals from continued education* is then the product of the output of the lookup function and the *Maximum theoretical reduction in referrals*, in turn multiplied by an adjusting factor: the *Referral rate* divided by the *GPs per practice*. This adjustment factor represents the fact that only the referral rate of participating GPs is reduced. Finally, the aggregate referral rate *Referral rate after learning effect* is simply the *Referral rate* minus the individual referral rates of GPs participating in teledermatology. As previously mentioned, I have not allowed for the possibility of knowledge transfer (i.e., GPs who practice teledermatology cannot transfer learning to GPs who do not practice in teledermatology).

8.3 Specification and parameters

The model was implemented in Vensim™ (Ventana® Simulation Environment). Comprehensive documentation for the full model was produced according to best practices (Martinez-Moyano, 2012) and is available in Appendix B. The time unit used in the simulations is a month (time step is 0.03125 of a month, which means the simulator estimates the values of all variables in the model approximately every simulated day). The model can be simulated at both practice-level and regional level, depending on the choice of parameters. Given the boundaries of the modelling effort, I defined a number of variables as constants or lookup functions dependent on simulated time (lookup functions are user-defined functions, whose output y is defined as a function of a specific input x , often simulated time).

Simulations were run for a period of 132 months (or 11 years). This period was chosen as the expected lifetime of teledermatology platforms is 10 years, a period after which the equipment should be replaced. In all simulations, practices start using teledermatology 12 months into the simulation, resulting in 1 year of only face-to-face care and precisely 10 years of teledermatology. Naturally, it is straightforward to run the simulations for a larger period of time (e.g., decades) and define the equipment be replaced every 10 years. This would affect costs but would not affect behaviour of the system, as discussed in Section 8.6.1. In the following sections, results are presented for a simulation period of 132 months. Behaviour over longer periods is discussed whenever appropriate. While the choice of simulation period introduces a somewhat arbitrary cut-off point in results tables, numerous graphs are provided so that the evolution of variables can be observed and potential patterns detected.

8.3.1 Simulation parameters

Numerous costs and consequences can be explored in the model. The approach in terms of costs was to estimate total costs as a function of unit costs for patients and their families, the Portuguese NHS, and society (defined as the sum of the two former). Table 8-1 contains the unit costs, including sources, used in the simulations, plus a number of auxiliary variables used to estimate total costs (e.g., distances and means of transport).

It is important to note that, at present, both the hospital and the primary care practice get reimbursed for teleconsultations at €83.79 each, so that each teleconsultation costs the NHS a total of €167.58. This is considered a financial incentive for practices to implement the technology, along with a lower co-payment for teledermatology. If one adds the investment and variable costs with the video-conferencing equipment, then it is hard to see how teleconsultation can be cheaper than face-to-face care from the perspective of the NHS. This issue is discussed in detail later.

The teleconsultation equipment used for teledermatology is used for a number of other specialties and purposes (see Chapter 5). As such, using the total costs of the equipment in the evaluation of teledermatology consultations leads to an overestimation of true costs. I assume that the ratio between teleconsultations in dermatology, and all teleconsultations performed remains at the average value between 2002 and 2012 (i.e., 31%), and include that percentage of total equipment costs in the evaluation.

In terms of consequences, I focus on process outcomes and intermediary effects, such as avoided waiting time and increased utilisation. Unfortunately, the scarcity of data does not allow the inclusion of final health outcomes in the model. Outcomes of interest include: unmet need/demand (patients), avoided referrals due to continued education (patients), avoided travelling due to difference in distances to teledermatology and the hospital (kilometres), avoided waiting due to differences in waiting times for teledermatology and face-to-face appointments (patient*days), total number of patients seen (patients), referral

rate per 1000 inhabitants, and capacity available (patients) per 1000 inhabitants, before and after the introduction of teledermatology.

Table 8-1 Unit costs and auxiliary costing variables used in the model

Parameter	Unit	Value	Source*
Median distance to hospital	Kilometre	59.54	Data set
Mean distance to practice	Kilometre	5.87	Survey
Median monthly wage	EUR	755.50	INE, I.P.
Cost of video-conferencing platform	EUR	9723.08	ARSA, I.P.
Maintenance costs, per platform per year [‡]	EUR	972.31	Estimate
Allocation of equipment to teledermatology	Dmnl	0.31	ARSA, I.P.
Fraction face-to-face patients using car	Dmnl	0.80	Survey
Fraction face-to-face patients using bus	Dmnl	0.08	Survey
Fraction face-to-face patients using taxi	Dmnl	0.01	Survey
Fraction face-to-face patients using ambulance	Dmnl	0.11	Survey
Fraction teleconsultation patients using car	Dmnl	0.57	Survey
Fraction teleconsultation patients using bus	Dmnl	0.02	Survey
Fraction teleconsultation patients using ambulance	Dmnl	0.02	Survey
Fraction teleconsultation patients walking	Dmnl	0.39	Survey
Share of population working	Dmnl	0.17	Survey
Teleconsultation patients being accompanied	Dmnl	0.46	Survey
Face-to-face patients being accompanied	Dmnl	0.67	Survey
Teleconsultation share of companions working	Dmnl	0.35	Survey
Face-to-face share of companions working	Dmnl	0.60	Survey
Time to attend face-to-face appointment	Minute	180	Survey
Time to attend teleconsultation	Minute	90	Survey
Face-to-face co-payment	EUR	7.50	ACSS, I.P.
Teleconsultation co-payment	EUR	3.00	ACSS, I.P.
Reimbursement, first appointments	EUR	83.79	ACSS, I.P.
Reimbursement, subsequent appointments	EUR	76.17	ACSS, I.P.

* Sources: INE, I.P. (Statistics Portugal); Data set (see Chapter 6); Survey (see Chapter 5); ARSA, I.P. (Alentejo Regional Health Authority); ACSS, I.P. (Central Administration of the Health System). [‡] Maintenance estimate as 10 percent of investment.

Using simulated data on costs and consequences, it is possible to estimate an average cost per patient (this is different from the unit costs since fixed costs are diluted across patients) and calculate cost per consequence measures (e.g, how much one less day a patient waits for an appointment costs the NHS). It is essential, however, to interpret the results with caution. The point is not to precisely predict cost-outcome ratios, but rather to describe how costs and consequences are likely to evolve under different scenarios and assumptions.

As mentioned, previous studies often make statements at a broader level based on findings from practice-level analyses (see, for example, Cusack et al. (2008)). Simulations can be performed to test whether practice-level assumptions actually extend to a wider context. Table 8-2 describes the parameters used in both practice and regional simulations. The difference between the two lies in the populations. As seen in Table 8-2, the median population per practice between 2002 and 2013 was 7,442 inhabitants, while the total average population for the same period was 297,718. At the practice-level model, the population in the practice is naturally 7,442 with the population at other practices set at 297,718 minus 7,442 equals 290,276. At the regional level, the population can be set at any value for up to three practices, the rationale is the same: three practices with 7,442 inhabitants and all other practices with 297,718 minus 22,326 equals 275,392. Figure 8-7 illustrates the model at the regional level.

Table 8-2 Parameters used in both practice-level and regional simulations

Parameter	Unit	Value	Source*
Median initial referral rate, per 1000	Patient/People	0.2738	Data set
Median initial non-completion rate, per 1000	Patient/People	0.0827	Data set
Median GPs, per practice	Physician	6	INE, I.P.
Average population in region, (2002-2013)	People	297,718	INE, I.P.
Median population, per practice (2002-2013)	People	7,442	INE, I.P.
GP demand per 1000, per year	Patient/People	3,671	INE, I.P.
Monthly dermatology capacity, total	Patient/Month	82	Data set
Initial face-to-face waiting time	Day	187	Data set
Time to perceive, face-to-face	Month	12	Literature
Time to perceive, teleconsultation	Month	6	Assumption
Reduction in psoriasis referrals from learning	Dmnl	0.42	DCE
Proportion of psoriasis referrals	Dmnl	0.16	HESE, E.P.E.
Reduction in keratosis referrals from learning	Dmnl	0.18	DCE
Proportion of keratosis referrals	Dmnl	0.18	HESE, E.P.E.

* Sources: Data set (see Chapter 6); INE, I.P. (Statistics Portugal); DCE (Discrete choice experiment, see Chapter 7); ACSS, I.P. (Central Administration of the Health System); Literature (van Ackere and Smith (1999)); HESE, E.P.E. (*Hospital do Espírito Santo de Évora*).

Feedback effects have been omitted for readability in Figure 8-7, yet they are still present as in Figure 8-2 and Figure 8-3.

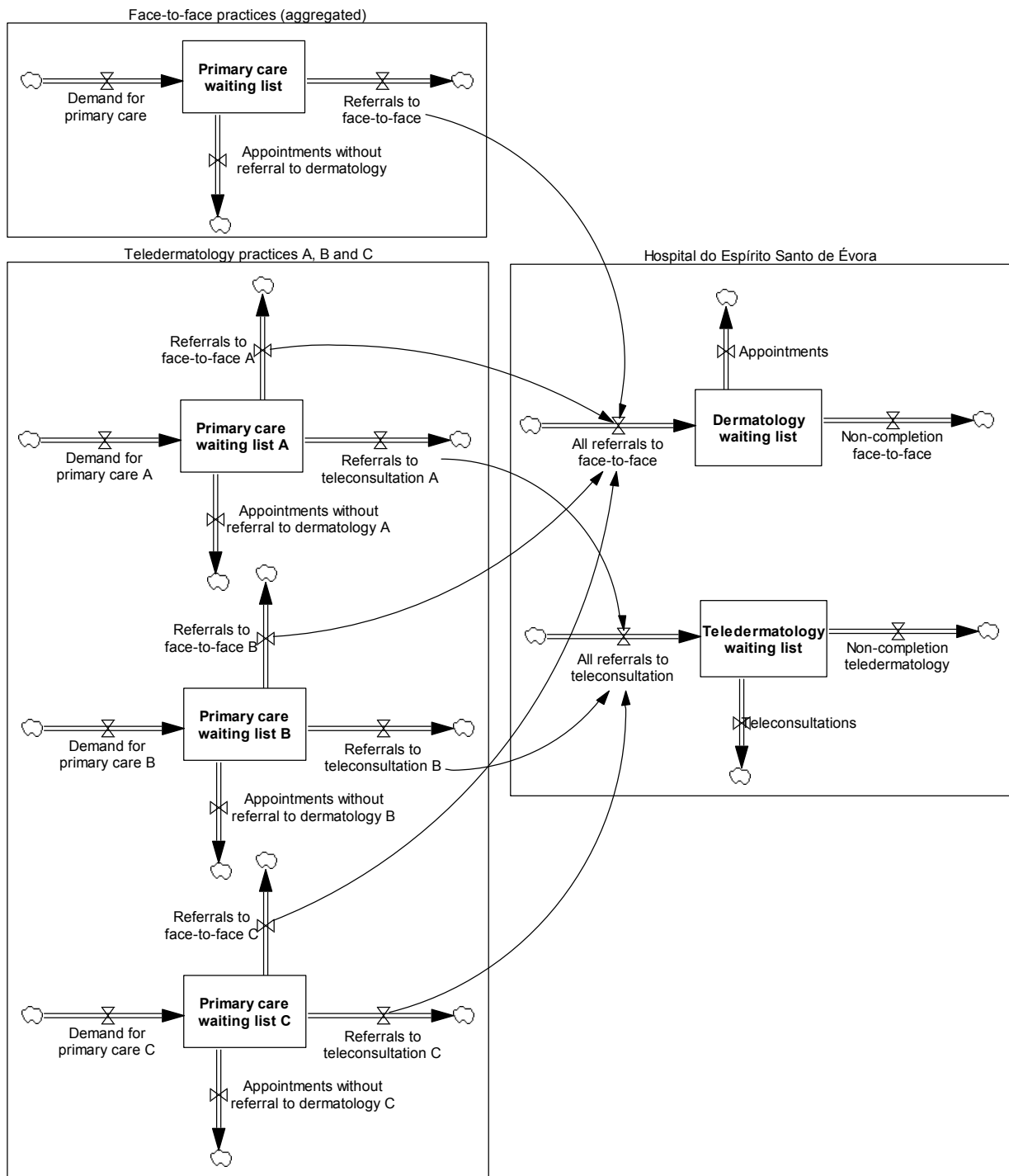


Figure 8-7 Stocks and flows at the regional level (feedback effects omitted for readability)

Each practice (or group of practices, in the case of face-to-face) reacts to the waiting time (teledermatology waiting time for practices A, B and C, and face-to-face waiting time for other practices). Each practice adapts its referral and non-completion rates accordingly. Each practice determines its monthly referrals by multiplying the referral rate by its population. All those numbers are added to the common hospital-level waiting list. The non-

completion flow from the hospital-level waiting list is calculated in the same way: by adding all the products of individual rates and populations. Teledermatology practices each have their own learning feedback loops and specific implementation dates can be chosen. For example, one practice may implement teledermatology at the beginning of the simulation, another at month 12, and the third a year later. This will naturally lead to different experience levels and different perceptions of waiting times at different practices. While three teledermatology practices may sound insufficient, the same aggregation strategy used for face-to-face practices can be used for the three teledermatology practices. That is, each practice can serve a different number of inhabitants.

8.4 Model validation and testing

Methods for validating system dynamics models (and simulation models, in general) have been the subject of much discussion (Barlas and Carpenter, 1990, Barlas, 1996, Lane, 1995, Landry et al., 1996, Kleijnen, 1995, Forrester and Senge, 1978, Mass and Senge, 1978, Sterman, 1984).

8.4.1 Validation in the context of system dynamics

While a cornerstone of econometric analyses, tests of significance and goodness of fit have been largely disregarded by system dynamicists. There are many reasons for this. A system dynamics model essentially embodies causal hypotheses about how different variables interact and produce certain patterns of system behaviour. For example, van Ackere and Smith (1999) propose that an increase in waiting time will cause demand to drop, holding other variables constant. While the pattern produced by their model may apply to a number of different health authorities across England, it does not capture all the historical idiosyncrasies of each authority, making point-prediction difficult at best, useless at worst. For example, Forrester shows that the use of random inputs in a system dynamics model significantly undermines tests of significance, yet it may not alter the model's policy

implications, in certain cases. Furthermore, any data can be fitted to any degree of precision (Sterman, 1984). Validating a system dynamics model by fitting its output to observed data is not necessarily useful, as there are many other models that produce a good fit using implausible (at the limit, absurd) causal relationships (Lane, 1995).

The technicalities of statistical testing have also been questioned. Frequently there are many variables of interest in a system dynamics model (e.g., in my model: waiting time, referrals and non-completion), leading to a simultaneous multiple hypotheses problem (Barlas, 1996). The choice of significance level should reflect a judgement about which is worse: to reject a valid model or to accept an invalid one. The ubiquitous choice of 0.05 is arbitrary and not necessarily useful for every situation. Finally, statistical tests often require that certain conditions hold (e.g., assumption of normality with zero mean and perfect measurement of data). Statistical tests are then as much a test of whether a model is valid as they are tests of underlying assumptions. Namely, a failed significance test indicates a parameter cannot be estimated reliably given available data, not that the variable is unimportant (Mass and Senge, 1978).

For all these reasons, statistical measures of historical fit have been considered weak tests of validity, very rarely undertaken by system dynamicists (Sterman, 1984). This has led to criticism, notably from econometricians, that system dynamics models are of dubious practical usefulness, and that assessing whether fit is merely *close enough* lacks rigour. For some years, it seemed these differences were irreconcilable (Meadows, 1980), but more recently there has been an acceptance that different philosophical approaches can help answer different questions, and may have their own merits (Radzicki, 1990). Barlas and Carpenter (1990) discuss the implications for model validation of adopting two different philosophical approaches: one which sees models as objective representations of reality, and another which considers models as just one possible description of reality. In the former, a model is either valid or not. Regarding the latter, they write:

A valid model is assumed to be only one of many possible ways of describing a real situation. No particular representation is superior to all others in any absolute sense, although one could prove to be more effective. No model can claim absolute objectivity, for every model carries in it the modeller's world view. Models are not true or false but lie on a continuum of usefulness. Model validation is a gradual process of building confidence in the usefulness of a model; validity cannot reveal itself mechanically as a result of some formal algorithms. (Page 157, Barlas and Carpenter (1990))

The idea that validation is a process of building confidence is central to validation in system dynamics. Models serve specific purposes, and validity can only be assessed in the context of a specific purpose or problem (Forrester and Senge, 1978). Tests of historical fit may be nonsensical when models are theoretical or pedagogic, if not impossible when data are unavailable. Tests of goodness of fit are essential when accuracy and calibration are important, but are less useful when exploring higher-level longer-term behaviour, when randomness and idiosyncrasies are less relevant.

Forrester and Senge (1978) propose three types of validation tests: tests of model structure, tests of model behaviour and tests of policy implications. Tests of model structure seek to build confidence that the model adequately represents the structure of the real system, given existing knowledge. Tests of model behaviour include historical fit, but also tests of extreme conditions and sensitivity to initial values. Barlas (1996) suggests that behavioural fit should only be assessed after structural testing, as there is nothing to gain from testing the fit of an implausible model. Finally, tests of policy implications seek to build confidence that the simulation findings are robust to different assumptions (e.g., different parameters).

8.4.2 Validation tests

There are a number of reasons for being confident in the structural validity of the model presented here. All three feedback mechanisms in the model are based on previous theoretical and empirical findings, confirmed in my own empirical work (exploratory interviews and econometric analyses). The *impact of waiting time on referrals* feedback loop is a replica of a loop present in van Ackere and Smith (1999). The use of an s-shaped curve for the effect of continued GP education and learning is also with precedent (Naim and Towill, 1993), and adequately reflects there are limits to reducing referrals. As for the parameters used in the model, they are virtually all based either on data analyses undertaken in Part 2, or on previous literature (see Table 8-1 and Table 8-2 for specific sources of data). Finally, adding to the face validity of the model's structure is the fact that it has been presented, in various forms and development phases, in scientific conferences and meetings, including several presentations to both physicians and senior managers in Alentejo.

With regards to model behaviour, I conducted a number of simulations to test the impact of different parameters, and shapes of implementation and learning curves, on the outcomes. Three different shapes were used to represent the *Effect of continued education on referral rate*: a 1st, 3rd (the default) and 9th order smooths of a step function of value 1, with smoothing constants of 0.4. The same shapes were used for the *Implementation of teledermatology at practice level* but with smoothing time of 1 (equivalent to approximately 3 months). Different times to achieve maximum effects were tested. The results were as expected: shorter and longer times to reach the maximum reduction in referrals and the maximum substitution effect, yet same general qualitative system behaviour. The parameter *Time to perceive waiting time* was changed from a maximum of 1 year to a minimum of 1 month, also with predictable effects: a quicker adaptation to past waiting times in the latter, but generally the same qualitative behaviour in both (this is discussed in greater detail

below). As seen in Table 8-2, I chose 6 months for the *Time to perceive waiting time* after the introduction of teleconsultation as it is expected that information delays are less important once appointments start taking place in the practice.

Finally, simulations were performed to explore behaviour under extreme conditions and test sensitivity to initial values. Results were as expected. When capacity is set to null, the waiting time and waiting list increase quite considerably, but eventually reach a plateau, as referrals adjust towards zero. This happens in both the face-to-face and teledermatology pathways. Unlike van Ackere and Smith (1999), elasticities in this model do not go to zero for longer waiting times. While in their model there is a residual demand for NHS regardless of waiting time, in my model it is possible for referrals to adjust to zero if capacity is null. Another test is to set the population at 10 times its original value. It is the demand side that is driving the results now. But unlike in the previous scenario, there is capacity. As such, referrals (which are determined by a referral rate per 1000 inhabitants) adjust towards a value that can be met by existing capacity. Before the new equilibrium is reached, there is a significant increase in the waiting list and in non-completion, as referrals adapt to the new situation. Unmet demand increases considerably. When waiting lists are set to have initially no patients, the system reaches equilibrium quite rapidly (the actual waiting times are discussed below in more detail).

With regards to historical fit, I encountered similar problems as van Ackere and Smith (1999), with respect to data availability and concurrent policy changes. For one, there are no available data for non-completion and referrals in teledermatology (only number of teleconsultations). Data on teledermatology waiting lists and times are based on exploratory interviews or self-reported by patients. No time-series data are available against which to test fit. Furthermore, there are very few variables in the model to characterise different practices. While this probably does not affect pattern behaviour, it does affect goodness of fit. Also, even in the relatively short period under study, the regional health system has seen considerable change. In 2003, many hospitals, notably the *Hospital do Espírito Santo de*

Évora, were given a status similar to that of public-interest companies, with management boards changing (Barros et al., 2011). Hospitals in Portugal are financed through contracts which, among other things, stipulate target waiting. In 2006, the concept of *Unidades de Saúde Familiar* – multidisciplinary teams at the primary care level who are partially paid through incentive mechanisms – was introduced. In 2008, groups of primary care practices were created. None of these events are accounted for in the model.

I experimented with different values for the implementation curve and time to perceive the waiting time, achieving a statistically significant fit for the evolution of the referral rate in *Vila Viçosa*, a practice in Alentejo. However, since these variables are exogenous and not independently verifiable, such a test adds little to the validity of the model. For example, in the validation of their model, van Ackere and Smith (1999) replicated behaviour for the Local Health Authority of Sunderland, by determining a “*plausible growth rate for potential demand and exogenous increase in beds (...) determined to match the pattern of the observed data*” (page 250), and acknowledging that their model did not include policy changes at the regional and national level. As the authors state, there is no reliable data to verify the rates used to fit the data. The model presented here performs well under extreme conditions and is built on previous knowledge, verified in empirical analyses (both in Alentejo and other settings). This contributes far more towards achieving confidence in the model.

8.5 Simulations and scenarios

Simulations were run at both the individual practice and regional levels. The base case, against which other cases are compared, is the use of face-to-face appointments. Different values and assumptions regarding potential policy levers are explored.

8.5.1 Practice-level simulations

The first set of simulations is intended to illustrate how the waiting time, referrals and non-completion evolve after the introduction of teledermatology in a median Alentejo practice

(practice A). Before the introduction of teledermatology, the system is in equilibrium. The initial waiting time is constant at 187 days, with the number of patients exiting the list (through completed and missed appointments) equal to the number of entrants (i.e., referrals). Primary care is also in steady-state equilibrium, with realised demand (number of monthly entrants to the GP waiting list) met by GP capacity. Now assume that practice A is the first practice to implement teledermatology in the region, doing so one year into the simulation. Every month, dermatologists allocate 1 hour of their schedule to perform teleconsultations. Because of the shorter duration of teleconsultations, dermatologists are able to see 10 patients in those 60 minutes, double what they could do in face-to-face.

Since there is no waiting list for teledermatology, the actual waiting time is initially zero (see Figure 8-8). The perceived waiting time starts at 187 days but decreases steadily to 100 days at the end of the simulation period (see Figure 8-9). This is still far from the actual waiting time which, at year 2, is about 15 days. Naturally, GPs should be aware of the actual waiting time, yet using 6 months for the variable *Time to perceive waiting time* in practice A is still deemed appropriate because patients take longer to become aware of these changes. As patients and GPs perceive the improvement in waiting time, two things happen: the referral rate increases (Figure 8-10) and the non-completion rate decreases (Figure 8-11). The true waiting time for teledermatology is 15 days throughout the simulation.

It is important to explain briefly why the teledermatology waiting time stabilises around 15 days and not some other value. I assume teleconsultations are performed once every month, as capacity is only 1 hour per month. It follows that the average lowest waiting time (assuming a normal distribution) is 15 days, or half a month (a patient referred just after teleconsultations take place will have a waiting time of one month, while the patient referred just before teleconsultations will have a waiting time close to zero).

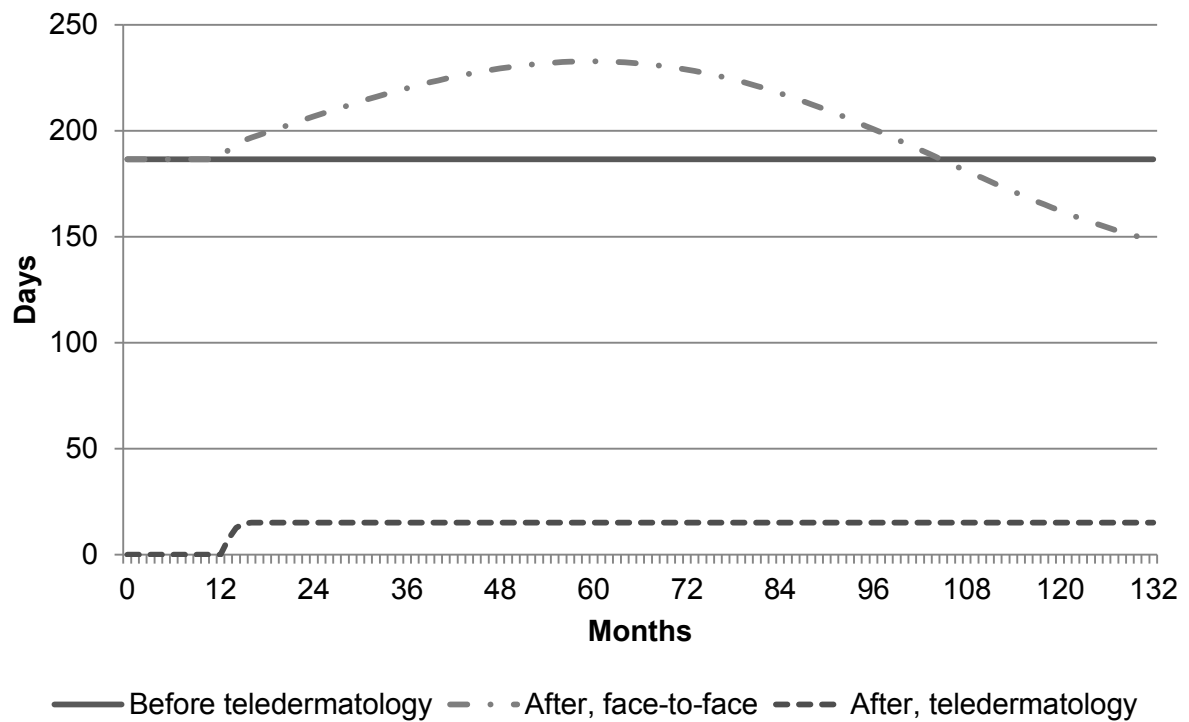


Figure 8-8 Actual waiting times for dermatology appointments

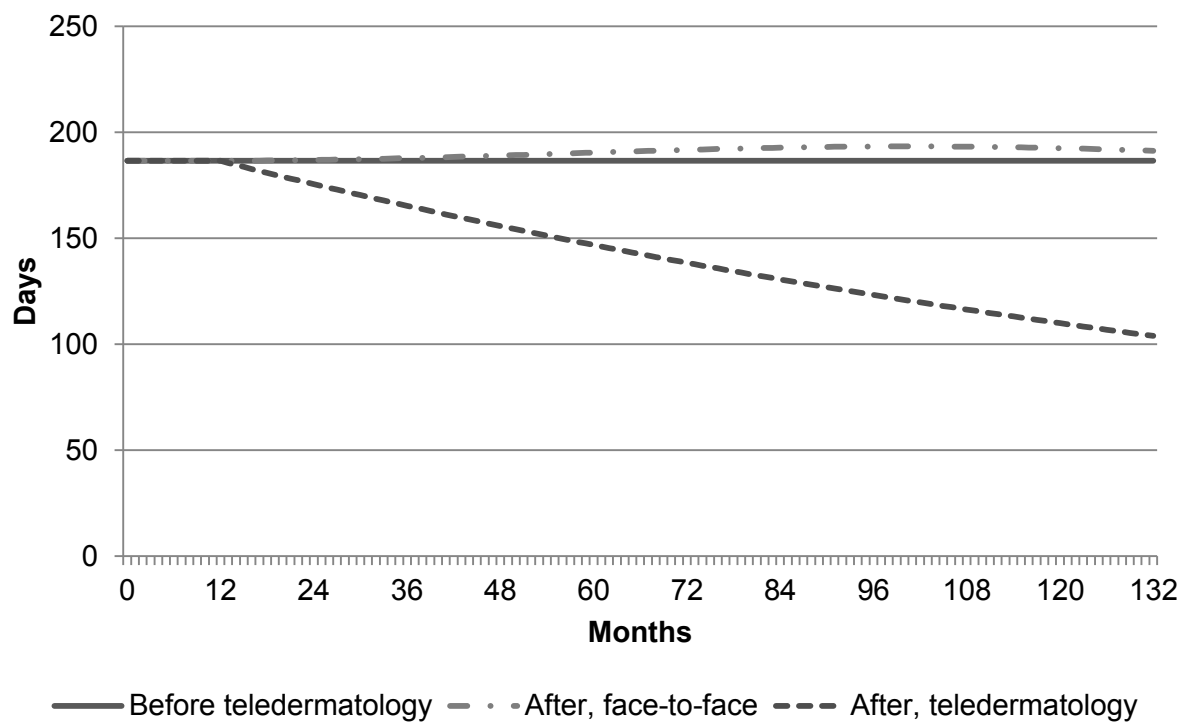


Figure 8-9 Perceived waiting times for dermatology appointments

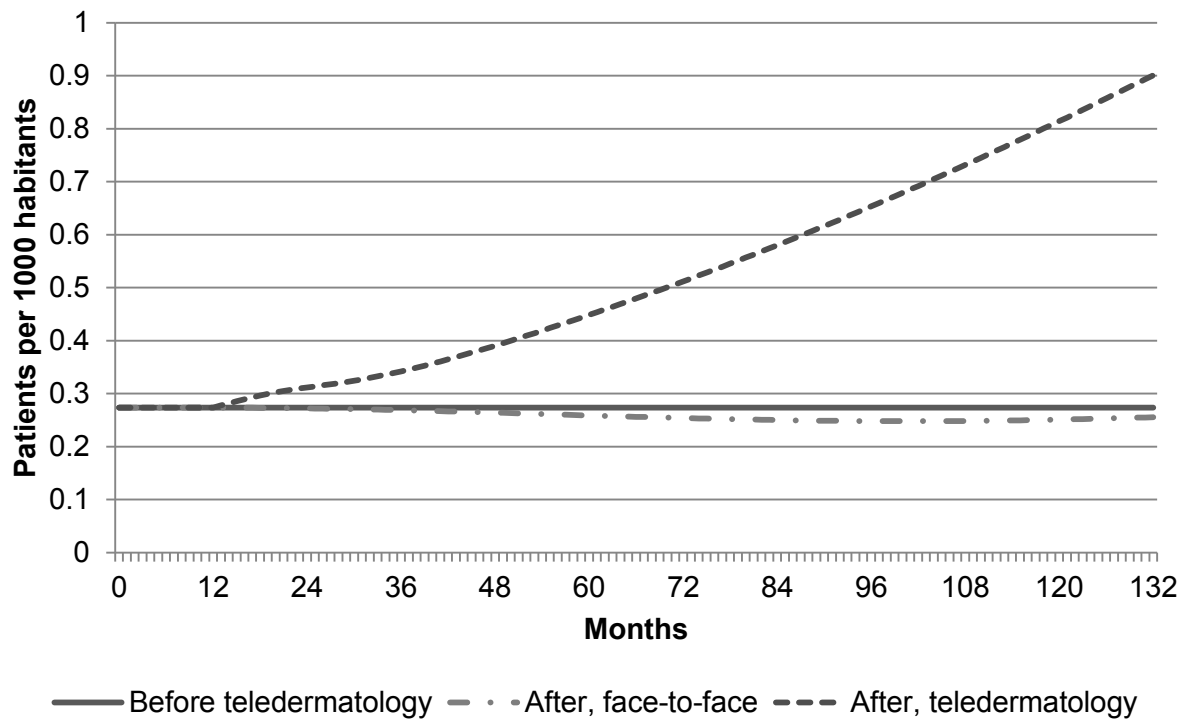


Figure 8-10 Referral rates per 1000 inhabitants

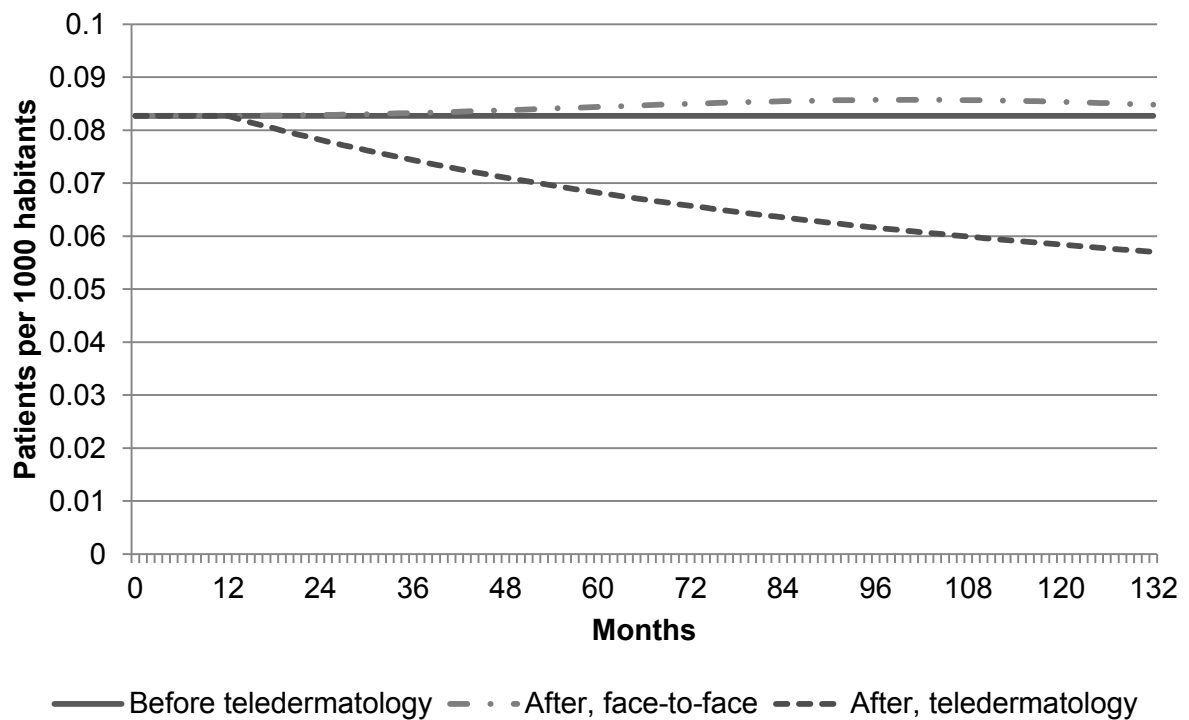


Figure 8-11 Non-completion rates per 1000 inhabitants

Figures 8-8 through 8-11 also show the impact of practice A adopting teledermatology on other practices (i.e., the practices that did not implement teledermatology). The effects

are negative: actual and perceived waiting times increase, referral rate decreases and non-completion rate increases. Perceived waiting time is slightly higher at the end of the simulation period, with actual waiting time reaching a maximum of 233 days (a 25 percent increase from the initial 187 days) around year 5. And, while Figure 8-8 shows that the true waiting time for face-to-face appointments is lower at the end of the simulation period, this is because fewer patients are being referred and non-completion is higher, as seen in Table 8-3. These effects are the consequence of two remaining connections between practice A and the face-to-face pathway: direct face-to-face referrals and subsequent face-to-face appointments. Both these connections add to the time dermatologists spend on patients from practice A.

This is an important result for various reasons. First, it shows in a very simple and clear way that the introduction of teledermatology creates enormous imbalances in the provision of dermatological outpatient care between practice A and other practices. Second, it is easy to see that these imbalances are due to the allocation of dermatologists' time to the new teledermatology pathway. Third, it illustrates the need to explore regional effects of wider use, as the allocation of 1 hour per month of teledermatology in only one practice (serving merely 2.5 percent of the region's population) is enough to increase waiting time for all other practices by 25 percent at a given point in time.

On the other hand, GP waiting time in practice A increases. Because GPs are now present in teleconsultations, they have less time to perform primary care appointments. However, the practical implications are limited for two reasons: 1) in the 60 minutes that GPs now spend doing teleconsultations, they can perform at most 3 GP appointments, since these are longer; and 2) those 3 appointments they stop performing represent little over 0.1 percent of all the primary care appointments they perform in a given month. While changes to dermatologists' time allocation can have important consequences, changes in GPs' time allocation seem to be considerably less relevant. There are, however, two caveats. One, a given practice can perform teleconsultations in more than one medical specialty (not only

teledermatology). As the number of teleconsultations increases, the impact on GP workload will become more important. Second, the impact depends naturally on the size of the practice. 60 minutes per month for the median sized practice may not be significant, yet for a smaller practice with only a couple of GPs, the impact is bigger.

Table 8-3 illustrates the astounding impact that 1 hour of dermatologists' time has on patients from practice A. Available dermatology capacity (i.e., appointments or patients per 1000 inhabitants), goes up by 391 percent. Still, this falls significantly short of meeting need/demand for dermatological outpatient appointments. The Central Administration of the Health System estimates that between 50 to 60 appointments per 1000 inhabitants are enough to cover the yearly need for outpatient dermatological care, 35 to 40 percent of which are first appointments (Administração Central do Sistema de Saúde I.P., 2008). I take the lower limit of 50 appointments per 1000 inhabitants and the estimated 40 percent of first appointments to calculate an underlying demand for dermatology in Alentejo. The numbers for unmet demand in Table 8-3 represent the cumulative difference between the underlying demand and referrals made to dermatology waiting lists. The results indicate a 21 percent drop in unmet demand, still failing to meet the needs of 146 patients per 1000 inhabitants at the end of the simulation period. These numbers are illustrative of how dire the supply of dermatologists in Alentejo is. In the future it would be interesting to investigate whether increased supply and access has led to demand inducement.

Table 8-3 Summary of practice-level results before/after teledermatology

Variable	Before	After	%Δ
Teledermatology waiting time (day)	187	15	-91.98%
Face-to-face waiting time (day)	187	148	-20.86%
GP waiting time, practice A (day)	8	10	25.00%
Referral rate, practice A (patients/1000 hab.)	0.274	0.901	228.83%
Referral rate, other practices (patients/1000 hab.)	0.274	0.256	-6.57%
Non-completion rate, practice A (patients/1000 hab.)	0.083	0.057	-31.33%
Non-completion rate, other practices (patients/1000 hab.)	0.083	0.085	2.41%
Capacity, practice A (patients/1000 hab.)	0.274	1.344	390.51%
Capacity, other practice (patients/1000 hab.)	0.274	0.262	-4.38%
Unmet demand, practice A (patients/1000 hab./year)	183.9	145.7	-20.77%
Unmet demand, other practices (patients/1000 hab./year)	183.9	188.0	2.23%
Avoided travelling, practice A (km)	-	21,498	-
Avoided waiting, practice A (patient*day)	-	72,153	-
Avoided referrals, practice A (patient)	-	49	-
Total costs to NHS (€)	617,635	647,003	4.75%
Total costs to NHS, per patient (€)	82.23	88.27	7.35%
Total costs to NHS, per km avoided (€)	-	1.37	-
Total costs to NHS, per patient day avoided (€)	-	0.41	-
Total costs to patients (€)	230,937	224,112	-2.96%

All values at end of simulation period (132 months)

Finally, it is possible to estimate the amounts of travelling and waiting that were avoided in practice A, following its implementation of teledermatology. In 10 years of teledermatology (between month 12 and 132 of the simulations), patients from practice A avoided travelling 21,498 km to the *Hospital do Espírito Santo de Évora* for face-to-face appointments, and saved 72,153 patient days of waiting. Furthermore, 49 patients avoided being referred to a teleconsultation altogether by having their condition treated by the GP, as a consequence of continued education and learning from participation in teleconsultations. The effect of continued education on the referral rate is visible in Figure 8-12.

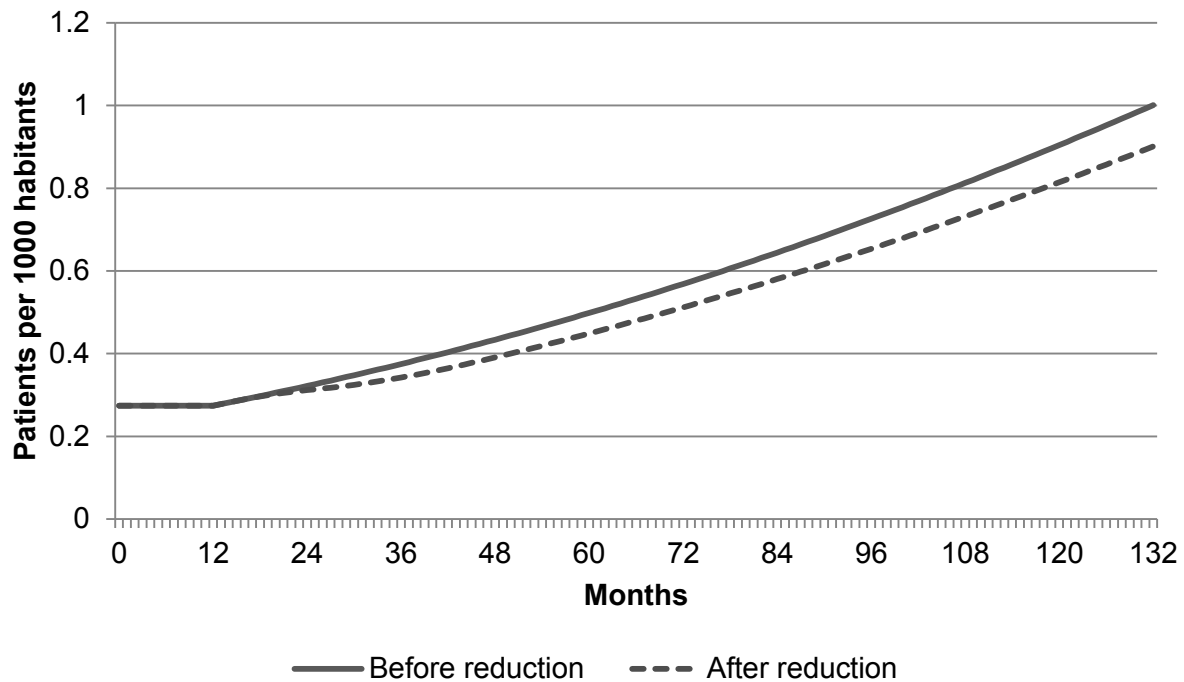


Figure 8-12 Referral rate to teledermatology before/after reduction from continued education

The simulation results at practice-level show that the allocation of a mere one hour per month of dermatologists' time is enough to cause huge imbalances in service quality, patient experience and potentially clinical outcomes. One practice's decision to implement teledermatology can have significant consequences for practices that remain in the face-to-face pathway, namely by increasing the waiting time for face-to-face appointments and reducing capacity and service utilisation. It is thus important to understand what the consequences at the regional level are, when several practices start using teledermatology.

8.5.2 Regional level simulations

As previously explained, the model has been extended to accommodate three practices that implement teledermatology at different times. I assume two more practices serving the same number of inhabitants (7,422) join practice A, at months 24 and 36, respectively one and two years after practice A implements teledermatology. All other assumptions remain the same, most importantly, dermatologists continue to allocate 1 hour every month to teleconsultations. The results are now clearly different (see Figures 8-13 through 8-16).

When teledermatology is implemented, capacity doubles and the rates of referral and non-completion react. The referral rate increases and the non-completion rate drops. The impact of these trends is that eventually, around month 84, the true waiting time starts to increase (see Figure 8-13). Because GPs and patients take 6 months to perceive the actual waiting time, referrals and non-completion do not react immediately and continue to respectively increase and decrease (see Figure 8-14). This delay means there will be an imbalance between supply and demand for teledermatology. When patients and GPs eventually perceive that waiting time is increasing, it will be too late: enough patients will have entered the list to cause the waiting time to increase, over the following years.

With time, perceived waiting time will peak (beyond the simulation period) and when this happens, referral rate will be at its lowest and non-completion at its highest. Waiting time will then drop and the process happens all over again, as when teledermatology was introduced. The fact the system oscillates merely reflects the *adaptive expectations* formulation, which accounts for delays in perceptions of true waiting times. This oscillatory behaviour goes on *ad aeternum*, driven by the impact of delayed perceptions of waiting time on referrals and non-completion. The choice of *time to perceive waiting time* affects the frequency and amplitude of the oscillation. Lower values for the *Time to perceive waiting time* (i.e., the lower the value the quicker patients and GPs perceive the true waiting time) lead to more frequent oscillation and of bigger amplitude, and vice-versa.

The actual face-to-face waiting time exhibits a trajectory similar to that in Figure 8-8, with the difference that waiting time drops more quickly. It initially increases as a consequence of shifting one hour of dermatologists' time to teledermatology. This causes the referral rate to drop. While this is happening, three practices leave the face-to-face pathway. They still refer a number of patients to face-to-face appointments and some teleconsultations do lead to subsequent face-to-face appointments, yet these represent only a fraction of the reduction in referrals from their exit. The combined effect is a drop in referrals and an ensuing fall in the waiting time. Because the *Time to perceive waiting time* in

the face-to-face pathway is one year, these changes in the true waiting time are, to a great extent, not reflected in the perceived waiting time (as seen in Figure 8-14).

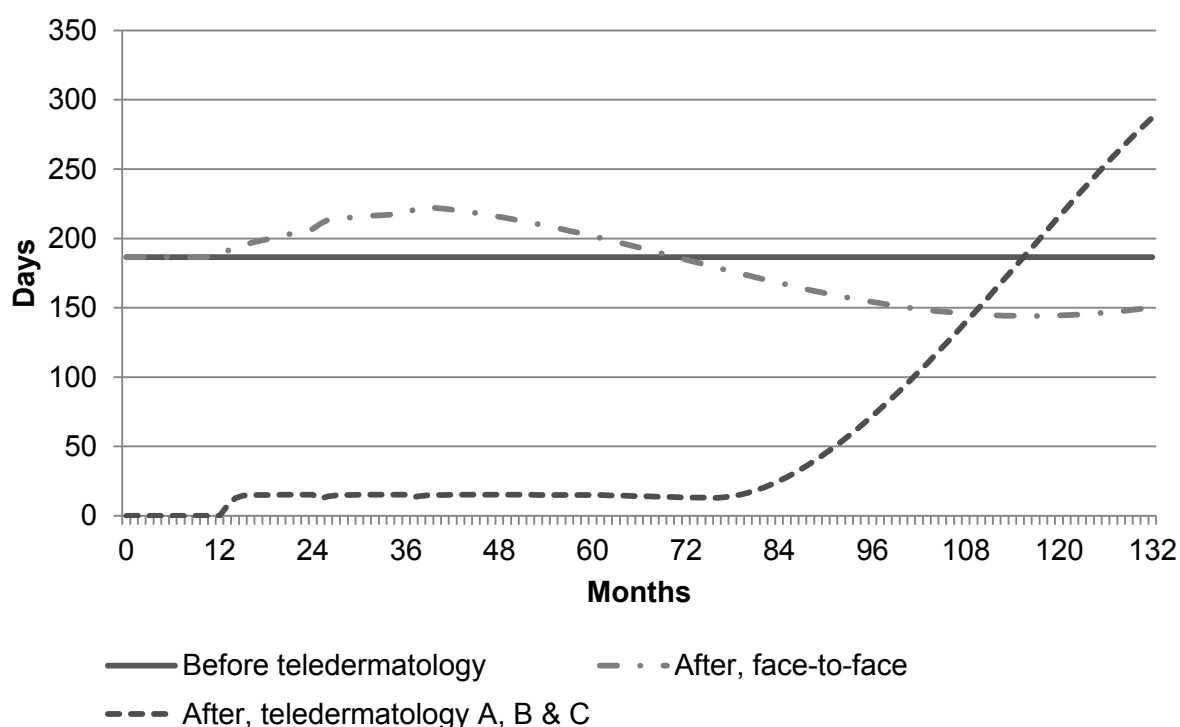


Figure 8-13 Actual waiting times for dermatology appointments

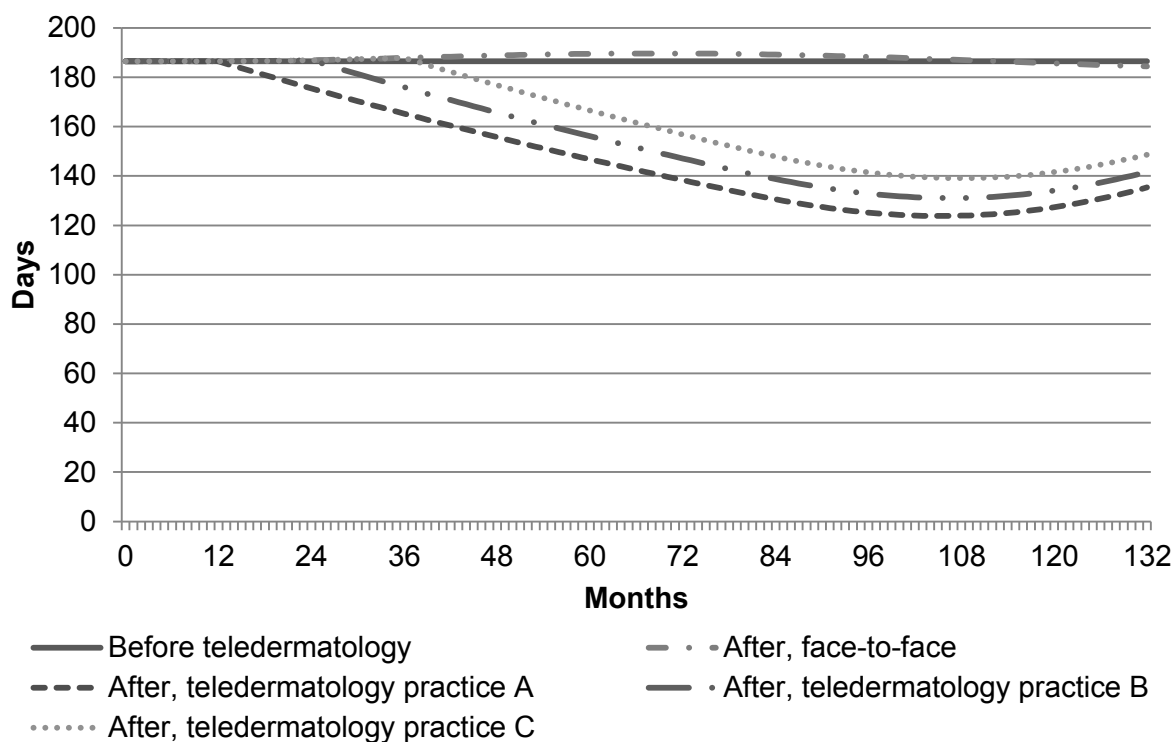


Figure 8-14 Perceived waiting times for dermatology appointments

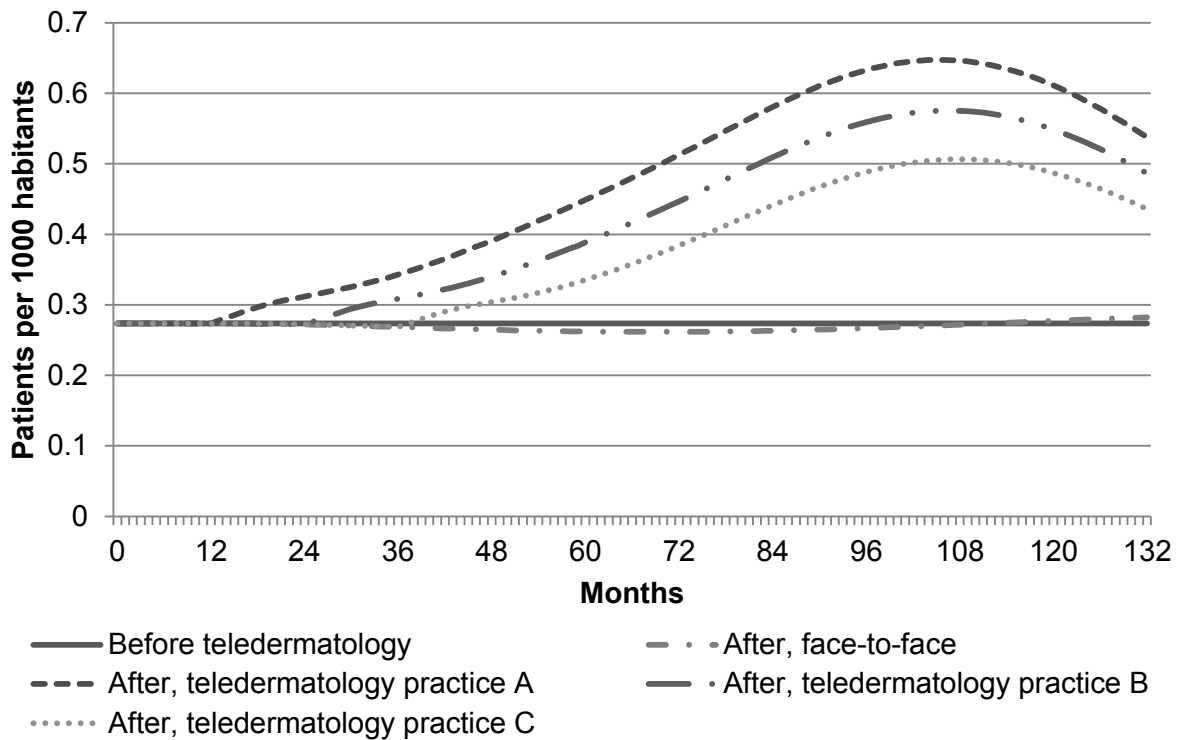


Figure 8-15 Referral rates per 1000 inhabitants

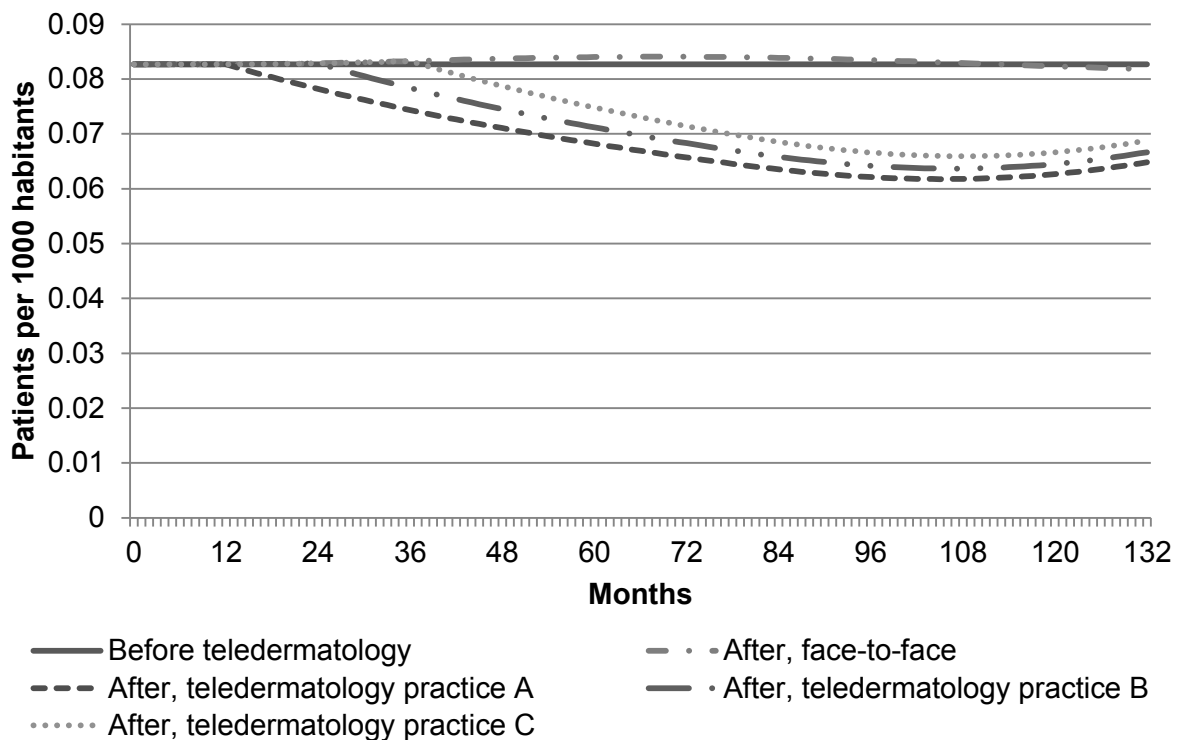


Figure 8-16 Non-completion rates per 1000 inhabitants

In Figures 8-14 through 8-16, the different implementation timings for the three teledermatology practices are clearly visible. It is also interesting to note that first movers

enjoy greater benefits than late comers. When practice C implements teledermatology, in month 36, practice A's referral rate has already increased by 25 percent. Eventually this first mover advantage disappears, especially for lower values of *Time to perceive waiting time*.

Even though the actual waiting time is higher at the end of the simulation period, the benefits for practices implementing teledermatology are still significant (see Table 8-4).

Table 8-4 Summary of regional level results before/after teledermatology

Variable	Before	After	%Δ
Teledermatology waiting time (day)	187	287	53.48%
Face-to-face waiting time (day)	187	150	-19.79%
Referral rate, practices ABC (patients/1000 hab.)*	0.274	0.487	77.74%
Referral rate, other practices (patients/1000 hab.)	0.274	0.282	2.92%
Non-completion rate, practice A (patients/1000 hab.)*	0.083	0.067	-19.28%
Non-completion rate, other practices (patients/1000 hab.)	0.083	0.082	-1.20%
Capacity, practices ABC (patients/1000 hab.)*	0.274	0.448	63.50%
Capacity, other practice (patients/1000 hab.)	0.274	0.275	0.36%
Unmet demand, practices ABC (patients/1000 hab./year)*	183.9	159.5	-13.27%
Unmet demand, other practices (patients/1000 hab./year)	183.9	187.7	2.07%
Avoided travelling, practices ABC (km)	-	42,500	-
Avoided waiting, practices ABC (patient*day)	-	47,952	-
Avoided referrals, practices ABC (patient)	-	109	-
Total costs to NHS (€)	617,635	736,423	19.23%
Total costs to NHS, per patient (€)	82.23	93.09	13.21%
Total costs to NHS, per km avoided (€)	-	2.80	-
Total costs to NHS, per patient day avoided (€)	-	2.48	-
Total costs to patients (€)	230,937	240,386	4.09%

All values at end of simulation period (132 months); * averaged across practices A, B and C.

Referral non-completion rates show improvement in teledermatology practices and unmet demand is lower. As before, unmet demand in face-to-face practices is higher. Even so, the fact that two more practices are now *sharing* the same one hour per month

teledermatology capacity is a slight improvement in terms of equity, as illustrated in the small 0.36 percent increase in face-to-face capacity. Table 8-5 illustrates how teledermatology benefits depend on the scale of implementation.

Table 8-5 Impact of joining teledermatology practices on regional-level outcomes

	Variable*	F2f	%Δ[‡]	Tele	%Δ[‡]
1 teledermatology practice	Waiting time	148	10.70%	15	-91.98%
	Referral rate ^θ	0.256	-6.57%	0.901	228.83%
	Non-completion rate ^θ	0.085	2.41%	0.057	-31.33%
	Dermatology capacity per 1000	0.262	-4.38%	1.344	390.51%
	Unmet demand per 1000 ^θ	188.0	2.23%	145.7	-20.77%
	Avoided travelling	-	-	21,498	-
	Avoided waiting	-	-	72,153	-
	Avoided referrals	-	-	49	-
2 teledermatology practices	Waiting time	158	-15.51%	75	-59.89%
	Referral rate ^θ	0.268	-2.19%	0.636	132.12%
	Non-completion rate ^θ	0.083	0.00%	0.067	-19.28%
	Dermatology capacity per 1000	0.267	-2.55%	0.672	145.26%
	Unmet demand per 1000 ^θ	187.5	1.96%	161.8	-12.02%
	Avoided travelling	-	-	37,797	-
	Avoided waiting	-	-	112,997	-
	Avoided referrals	-	-	89	-
3 teledermatology practices	Waiting time	150	-19.79%	287	53.48%
	Referral rate ^θ	0.282	2.92%	0.487	77.74%
	Non-completion rate ^θ	0.082	-1.20%	0.067	-19.28%
	Dermatology capacity per 1000	0.275	0.36%	0.448	63.50%
	Unmet demand per 1000 ^θ	187.7	2.07%	159.5	-13.27%
	Avoided travelling	-	-	42,500	-
	Avoided waiting	-	-	47,952	-
	Avoided referrals	-	-	109	-

All values at end of simulation period (132 months); F2f – face-to-face practices; Tele – teledermatology practices; * units as in Table 8-4; ^θ averaged across practices; [‡] percentage change relative to before teledermatology, see Table 8-4.

When only practice A is using teledermatology, the benefits are astounding (as discussed in the previous section). When practice B joins, the teledermatology waiting time is higher but is still quite attractive (75 days for a teleconsultation compared to 158 days for a face-to-face appointment). Despite the fact that benefits are now less clear as measured in terms of teledermatology referral and non-completion rates, unmet demand, and capacity per 1000 inhabitants, conditions are still significantly better than in the face-to-face pathway. In fact, 112,997 patient days of waiting are avoided. When practice C joins, 2 years after practice A began using teledermatology, the one hour capacity becomes clearly insufficient. In other words, there are diminishing returns from wider use of teledermatology under a capacity constraint. This is clearly visible in the rates of growth for avoided travelling, waiting and referrals. For example, there is a 76 percent increase in avoided travelling when practice B joins, but only a 12 percent increase when practice C joins.

Also of interest are the outcomes for practices that remain in the face-to-face pathway. The results indicate that waiting times drop as more practices join teledermatology. However, in all three scenarios, unmet demand in face-to-face practices is higher. Even so, the more practices join teledermatology the less worse off face-to-face care becomes. Naturally, since teledermatology capacity is constrained to one hour per month, if enough practices implemented teledermatology, face-to-face practices would eventually be better off.

8.5.3 Policy scenarios

As mentioned in Chapter 5, some practices in Alentejo appoint a coordinator who is present in every teleconsultation for a specific specialty, while in others the patient's own GP is present. This can be regarded as a policy lever. There are certain advantages to having a coordinator. For one, the coordinator sees more patients, thus acquiring experience more rapidly. Second, it can be organisationally challenging to have all GPs present for one hour every month to present one or two patients each. Having one GP perform every

teleconsultation is considerably easier. On the other hand, since it is assumed the effect on referral rates is not transferable between GPs, having a coordinator will limit the number of patients who benefit from continued GP education to those seen by the coordinator.

I tested the impact of having a coordinator at the practice-level. The difference in the referral rate is illustrated in Figure 8-17. When every GP participated in teleconsultations, 78 referrals were avoided (see Table 8-3). If the practice appoints a coordinator, that number goes down to 13.

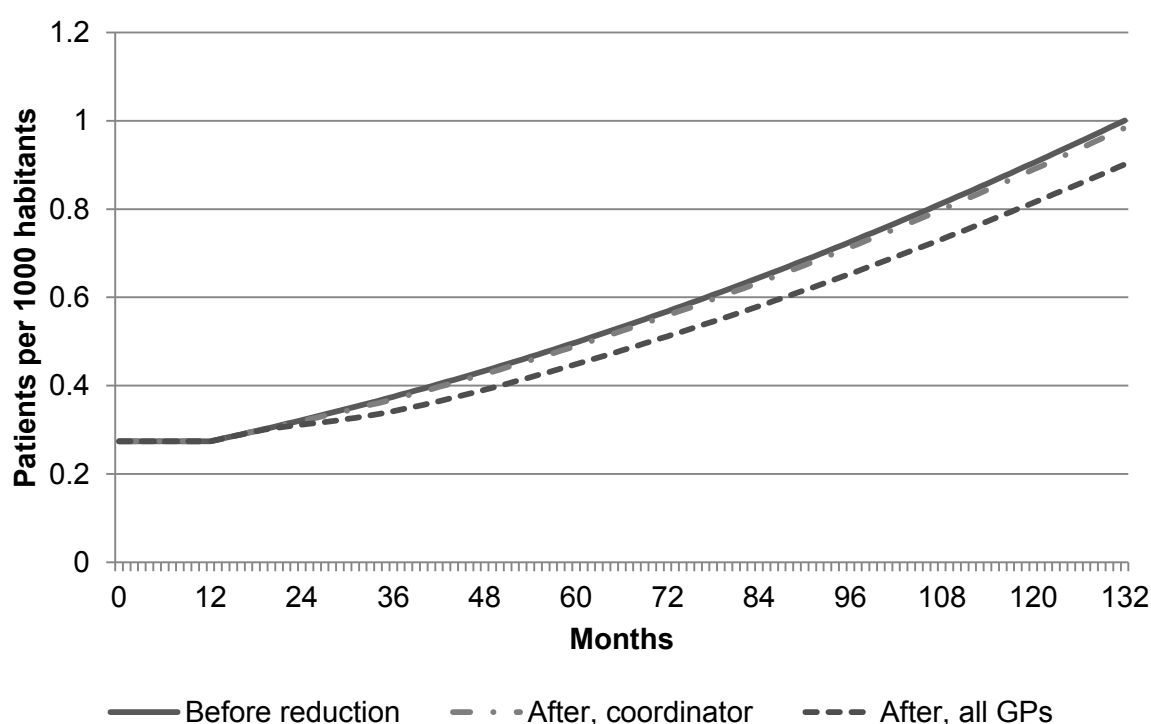


Figure 8-17 Referral rate to teledermatology before/after reduction from continued education, with and without a coordinator who performs all teleconsultations

A second policy decision is the reimbursement of teleconsultations by the NHS. Currently, both the hospital and the primary care practice get reimbursed at €83.79 for each teleconsultation. This is effectively a financial incentive for practices to implement teledermatology, a strategy to promote wider use. It is however, costing the NHS a great deal, as seen in Table 8-3 and Table 8-4. It is important to note that physician total work time (dermatologists' and GPs') does not actually increase, it merely shifts from face-to-face and

GP appointments to teleconsultations. With regard to secondary care, the number of appointments performed in the same amount of dermatologist time doubles. Accordingly, hospital income doubles as well. Concerning primary care, the impact is less clear, as practices are funded through contracts. The per capita funding for the *Unidade Local de Saúde do Baixo Alentejo* (an *Unidade Local de Saúde* is a public entity which integrates secondary and primary care units), for example, was €601 for 2013 (Administração Central do Sistema de Saúde I.P., 2012). Every teleconsultation represents approximately 14 percent of this per capita.

Given the clear benefits of teledermatology for practices using the service, and the obvious disadvantages to those practices still using face-to-face care, it would seem unnecessary to further incentivise practices financially. While at the primary care level, it is difficult to understand the financial impact of substituting GP appointments for teleconsultations, at the secondary care level, teleconsultations represent a clear burden to NHS budgets. I tested the impact of halving the reimbursement fee, so that both practice and hospital receive €41.90 (in other words, the reimbursement for each patient seen remains constant after the introduction of teledermatology). Furthermore, given the significant benefits to patients from avoided travelling and waiting, I also changed the co-payment for teleconsultations from its original €3 to that of a face-to-face appointment: €7.50. The results of this policy are illustrated at the regional level (3 practices using teledermatology) in Table 8-6. The results show that teledermatology still increases expenditure, yet by significantly less. Furthermore, even with the increase in co-payment, costs to patients are still lower.

Table 8-6 Impact of changes to reimbursement and co-payment of teledermatology

Variable	F2f	Tele	%Δ*	Policy	%Δ*
Total costs to NHS, per patient (€)	82.23	183.7	118.13%	103.4	17.03%
Total costs to NHS, per km (€)	-	2.80	-	1.15	-
Total costs to NHS, per patient day (€)	-	2.48	-	1.02	-
Costs to patients, per appointment (€)	30.75	6.53	-78.76%	11.03	-64.13%

All values at end of simulation period (132 months); F2f – face-to-face practices; Tele – teledermatology practices; * percentage change relative to face-to-face.

The final policy lever is the allocation of dermatologists' time to teledermatology. Curiously, this has not been mentioned in the interviews or in the literature. Yet, the simulations show quite clearly this is an important issue. Setting an arbitrarily high value for the time spent by dermatologists on teledermatology can lead to significant imbalances across the region, making those living in teledermatology practices better off and those living in other practices even worse off. This is an important insight. Teleconsultations are often hailed as a means to increase equality through increased access for those living in remote locations. The simulation results suggest that teleconsultations may actually aggravate the situation. With the teledermatology referral rate increasing by 229 percent and the referral rate of face-to-face practices dropping by 7 percent (see Table 8-3), inequity is actually much higher.

Setting an adequate time for teledermatology is not trivial. A good starting point is to estimate some measure of required capacity per 1000 inhabitants. Using the monthly dermatological capacity and the average population in the region in Table 8-2, the estimated average monthly capacity per 1000 inhabitants is 0.275 patients. For a median-sized practice of 7,442 inhabitants, this results in roughly 2 appointments per month, or 12 minutes of dermatologist time. Setting up two video-conferencing platforms (one for the hospital and one for the practice), plus making all the organisational arrangements, to perform two teleconsultations is nonsensical. One solution would be to implement the technology in multiple median-sized practices, or in fewer but bigger practices. To reach 1 hour of

teleconsultations per month, a population of 36,364 is needed. Assuming that three equally-sized practices, with 12,121 inhabitants each, implement teledermatology in month 12, and that 1 hour of dermatologists' time is made available to them, the result is oscillation (see Figure 8-18).

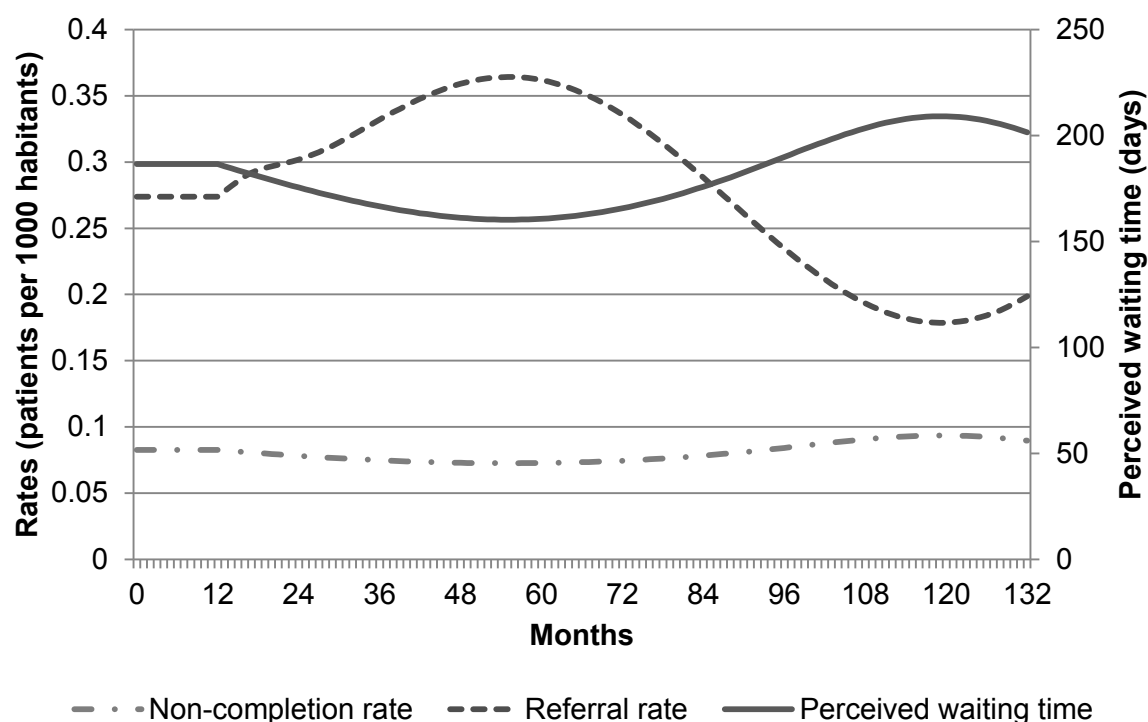


Figure 8-18 Impact of equitable allocation policy on teledermatology waiting time and rates

The behaviour in Figure 8-18 may not be realistic. While a direct consequence of the assumptions contained in the model (the same which produced the results in Tables 8-3 through 8-5), the behaviour would probably not occur in practice. There are two possible explanations for this. The first, and simpler, one is that understanding of behaviour in lower waiting times in this thesis is poor. Anecdotal evidence suggests waiting times for teledermatology are between one week and two months, but in my sample (in Chapter 6) I only observe waiting times longer than 100 days. While the elasticities at lower waiting times estimated in Chapter 6 are statistically significant, they are not based on actual observed behaviour, but rather inferred from a sample which does not include those values. The oscillatory behaviour in the regional-level simulations may simply reflect poor

representations of the elasticities of referrals and non-completion with respect to waiting time.

A second explanation is that supply is not constant in real life. It is possible that sometime after the introduction of teledermatology, policy makers and physicians would feel motivated by the positive results (e.g., reduced waiting time and non-completion, and increased referrals) and would increase teledermatology capacity to 2 hours per month. Assuming this would happen one year after the introduction of teledermatology, the oscillations would almost disappear (see Figure 8-19). Again, encouraged by the positive results of increasing teledermatology capacity, policy makers and physicians might decide to extend teledermatology capacity once again one year later to 3 hours per month. The oscillatory behaviour now seems to have disappeared. Assuming a further increase a year later to 4 hours per month (on average, 1 hour per week), the improvement in perceived waiting time would be negligible.

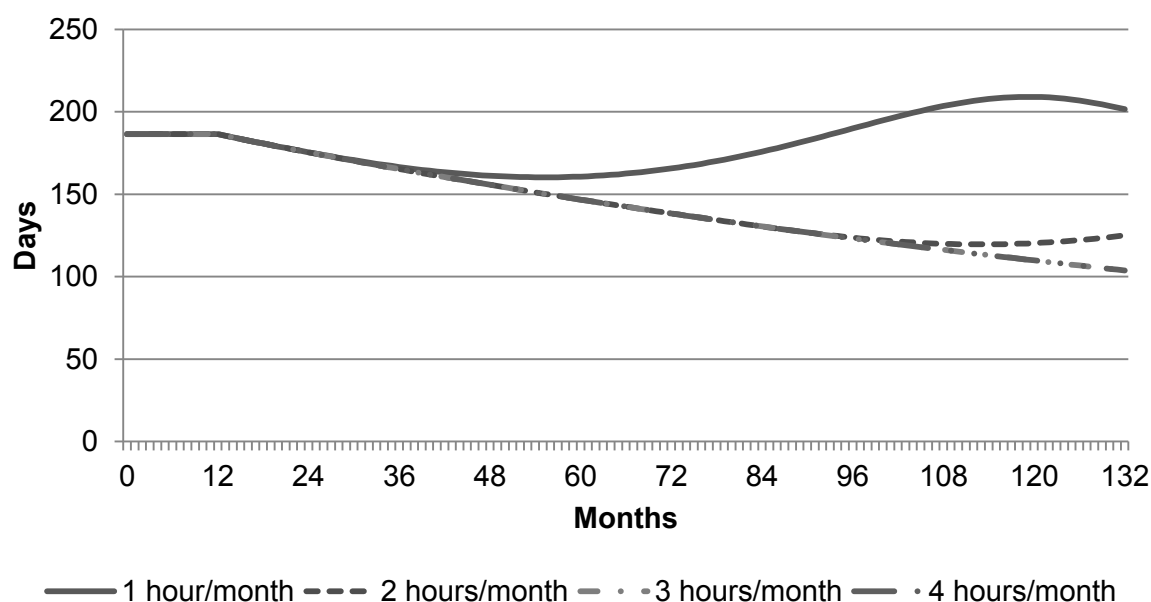


Figure 8-19 Perceived teledermatology waiting time for expanding allocation policy

The expanding allocation policy is interesting because it is realistic. In Alentejo, the allocation of dermatologists' time to teledermatology has been increasing since the initial pilot in 1998. For the majority of the period under analysis, there was approximately 1 hour

per week available to teledermatology practices. Recently, there have been calls to further increase teledermatology capacity, in response to practices joining the teledermatology pathway. It is also appropriate to consider the effects of these allocation policies on face-to-face care. Table 8-7 shows the impacts of both the equitable and expanding allocation policies on both face-to-face and teledermatology outcomes (values are for the end of the simulation period).

Table 8-7 Impact of two allocation policies: equitable and expanding

	Variable*	F2f	%Δ[‡]	Tele	%Δ[‡]
Equitable allocation policy	Perceived waiting time	178	-4.81%	202	8.02%
	Referral rate ^θ	0.308	12.41%	0.199	-27.37%
	Non-completion rate ^θ	0.079	-4.82%	0.090	8.43%
	Dermatology capacity per 1000	0.291	6.20%	0.275	0.36%
	Unmet demand per 1000 ^θ	184.7	0.44%	179.5	-2.39%
	Avoided travelling	-	-	44,728	-
	Avoided waiting	-	-	-99,176	-
	Avoided referrals	-	-	123	-
Expanding allocation policy	Perceived waiting time	219	17.11%	104	-44.39%
	Referral rate ^θ	0.173	-36.86%	0.901	228.83%
	Non-completion rate ^θ	0.099	19.28%	0.057	-31.33%
	Dermatology capacity per 1000	0.224	-18.25%	1.100	301.46%
	Unmet demand per 1000 ^θ	191.4	4.08%	141.7	-22.95%
	Avoided travelling	-	-	105,641	-
	Avoided waiting	-	-	594,325	-
	Avoided referrals	-	-	228	-

All values at end of simulation period (132 months); F2f – face-to-face practices; Tele – teledermatology practices; * units as in Table 8-4; ^θ averaged across practices; [‡] percentage change relative to before teledermatology, see Table 8-4.

Because there is oscillation in the equitable allocation scenario, the numbers in Table 8-7 should be interpreted with caution. They show higher waiting times and non-completion, lower referrals and even a negative value for avoided waiting time (i.e., teledermatology did not overall lead to a reduction in waiting times). Yet, looking at Figure 8-18, it is clear that the

end of the simulation coincides with the negative part of the oscillation. Once again, unmet demand shows improvement.

More interesting are the results for face-to-face practices when the expanding allocation policy is adopted. The situation gets worse on all accounts: waiting time, non-completion and unmet demand increase, while capacity and referral rate drop. As with the oscillatory behaviour, it is not realistic to assume that these changes would be allowed to persist. In practice, face-to-face capacity would be changed. In the past, while the number of full-time dermatologists has remained constant, a number of interns have come and gone, as part of their medical training, temporarily increasing capacity and providing some relief. Furthermore, in 2011 and 2012, the hospital hired a consultant on a flexible basis. This consultant left in 2013. These changes to supply would explain how the region has managed to keep the waiting time for face-to-face appointments under relative control.

8.6 Sensitivity analyses

It is important to test the robustness of the simulation results with respect to uncertainty in the parameters and initial conditions. There are many potential tests, yet only a number of parameters will actually have an important impact on findings and thus merit a sensitivity test. For example, the teleconsultation literature frequently tests the sensitivity of total costs to changes in distance (Wootton et al., 2000) and cost of video-conferencing equipment (Lamminen et al., 2011).

8.6.1 Sensitivity to distance and equipment costs

In Alentejo, distance from practices to the hospital varies between 13.6 and 152.8 kilometres. Those values do not affect the unit costs of teleconsultations to patients, but they do affect the unit costs of face-to-face care, making teledermatology more attractive the farther away from the hospital the practice is. For the shortest distance the difference in unit costs is €10.37 (in favour of teleconsultations), while for the longest distance it is €52.32. Put

simply, a 157 percent increase in distance (from the median 59.5 to 152.8 kilometres) corresponds to an 89 percent drop in the difference between the unit costs of teleconsultation and those of face-to-face.

As for the video-conferencing equipment, there is no sample from which to choose a value. Lamminen et al. (2011) choose a somewhat arbitrary value of €2,000 for the total costs of equipment in 2009, a reduction of 94 percent from the investment in 1997. Applying a 95 percent reduction in the costs of equipment in Alentejo, the total costs per teledermatology patient for the NHS drop from €172.84 to €157.85, compared to €82.23 for a face-to-face patient. Since both practices and hospitals get reimbursed for teleconsultations, this is not enough to make teledermatology cheaper for the NHS. However, if the hospitals and practices got reimbursed at half that value (see previous section), patients paid the usual co-payment fee of €7.50, and assuming a 95 percent reduction in equipment costs, the NHS costs per teledermatology patient would be €78.14, now cheaper than those of a face-to-face patient.

The total costs for the NHS following the introduction of teledermatology, at the practice-level (one practice serving 7,442 inhabitants, same as in section 8.5.1), would be €600,185, compared to €617,635 before the introduction of teleconsultation. All other results would hold, as in Table 8-3. In such a scenario, the use of teleconsultation would result in fewer costs, increased utilisation and better patient experience. It should be noted that a 95 percent reduction in costs is not unrealistic as there are currently open source solutions at no cost (such as Skype™) with very good quality.

Both distance and equipment costs are static variables in the model. In other words, behaviour does not change as a consequence of distance or equipment costs. It is changes in perceptions of waiting time and consequently changes to referral and non-completion rates that actually drive outcomes. As such, it is of the essence that assumptions concerning behaviour are tested. This means exploring the three feedback loops in the model: the

elasticities of referrals and non-completion with respect to waiting times and the reduction in referrals from continued GP education and learning.

8.6.2 Sensitivity to effect of continued GP education on referrals

The 95 percent confidence intervals for the expected probability reductions estimated in Chapter 7, are used to test the impact on the referral rate of GPs participating in teledermatology, at practice-level. The reductions in the referral rates are shown in Figure 8-20. At the lowest reduction level, only 18 referrals are avoided, compared to 80 referrals at the highest reduction.

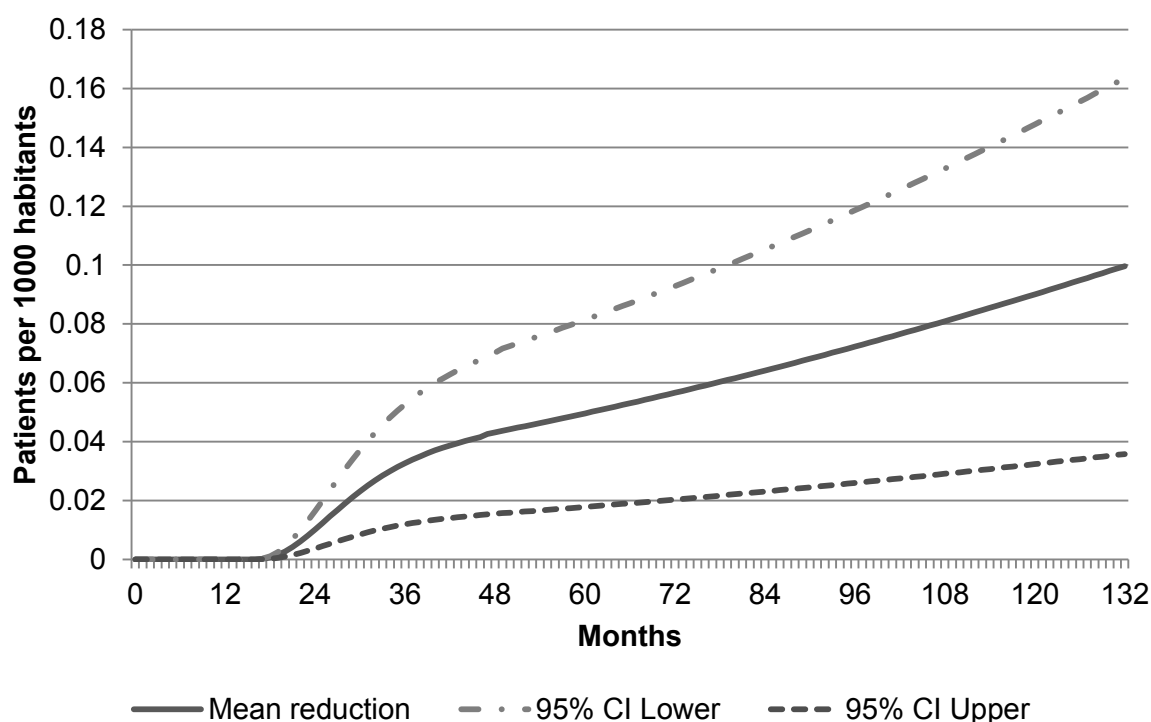


Figure 8-20 Sensitivity to different assumptions concerning the learning effect

8.6.3 Sensitivity to elasticities of demand and non-completion

Regarding the referral and non-completion loops, the key assumptions relate to the shapes of the elasticities with respect to waiting time. The elasticities of referrals and non-completion were estimated as functions of waiting time in Chapter 6 with 95 percent confidence

intervals. The lower and upper bound curves were used to test the sensitivity of the model to different elasticities, at practice-level (see Table 8-8).

Table 8-8 Impact of different elasticities on perceived waiting times and unmet demand

Perceived waiting time, F2f		Elasticity of non-completion <i>wrt</i> waiting time		
		Less elastic*	Mean	More elastic [¥]
Elasticity of referrals <i>wrt</i> waiting time	Less elastic [¥]	202	202	202
	Mean	191	191	191
	More elastic*	187	187	187
Perceived waiting time, Tele		Elasticity of non-completion <i>wrt</i> waiting time		
		Less elastic*	Mean	More elastic [¥]
Elasticity of referrals <i>wrt</i> waiting time	Less elastic [¥]	104	104	104
	Mean	104	104	104
	More elastic*	114	114	114
Unmet demand, F2f		Elasticity of non-completion <i>wrt</i> waiting time		
		Less elastic*	Mean	More elastic [¥]
Elasticity of referrals <i>wrt</i> waiting time	Less elastic [¥]	187.2	187.2	187.2
	Mean	188.0	188.0	188.0
	More elastic*	188.0	188.0	188.0
Unmet demand, Tele		Elasticity of non-completion <i>wrt</i> waiting time		
		Less elastic*	Mean	More elastic [¥]
Elasticity of referrals <i>wrt</i> waiting time	Less elastic [¥]	176.4	176.4	176.4
	Mean	145.7	145.7	145.7
	More elastic*	88.9	88.9	88.9

All values at end of simulation period (132 months); F2f – face-to-face practices; Tele – teledermatology practices; *wrt* – “with respect to”; waiting times in days, unmet demand in patients per 1000 inhabitants; * 95% confidence interval lower bound; [¥] 95% confidence interval upper bound.

It is important to note that the impact is now quite complex. While with continued education only two parameters were changing, here a whole curve is changing. Also, there are many possible combinations for the true curve, somewhere between the lower and upper bound curves. And finally, the impact depends on initial conditions, since elasticities are a function of waiting time. Having said that, the values shown in Table 8-8 are merely

indicative; they are in no way exhaustive. Yet, they clearly indicate that the elasticity of referrals is more important than the elasticity of non-completion. They also suggest that the more elastic the referral rate, the less unmet demand for tele dermatology there is.

8.7 Discussion

In this chapter, I have sought to build on previous evaluations of teleconsultation by explicitly including a number of causal mechanisms and contextual influences in a simulation model of teleconsultation in Alentejo. Using a model which can only be considered a simple representation of a much more complex reality, I have estimated multiple effects, from different perspectives, at different scales of use, and under different assumptions. Furthermore, I have identified key policy levers and their likely effects, and signalled areas which justify further research and data collection given their impact on outcomes. All this now makes it possible to discuss generalizability of findings.

The simulation model includes only three feedback loops and focuses only on a handful of key outcomes. And yet the number of simulations, scenarios and tests that can be performed far exceeds those that can be adequately detailed and discussed here. This is illustrative of how difficult it is to understand complex dynamic systems in all their richness. The model makes the consequences of researchers and practitioners' assumptions and worldviews transparent. It is worth emphasising that all of the model's components are informed, at some level, by pre-existing knowledge and data. The resulting behaviours emerge not from some obscure set of formulations, but from established researchers and practitioners' beliefs and hypotheses about how patients and physicians make decisions in outpatient care pathways. The simulation model allows these hypotheses to be tested, producing insights into existing knowledge.

An important finding is that teleconsultation – frequently hailed as a means to reduce imbalances – may actually exacerbate them. The many benefits reported in the literature are

enjoyed by only a few, and, to a degree, at the expense of those still using face-to-face appointments. Even in an equitable allocation scenario, face-to-face capacity is diminished as a consequence of both face-to-face referrals from teledermatology practices, and subsequent face-to-face appointments following teleconsultations. While supply was clearly insufficient before the introduction of teledermatology, it was allocated across practices on the basis of clinical need. The face-to-face waiting list gave priority to those individuals most in need, regardless of which practice they came from. With the introduction of teledermatology, a new pathway and waiting list is created, one which necessarily competes for dermatologists' time. This new waiting list is initially empty and exit is quicker as capacity is doubled. On top of enjoying these benefits (plus all the benefits associated with reduced travelling), teledermatology practices continue to refer urgent and severe patients to the face-to-face waiting list. There are now two concurrent mechanisms for gaining access to outpatient dermatological care, competing for limited consultant time.

While it is hardly surprising that ring-fencing a disproportionate amount of consultant time for a relatively small number of patients will have negative consequences for other practices, it is remarkable this has not been discussed previously in the literature. Moreover, none of the physicians and managers interviewed mentioned this issue. To my knowledge, not one study or interviewee considered the potential imbalances resulting from the creation of a parallel pathway. Indeed, assertions that implementation of teledermatology at a wider level would naturally lead to substantial benefits may not hold. Whether wider use effectively results in broader benefits will depend on how well capacity meets underlying demand to begin with. A doubling of capacity will not be sufficient if it still fails to meet demand. As the practice-level simulations illustrate, under a disproportionate allocation of dermatologists' time to only 2.5 percent of the population, the resulting 229 percent increase in the referral rate would not be enough to eliminate unmet demand in that population. It follows that even if all practices adopted teledermatology, waiting lists and unmet demand would persist, albeit possibly to a lesser extent.

This is confirmed in the regional level simulations. There are diminishing returns to the wider adoption of teledermatology. The only explanation for the astounding advantages enjoyed by teledermatology practices is that they are benefiting from disproportionate use of resources, an unsustainable situation that must eventually be corrected. The simulations indicate that a more comprehensive use of teledermatology would indeed result in greater service capacity, productivity and utilisation, but waiting times would not disappear. Broader implementation would limit the inequity introduced by teledermatology, since all practices would enjoy the same level of care. As it stands, that is not the case.

The simulations indicate that enclosing all these dynamic effects into black boxes is unrealistic and can lead to gross misinterpretations of impact and cost-effectiveness. Looking only at final measures of utilisation would ignore the number of patients who did not get referred as a consequence of continued GP education and learning. These patients did not have to come back for a teleconsultation saving money and time. Furthermore, they saved precious dermatologist time allowing other patients to be seen in their place. Effectively, the system is now treating more patients with the same level of resource utilisation (i.e., number of appointments). These changes would go unnoticed in standard economic evaluations as the focus would be on final outcomes, not intermediate effects or changes. The simulation results suggest the reduction in referral rates from continued learning is a key variable of interest in the study of teleconsultation. Again, it has been largely overlooked by previous research.

Finally, the model confirms the importance of baseline conditions and context when evaluating complex interventions. The patterns of behaviour described here are true for Alentejo, where there is substantial undersupply of dermatologists. Even a 229 percent increase in referrals falls short of meeting the level of demand that is met in other regions of the country, such as Lisbon (Administração Central do Sistema de Saúde I.P., 2008). While some demand inducement is a possibility, especially whilst a few practices continue to enjoy a disproportionate amount of resources, it is difficult to make the point that teledermatology

is inducing utilisation in Alentejo. It is more likely responding to unmet demand. This is an important point. The question in Alentejo is not whether teledermatology increases costs. It is whether there is any other intervention that would increase referrals by 229 percent at lower cost. There should be no doubt that costs must increase if the Portuguese NHS is to meet the principles of universality and equitable access.

Hiring more dermatologists would be an obvious solution, yet physicians have repeatedly declined positions in the region in the past. Thus, comparing teledermatology to current costs of face-to-face care is unrealistic. Such a comparison would presume that recruiting dermatologists at current prices is an option, when, in reality, in the past 10 years this has not happened. Teledermatology can undoubtedly increase existing dermatologists' output, and under certain conditions (e.g., same reimbursement as face-to-face, higher co-payment, lower equipment costs, etc.) it may be cheaper than recruiting even at current prices, but it will not solve the problem. Eventually, capacity will be insufficient. A second solution would be to perform teleconsultations with physicians in other parts of the country, such as Lisbon or Oporto, where there is abundant supply of dermatologists.

8.7.1 Limitations and future work

There are a number of limitations which must be considered. First, some of the parameters are based on expert opinion (e.g., the variable *Time to achieve maximum reduction*) and others on previous research not necessarily related to teleconsultations (e.g., the specification of the *Effect of continued education on referral rate* as an s-shaped curve). Ideally, one would collect experimental data on these variables but this was beyond the resources available for this thesis. A second potential limitation is the choice of model boundaries. It was decided to not include implementation as this would significantly change the scope of the model and the questions being asked. Third, for reasons discussed previously, supply was not expressed as a function of waiting time, as in van Ackere and Smith (1999). Indeed, it was not a function of any other variable in the model, except

simulated time. It would be important to understand how and why teleconsultation affects the supply of outpatient appointments. Finally, the impact on the use of private care, self-care and A&E department was not explicitly modelled.

In the future, these limitations can be addressed by expanding the model to include more behaviours and less restrictive assumptions. Furthermore, it would be interesting to explore the use of other simulation modelling techniques to answer more focused research questions, and combine the insights from those models with the system dynamics model presented here.

8.8 Conclusion

While there is no silver bullet method for the evaluation of complex health interventions, it should now be clear that assessing the impact of teledermatology in one practice in Alentejo, using conventional economic evaluation methods, focusing on final clinical outcomes, and enclosing everything else in black boxes, will be of limited value to other practices in the region, the regional health authorities, and other settings within and beyond Portugal. Simulation modelling, informed by previous theoretical and empirical research, can provide insights which are more generalizable than single measures of economic impact.

By explicitly accounting for the behaviours of patients and physicians explored in previous research, and acknowledging that these are contained within a wider context, I have shown that practice-level statements of impact may not be applicable to the broader regional level. I have also identified instances of counter-intuitive, unexpected and unintended consequences. I have done this by transparently testing instituted mental models using a simulation model. The evaluation of complex health interventions must be performed using the right techniques and methodologies. In Part 2 of this thesis, culminating in this chapter, I have shown how the approach suggested in Chapter 3 can be applied in practice.

8.8.1 Implementation of results and stakeholder acceptance

As mentioned throughout this chapter, numerous stakeholders contributed in various stages of model conception, development, validation and testing. It is important to note this was not a group model building exercise. While parts of initial models were shown to stakeholders in Alentejo, it was clear early on they were uncomfortable when presented with actual system dynamics diagrams and had no intention of actually participating in model development. As with Lane et al. (2003), showing high-level simplified causal diagrams rather than stock and flow diagrams resulted in more involvement from stakeholders. No concerns were ever expressed over the validity of preliminary results or findings. A possible reason for this is that stakeholders were thinking about the system as whole as opposed to their own limited remit, perhaps for the first time. It is difficult to say exactly what stakeholders were feeling, as I did not collect their views of, and opinions on, the modelling process. While undoubtedly interesting, such an exploration was considered beyond the resources available for this study.

Besides numerous meetings with specific stakeholders (e.g., the chief dermatologist at Hospital do Espírito Santo de Évora and the Director of Alentejo Telemedicine Programme) throughout the modelling effort, I presented my work at yearly regional meetings, in which specialists, GPs and managers were present. There was frequently a disposition to focus on the positive findings (e.g., reductions in waiting times and patient costs) and disregard the potentially problematic areas (e.g., imbalances between teledermatology and face-to-face practices, costs, etc.).

It is too early to tell whether the findings presented here have had, or will have, any impact in practice. A ministerial working group – set up by the Ministry for Health and chaired by the Director of the Alentejo Telemedicine Programme – is currently developing a national Telemedicine Programme building on the experiences in Alentejo. The group's priorities include: advancing the use of teleconsultation in other regions of Portugal; piloting a remote

monitoring solution for chronic obstructive pulmonary disease patients; and ensuring lawmakers and payers develop legislation and policies that promote the use of telemedicine. It would be desirable to incorporate the findings presented here into discussions concerning the use of teleconsultation in other parts of Portugal.

Chapter 9

DISCUSSION AND CONCLUDING

REMARKS

Complex interventions have the potential to transform health care. They shift the location of care, affecting multiple parts of the system simultaneously. They challenge professional boundaries, roles and responsibilities. Their scope is diffuse and dynamic as a consequence of learning and reinvention. They have broad effects, both clinical and non-clinical, and within and beyond the health care system. They are flexible and contextual, tailored to meet the needs and preferences of specific populations and settings. It is their transformative power that creates challenges to evaluation.

The use of conventional economic evaluation methods in the context of complex interventions is hindered in important ways. The choice of scale, time frame and comparator can make the difference between negative and positive results. Findings can differ considerably across settings, and generalizability is affected by many factors, from demographics to baseline and initial conditions (Hauck et al., 2003). Because complex interventions have multiple effects, focusing on a single outcome of interest is too narrow a

measure of impact (Craig et al., 2008). It is arguable whether economic evaluation can inform treatment priorities, since the same intervention can have different costs and consequences in different settings, making it difficult to establish priorities based on a single evaluation study. Finally, the assumption that interventions can be standardised, defined in terms of essential components which applied anywhere will produce the same effects, does not hold for complex interventions, where it is precisely diversity that makes them succeed or fail (MacKenzie et al., 2010).

It is unrealistic to assess complex health interventions without taking into account the behaviours of those delivering and receiving them; it is precisely their actions and choices that determine how, why and whether complex interventions work. I have suggested that behaviours do not have to be locked within black boxes in evaluations of complex interventions (Anderson, 2008). It is possible to start opening those boxes, exploring what lies within. It is unlikely that a single method will provide comprehensive answers. A mix of methodologies and approaches, and the use of both experimental and non-experimental data, can be beneficial (Drummond et al., 2007, Canadian Agency for Drugs and Technology in Health, 2006, Walach et al., 2006). And, although the use of multiple methodologies is challenging, it leads to a more comprehensive and robust assessment of impact, building on the strengths of different approaches (McQueen and Anderson, 1999).

In this thesis, I have proposed an approach to the evaluation of complex health interventions which explicitly addresses the requirements set out above. The approach is illustrated in Figure 9-1. The overarching requirement is that evaluation should be about how and why an intervention works, not just whether it works. The approach suggested here brings together tools which can explicitly deal with complexity, answering how and why questions.

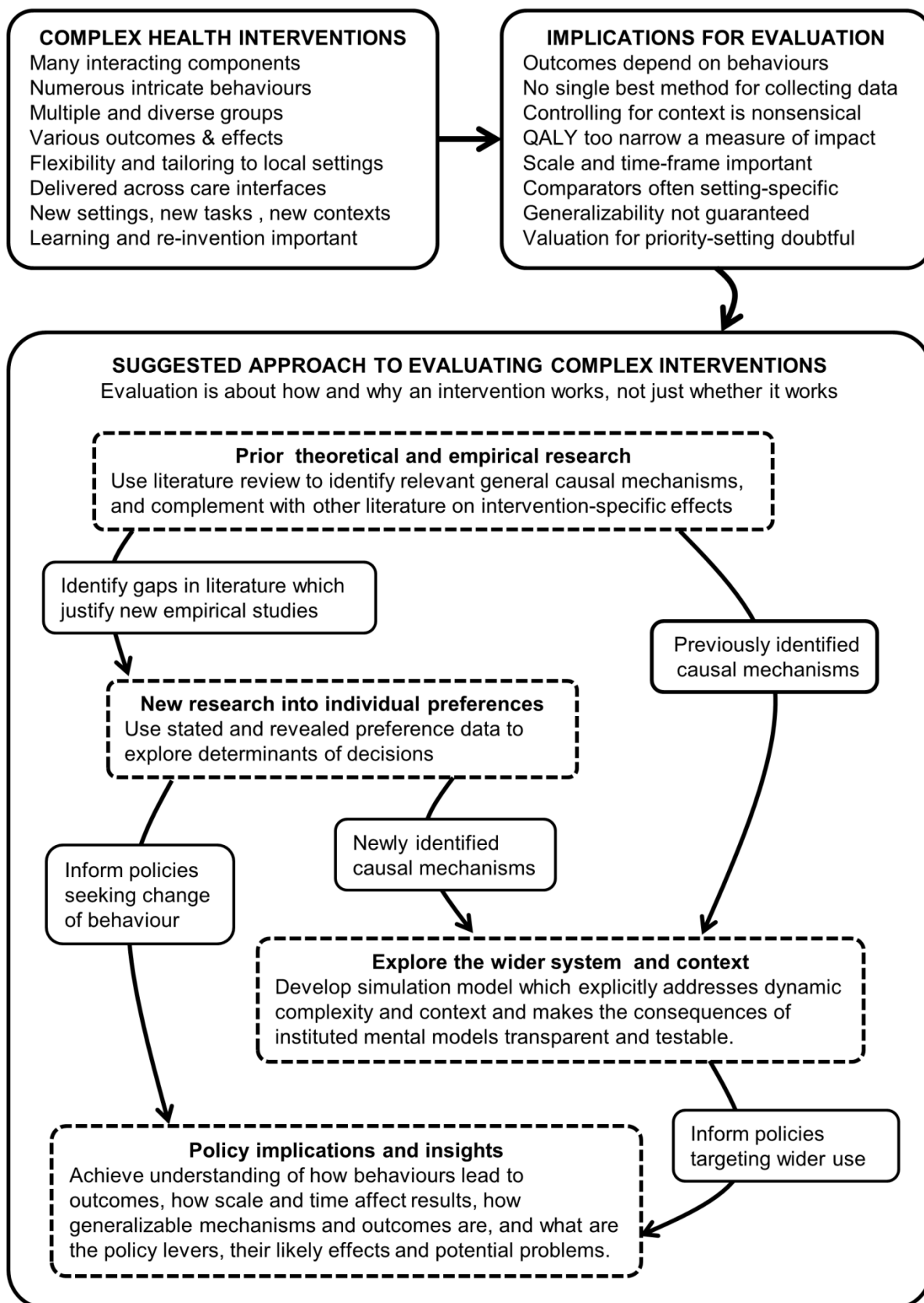


Figure 9-1 Suggested approach to evaluating complex health interventions

9.1 Summary

As depicted in Figure 9-1, to assess the impact of a complex intervention, one must understand how its costs and consequences emerge from the decisions and behaviours of those delivering and receiving it. The causal mechanisms behind the observed outcomes must be made explicit and their system-wide impact explored. As with other evaluation methods, one should begin by reviewing previous evidence on the impact of the intervention. This should be complemented and informed by an understanding of general causal mechanisms in the outpatient setting. Reviewing previous knowledge provides a picture of instituted mental models about how and why an intervention works, as well as an identification of gaps in existing empirical and theoretical research.

New studies can then be undertaken to fill these gaps, informing the evaluation of the specific intervention, and potentially leading to wider contributions to knowledge. The causal mechanisms identified in the reviews, and in the new studies, can then inform the development of a simulation model, which adopts a holistic systems perspective. Previous assertions of how and why an intervention works can then be tested, and the conditions under which it could work elsewhere, explored. By explicitly addressing dynamic complexity and context, the suggested approach allows causal mechanisms identified previously to be incorporated in the evaluation, rather than controlled for, or simply *hidden* in black boxes.

The literature review in Chapter 2 provides a starting point for identifying potentially important causal mechanisms. It describes a wealth of empirical and theoretical research that is so often overlooked by conventional economic evaluation methods. As shown in Figure 2-2, the number of factors investigated in the literature is astounding. Furthermore, there are numerous indications that behaviours are neither static nor linear. Siciliani and Hurst (2005), and van Ackere and Smith (1999), talk of feedback effects from waiting times to the quantities of health care demanded and supplied. Cutler and Huckman (2003), and Van de Wetering et al. (2012), caution that the instantaneous effects of a medical

intervention can differ substantially from impact at a later date. Schmalzried and Liszak (2012), Butler et al. (2001), and Skinner et al. (2006) all provide examples of well-intentioned actions by patients, physicians and policy makers that have unexpected and unintended outcomes. There is a recognition that feedback and dynamic complexity are present and relevant, yet approaches that take a wider systems perspective are found lacking.

Naturally, the literature review in Chapter 2 must be complemented with a review of the evidence on the specific intervention being assessed. In Part 2 of this thesis, I explored how the suggested approach can be applied to the evaluation of teleconsultation in Alentejo, Portugal. In Chapters 4 and 5, I reviewed previous knowledge on the impact of teleconsultation and explored its use in Alentejo. These chapters provide a picture of instituted mental models of researchers and practitioners about how and why the intervention works, and often numerous statements about whether it would work in other settings. The lack of a holistic systems perspective greatly limits the value of these statements, which frequently go untested.

Reviewing prior studies, and exploring how they inform our understanding of impact in a specific setting, also leads to an identification of the gaps in empirical and theoretical knowledge. These can be gaps in the literature about behavioural mechanisms and preferences which have implications for a wider range of interventions, as well as gaps in our understanding of intervention-specific mechanisms. New studies can then be undertaken to address those limitations. In Chapters 6 and 7, data was collected on revealed and stated preferences and appropriate econometric and statistical methods were used based on the characteristics of the data and the questions being asked. The studies informed the evaluation of teleconsultation and contributed more broadly to existing knowledge on patients and physicians' behaviours. Some of the insights produced by undertaking these studies are generalizable to evaluations of other complex interventions.

The causal mechanisms identified in the literature review and in the new studies of preferences can then inform the development of a simulation model which takes into account the wider system and context in which individual actions take place. A holistic systems perspective was adopted in Chapter 8, which explicitly acknowledged that physicians delivering, and patients receiving, teleconsultation operate in a wider context and affect, and are affected by, the system around them. Simulation modelling allowed previous assertions of how and why teleconsultation works, and whether it would work in other settings and contexts, to be tested. Simulation models make the consequences of existing mental models transparent and testable.

The end product is a greater understanding of how and why teleconsultation works; what are the dominant causal mechanisms, preferences and behaviours at play; how contextual differences, scales and time-frames affect the outcomes; and what are the likely effects of implementing teleconsultation elsewhere. The reasons behind counter-intuitive behaviours and unexpected outcomes are now clearer. Furthermore, insights from the simulation modelling exercise and empirical studies of preferences make contributions to previous research, informing future work.

9.2 Contributions

This thesis makes a number of contributions to extant research. The literature review in Chapter 2 provides a comprehensive overview of theoretical and empirical research into the various decisions available to physicians and patients in the outpatient or ambulatory setting. In it, numerous determinants are identified and their expected impact on decisions discussed. Besides providing a pool of potentially important mechanisms for inclusion in subsequent evaluations of complex interventions, the review may potentially inform future studies of decision-making and preferences.

The study undertaken in Chapter 6 illustrates how to address analytical challenges in health care utilisation data using panel models of count data. The use of a count model obviates the need to transform the data to deal with zeros which remain after temporal and spatial aggregation. Recent extensions of count models to panel data (Trivedi, 2010) allow unobserved heterogeneity and state dependence to be addressed explicitly. The elasticity of demand estimated in the study is comparable to those previously reported in the literature (Martin and Smith, 1999, Martin and Smith, 2003, Windmeijer et al., 2005), with the added benefit that actual referrals are used in the estimation, rather than appointments. The study also contributes to the literature on non-completion (Forrest et al., 2006, Forrest et al., 2007), which has not previously provided estimates of the elasticity of non-completion with respect to waiting time.

Chapter 7 describes what is to my knowledge, the first application of a DCE to the study of GP referrals. The method allows patients and physicians' preferences to be assessed separately, which is typically not possible with revealed preference data. By exploring only physicians' decisions, it is possible to determine whether they value the same attributes as patients. Findings suggest that measures of access such as distance and waiting time do not have a statistically significant impact on GPs' decisions to refer. Yet, GPs who practice farther away from the hospital are overall more likely to refer. I also find that having a special interest in dermatology, and practising teleconsultation, reduce the likelihood of a referral being made, motivating further work into the impact of continued GP education on referral rates. This study also analyses the determinants of referral priority, finding that GPs are more likely to refer clinically less severe patients as urgent when average waiting times increase.

The simulation model presented in Chapter 8 builds on previous work by seeking to *"open up the 'black boxes' that we label demand"* (van Ackere and Smith (1999), page 245). Namely, the model incorporates findings from econometric analyses of *"the GP's decision as to whether or not to refer to an NHS specialist"* and *"the patient's decision as to whether or*

not to attend the specialist appointment" (van Ackere and Smith (1999), page 245). It builds on previous work demonstrating "*how it is possible to embed a simple static economic model within a dynamic framework using the systems dynamics methodology*" ((Smith and van Ackere, 2002), page 8). Estimates from Chapters 6 and 7 are used to parameterise and calibrate the model, and confidence intervals are used to test the sensitivity of simulation results to uncertainty in the true value of parameters. The work described in Chapter 8 further contributes to the literature using system dynamics, and generally simulation modelling, to conduct economic evaluations of health interventions (NHS Improving Quality, 2013).

The approach suggested in this thesis has the potential to address the difficulties in using conventional economic evaluation methods to assess complex health interventions. Feedback, delays and non-linearities are explicitly handled rather than treated as a nuisance. A large number of outcomes and effects can be assessed in isolation or combined. Data from multiple sources, analysed through numerous methodologies, can be used. The sensitivity of the results to assumptions can be tested. Simulations can be performed at different scales and time-frames. Multiple contextual comparators can be included. Tailoring to specific contexts can be made explicit using different formulations, parameters or both. Using stated preference studies, insights about rare events can be generated, and regions where existing data are sparse can be explored. And finally, any valuation method can be used, allowing cost-effectiveness and cost-benefit ratios to be calculated, as well as the use of other decision-making aids, such as multi-criteria decision analysis (Santos et al., 2002). As Part 2 of this thesis illustrates, the approach can lead to a number of policy implications.

9.3 Policy implications

Teleconsultations are often hailed as a means to reduce imbalances in supply of health care services but they may actually exacerbate their effects. While the many benefits of

teleconsultation reported in the literature are confirmed, they are enjoyed by only a few, with face-to-face patients worse off. Face-to-face capacity is diminished as a consequence of remaining flows from teledermatology practices to the face-to-face pathway. The previous method of allocating capacity on the basis of a waiting list is effectively shut down by the introduction of a parallel teledermatology waiting list. The two mechanisms compete for access to outpatient dermatological care, with teledermatology seemingly having the upper hand.

Assertions that implementation of teledermatology at a wider level would lead to substantial benefits are questionable as they depend on how well capacity meets underlying demand. If capacity is insufficient, there are diminishing returns to the wider adoption of teledermatology. Service capacity, productivity and utilisation should improve, but waiting times would not disappear. Also, enclosing dynamic effects and behaviour within black boxes is both unrealistic and unnecessary. Focusing on final measures of utilisation would ignore the impact of continued GP education and learning on referrals, underestimating the economic impact of teleconsultation.

The application of the proposed approach in practice indicates that quantitative costs and consequences can be highly contextual but mechanisms are potentially generalizable. There are at least two examples: the level of supply and the choice of comparator. Regarding the first, the simulation results are generalizable to other settings where supply struggles to meet demand. When there is spare capacity or waiting times are already low, wider use of teledermatology will indeed result in wider benefits. Naturally, the actual numbers will be setting-specific. As for the choice of comparator, it is clearly dependent on context. Recruiting dermatologists might be an obvious solution in Lisbon, yet it is hardly an option in Alentejo, a region with a chronic undersupply of dermatologists. Comparators must make sense in the setting in which evaluation is taking place.

9.4 Limitations and future work

Limitations were discussed at the end of each study in Chapters 6 through 8, and were mostly related to the quality and completeness of the data. Indeed, evaluating complex interventions is difficult in part due to poor data availability. However, this thesis shows that it is possible to achieve insights into the impact of interventions even when data is limited, and that methods and techniques are available to make the most of what information is available.

As illustrated here, stated preference methods can be used to complement poor or limited revealed preference data. The use of stated preference data and simulation modelling can confirm whether further data collection is justified or not, by identifying the mechanisms which are likely to have a significant impact on results. This way, future data collection can focus on these priority areas. Evaluation of complex health interventions should be seen as an ongoing iterative process, whereby older assessments inform future studies.

Another avenue for future research is the use and combination of multiple simulation modelling approaches, informed by more sophisticated econometric studies. For example, in Chapter 6, I suggested that more complex designs could be used to explore respondent heterogeneity (i.e., the possibility that different GPs had different preferences). These might be incorporated into a simulation model through the use of cohorts in system dynamics or by adopting a different approach altogether, such as agent-based modelling.

At the more general level of the approach proposed here, there is uncertainty regarding whether it can inform the evaluation of other complex interventions in the outpatient setting and whether it is generalizable to complex interventions in other settings, beyond outpatient care. Although the suggested approach should be flexible enough to account for the contextual specificities of any intervention, this should be explored in the future.

REFERENCES

- ADMINISTRAÇÃO CENTRAL DO SISTEMA DE SAÚDE I.P. 2008. Rede de referência hospitalar de dermatologia. *In: HEALTH, M. O. (ed.)*.
- ADMINISTRAÇÃO CENTRAL DO SISTEMA DE SAÚDE I.P. 2012. Contrato-programa 2013. *In: HEALTH, M. O. (ed.)*.
- ADMINISTRAÇÃO REGIONAL DE SAÚDE DO ALENTEJO I.P. 2012. Telemedicina – Actividade 1998-2011. Administração Regional de Saúde do Alentejo I.P.,.
- AGHA, Z., SCHAPIRA, R. M. & MAKER, A. H. 2002. Cost effectiveness of telemedicine for the delivery of outpatient pulmonary care to a rural population. *Telemedicine Journal and E-Health*, 8, 281-291.
- ALKMIM, M. B., FIGUEIRA, R. M., MARCOLINO, M. S., CARDOSO, C. S., ABREU, M. P. D., CUNHA, L. R., CUNHA, D. F. D., ANTUNES, A. P., RESENDE, A. G. D. A. & RESENDE, E. S. 2012. Improving patient access to specialized health care: the Telehealth Network of Minas Gerais, Brazil. *Bulletin of the World Health Organization*, 90, 373-378.
- ANDERSEN, R. M. 1995. Revisiting the behavioral model and access to medical care: does it matter? *Journal of health and social behavior*, 1-10.
- ANDERSON, J., BUCK, C., DANAHAR, K. & FRY, J. 1977. Users and non-users of doctors—implications for self-care. *The Journal of the Royal College of General Practitioners*, 27, 155.
- ANDERSON, R. 2008. New MRC guidance on evaluating complex interventions. *BMJ*, 337.
- ARMSTRONG, A. W., DORER, D. J., LUGN, N. E. & KVEDAR, J. C. 2007. Economic evaluation of interactive teler dermatology compared with conventional care. *Telemedicine and e-Health*, 13, 91-99.
- ARMSTRONG, D., FRY, J. & ARMSTRONG, P. 1991. Doctors' perceptions of pressure from patients for referral. *BMJ: British Medical Journal*, 302, 1186.
- ARROW, K. J. 1963. Uncertainty and the welfare economics of medical care. *The American economic review*, 53, 941-973.
- ASSOCIAÇÃO PORTUGUESA PARA A PROMOÇÃO E DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO & ASSOCIAÇÃO PARA O DESENVOLVIMENTO DA TELEMEDICINA 2007. Telemedicina – Onde estamos e para onde vamos.
- AUCHINCLOSS, A. H. & ROUX, A. V. D. 2008. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. *American journal of epidemiology*, 168, 1-8.
- AYANIAN, J. Z. 2006. Rising Rates of Cardiac Procedures in the United States and Canada Too Much of a Good Thing? *Circulation*, 113, 333-335.
- BAILEY, A. M., THORNE, B. C. & PEIRCE, S. M. 2007. Multi-cell agent-based simulation of the microvasculature to study the dynamics of circulating inflammatory cell trafficking. *Annals of biomedical engineering*, 35, 916-936.

- BAKER, L., BIRNBAUM, H., GEPPERT, J., MISHOL, D. & MOYNEUR, E. 2003. The relationship between technology availability and health care spending. *HEALTH AFFAIRS-MILLWOOD VA THEN BETHESDA MA-*, W3-537.
- BAKER, L. C., AFENDULIS, C. C. & ATLAS, S. W. 2010. Assessing Cost-Effectiveness And Value As Imaging Grows: The Case Of Carotid Artery CT. *Health Affairs*, 29, 2260-2267.
- BALDWIN, L., CLARKE, M., HANDS, L., KNOTT, M. & JONES, R. 2003. The effect of telemedicine on consultation time. *Journal of telemedicine and telecare*, 9, 71-73.
- BARLAS, Y. 1996. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, 12, 183-210.
- BARLAS, Y. & CARPENTER, S. 1990. Philosophical roots of model validation: two paradigms. *System Dynamics Review*, 6, 148-166.
- BARRETT, M. S., CHUA, W.-J., CRITS-CHRISTOPH, P., GIBBONS, M. B. & THOMPSON, D. 2008. Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy: Theory, Research, Practice, Training*, 45, 247.
- BARROS, P. P., MACHADO, S. R. & SIMÕES, J. A. 2011. Portugal. Health system review. *Health systems in transition*, 13, 1.
- BARTON, P., BRYAN, S. & ROBINSON, S. 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy*, 9, 110-118.
- BAUM, C. F., SCHAFFER, M. E. & STILLMAN, S. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal*, 3, 1-31.
- BAYER, S., BARLOW, J. & CURRY, R. 2007. Assessing the impact of a care innovation: telecare. *System Dynamics Review*, 23, 61-80.
- BAYER, S., BOLT, T., KAPSALI, M. & BRAILSFORD, S. The social role of simulation models. International System Dynamics Conferences, 2010.
- BAYER, S., BRAILSFORD, S. & BOLT, T. Examining the role of simulation models in health planning. Proceedings of the 27th International Conference of the System Dynamics Society, 2009. 26-30.
- BELONGIA, E. A. & SCHWARTZ, B. 1998. Strategies for promoting judicious use of antibiotics by doctors and patients. *BMJ: British Medical Journal*, 317, 668.
- BERGMO, T. S. 2009. Can economic evaluation in telemedicine be trusted? A systematic review of the literature. *Cost Effectiveness and Resource Allocation*, 7, 18.
- BERGMO, T. S. 2010. Economic evaluation in telemedicine—still room for improvement. *Journal of Telemedicine and Telecare*, 16, 229-231.
- BERTAKIS, K. D., CALLAHAN, E. J., AZARI, R. & ROBBINS, J. A. 2001. Predictors of patient referrals by primary care residents to specialty care clinics. *Fam Med*, 33, 203-9.
- BLACKWOOD, B. 2006. Methodological issues in evaluating complex healthcare interventions. *Journal of Advanced Nursing*, 54, 612-622.
- BLAND, J. M. & ALTMAN, D. G. 1996a. Transformations, means, and confidence intervals. *BMJ: British Medical Journal*, 312, 1079.
- BLAND, J. M. & ALTMAN, D. G. 1996b. The use of transformation when comparing two means. *BMJ: British Medical Journal*, 312, 1153.
- BORSHCHEV, A. & FILIPPOV, A. From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. Proceedings of the 22nd international conference of the system dynamics society, 2004.
- BOSANQUET, N. 2009. Technology: scientific force or power force? *The Economics of New Health Technologies: Incentives, Organization, and Financing*, 43.
- BRAILSFORD, S. & HILTON, N. 2001. A comparison of discrete event simulation and system dynamics for modelling health care systems.
- BRAILSFORD, S. C., DESAI, S. M. & VIANA, J. Towards the holy grail: combining system dynamics and discrete-event simulation in healthcare. Simulation Conference (WSC), Proceedings of the 2010 Winter, 2010. IEEE, 2293-2303.

- BRAILS福德, S. C., LATTIMER, V., TARNARAS, P. & TURNBULL, J. 2004. Emergency and on-demand health care: modelling a large complex system. *Journal of the operational research society*, 55, 34-42.
- BREHMER, B. 1987. Systems design and the psychology of complex systems. *Empirical Foundations of Information and Software Science III*. Springer.
- BRENNAN, A., CHICK, S. E. & DAVIES, R. 2006. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics*, 15, 1295-1310.
- BROWN, E. 1999. Assessing the value of volunteer activity. *Nonprofit and Voluntary Sector Quarterly*, 28, 3-17.
- BURGISS, S. G., JULIUS, C. E., WATSON, H. W., HAYNES, B. K., BUONOCORE, E. & SMITH, G. T. 1997. Telemedicine for dermatology care in rural patients. *Telemedicine Journal*, 3, 227-233.
- BURNAY, E., CRUZ-CORREIA, R., JACINTO, T., SOUSA, A. S. & FONSECA, J. 2013. Challenges of a Mobile Application for Asthma and Allergic Rhinitis Patient Enablement—Interface and Synchronization. *Telemedicine and e-Health*, 19, 13-18.
- BUTLER, C. C., KINNERSLEY, P., PROUT, H., ROLLNICK, S., EDWARDS, A. & ELWYN, G. 2001. Antibiotics and shared decision-making in primary care. *Journal of Antimicrobial Chemotherapy*, 48, 435-440.
- BYNG, R., NORMAN, I., REDFERN, S. & JONES, R. 2008. Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation. *BMC Health Serv Res*, 8, 274.
- CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press.
- CAMERON, A. C. & TRIVEDI, P. K. 2007. Essentials of Count Data Regression. A *Companion to Theoretical Econometrics*. Blackwell Publishing Ltd.
- CAMERON, A. C. & TRIVEDI, P. K. 2013. Count Panel Data.
- CAMERON, A. E., BASHSHUR, R. L., HALBRITTER, K., JOHNSON, E. M. & CAMERON, J. W. 1998. Simulation methodology for estimating financial effects of telemedicine in West Virginia. *Telemedicine Journal*, 4, 125-144.
- CAMERON, C. 1996. Patient compliance: recognition of factors involved and suggestions for promoting compliance with therapeutic regimens. *Journal of advanced nursing*, 24, 244-250.
- CAMPBELL, S. M. & ROLAND, M. O. 1996. Why do people consult the doctor? *Family practice*, 13, 75-83.
- CANADIAN AGENCY FOR DRUGS AND TECHNOLOGY IN HEALTH 2006. Guidelines for the economic evaluation of health technologies: Canada. 3rd ed. Ottawa: CADTH.
- CARLILE, P. R. 2002. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization science*, 13, 442-455.
- CARLSEN, B., AAKVIK, A. & NORHEIM, O. F. 2008. Variation in practice: a questionnaire survey of how congruence in attitudes between doctors and patients influences referral decisions. *Medical Decision Making*, 28, 262-268.
- CASTELA, E., RAMALHEIRO, G., PIRES, A., CARREIRA, L. M., SANTOS, I., COSTA, H., MOTA, A. & RIBEIRO, L. 2005. Five years of teleconsultation: experience of the Cardiology Department of Coimbra Pediatric Hospital. *Revista portuguesa de cardiologia: orgao oficial da Sociedade Portuguesa de Cardiologia= Portuguese journal of cardiology: an official journal of the Portuguese Society of Cardiology*, 24, 835-840.
- CENTRE FOR REVIEWS AND DISSEMINATION 2012. Evidence briefing on teleconsultation. Centre for Reviews and Dissemination.
- CHAMBERLAIN, G. 1992. Comment: sequential moment restrictions in panel data. *Journal of Business & Economic Statistics*, 10, 20-26.
- CHANDRA, A., CUTLER, D. & SONG, Z. 2012. Who ordered that? The economics of treatment choices in medical care. *Handbook of health economics*, 2, 397-432.
- CHEN, F. M. 2001. Dermatology referrals in an academic family medicine clinic. *Southern medical journal*, 94, 475&hyphen.

- CHEN, F. M., FRYER, G. E. & NORRIS, T. E. 2005. Effects of comorbidity and clustering upon referrals in primary care. *The Journal of the American Board of Family Practice*, 18, 449-452.
- CHERAGHI-SOHI, S., HOLE, A. R., MEAD, N., MCDONALD, R., WHALLEY, D., BOWER, P. & ROLAND, M. 2008. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. *The Annals of Family Medicine*, 6, 107-115.
- CHERNEW, M. E., HIRTH, R. A. & CUTLER, D. M. 2003. Increased spending on health care: how much can the United States Afford? *Health Affairs*, 22, 15-25.
- CHERNEW, M. E., HIRTH, R. A. & CUTLER, D. M. 2009. Increased spending on health care: long-term implications for the nation. *Health Affairs*, 28, 1253-1255.
- CLAXTON, A. J., CRAMER, J. & PIERCE, C. 2001. A systematic review of the associations between dose regimens and medication compliance. *Clinical therapeutics*, 23, 1296-1310.
- CLAXTON, K., SCULPHER, M. & DRUMMOND, M. 2002. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). *The Lancet*, 360, 711-715.
- COCKBURN, J. & PIT, S. 1997. Prescribing behaviour in clinical practice: patients' expectations and doctors' perceptions of patients' expectations--a questionnaire study. *BMJ: British Medical Journal*, 315, 520.
- COLLINS, K., WALTERS, S. & BOWNS, I. 2004. Patient satisfaction with teledermatology: quantitative and qualitative results from a randomized controlled trial. *Journal of telemedicine and telecare*, 10, 29-33.
- COOPER, K., BRAILSFORD, S. & DAVIES, R. 2006. Choice of modelling technique for evaluating health care interventions. *Journal of the operational research society*, 58, 168-176.
- COSTA, P. D., REIS, A. H. & RODRIGUES, P. P. 2013. Clinical and Economic Impact of Remote Monitoring on the Follow-Up of Patients with Implantable Electronic Cardiovascular Devices: An Observational Study. *Telemedicine and e-Health*, 19, 71-80.
- CRAIG, P., DIEPPE, P., MACINTYRE, S., MICHIE, S., NAZARETH, I. & PETTICREW, M. 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ: British Medical Journal*, 337.
- CRAMER, J. A. 2004. A systematic review of adherence with medications for diabetes. *Diabetes care*, 27, 1218-1224.
- CRAVO OLIVEIRA, T., BARLOW, J., GONÇALVES, L. & BAYER, S. 2013. Teleconsultations reduce greenhouse gas emissions. *J Health Serv Res Policy*, 18, 209-214.
- CRAVO OLIVEIRA, T., BRANQUINHO, M. J. & GONÇALVES, L. 2012. State of the Art in Telemedicine—Concepts, Management, Monitoring and Evaluation of the Telemedicine Programme in Alentejo (Portugal). *Perspectives on Digital Pathology: Results of the Cost Action IC0604 Euro-Telepath*, 179, 29.
- CULLIS, J. G., JONES, P. R. & PROPPER, C. 2000. Waiting lists and medical treatment: analysis and policies. *Handbook of health economics*, 1, 1201-1249.
- CUSACK, C. M., PAN, E., HOOK, J. M., VINCENT, A., KAELEBER, D. C. & MIDDLETON, B. 2008. The value proposition in the widespread use of telehealth. *Journal of Telemedicine and Telecare*, 14, 167-168.
- CUTLER, D. M. 1995. Technology, health costs, and the NIH. *National Institutes of Health Roundtable on the Economics of Biomedical Research*.
- CUTLER, D. M. & HUCKMAN, R. S. 2003. Technological development and medical productivity: the diffusion of angioplasty in New York state. *Journal of Health Economics*, 22, 187-217.
- CUTLER, D. M. & MCCLELLAN, M. 2001. Is technological change in medicine worth it? *Health Affairs*, 20, 11-29.

- DAFNIS, G., GRANATH, F., PÅHLMAN, L., EKBOM, A. & BLOMQVIST, P. 2005. Patient factors influencing the completion rate in colonoscopy. *Digestive and liver disease*, 37, 113-118.
- DARZI, A. 2008. High quality care for all: NHS next stage review final report. *London: Department of Health*, 2009.
- DÁVALOS, M. E., FRENCH, M. T., BURDICK, A. E. & SIMMONS, S. C. 2009. Economic evaluation of telemedicine: review of the literature and research guidelines for benefit–cost analysis. *Telemedicine and e-Health*, 15, 933-948.
- DAVIS, J. P., EISENHARDT, K. M. & BINGHAM, C. B. 2007. Developing theory through simulation methods. *Academy of Management Review*, 32, 480-499.
- DE BEKKER-GROB, E., ESSINK-BOT, M., MEERDING, W., POLS, H., KOES, B. & STEYERBERG, E. 2008. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. *Osteoporosis International*, 19, 1029-1037.
- DE BEKKER-GROB, E. W., RYAN, M. & GERARD, K. 2012. Discrete choice experiments in health economics: a review of the literature. *Health economics*, 21, 145-172.
- DEMARTINES, N., FREIERMUTH, O., MUTTER, D., HEBERER, M. & HARDER, F. 2000. Knowledge and acceptance of telemedicine in surgery: a survey. *Journal of telemedicine and telecare*, 6, 125-131.
- DEPARTMENT OF HEALTH 2013a. Government Response to the House of Commons Health Select Committee Report into Public Expenditure on Health and Care Services (Eleventh Report of Session 2012–13) *In: HEALTH, D. O. (ed.). TSO*.
- DEPARTMENT OF HEALTH 2013b. The NHS Constitution - the NHS belongs to us all.
- DESHPANDE, A. 2008. *Real-time (synchronous) telehealth in primary care: systematic review of systematic reviews*, Canadian Agency for Drugs and Technologies in Health - Agence Canadienne des Médicaments et des Technologies de la Santé.
- DOARN, C. R. & MERRELL, R. C. 2013. Have you seen the light? *Telemedicine and e-Health*, 19, 329-329.
- DODGSON, M., GANN, D. M. & SALTER, A. 2007a. The impact of modelling and simulation technology on engineering problem solving. *Technology Analysis & Strategic Management*, 19, 471-489.
- DODGSON, M., GANN, D. M. & SALTER, A. 2007b. "In Case of Fire, Please Use the Elevator": Simulation Technology and Organization in Fire Engineering. *Organization science*, 18, 849-864.
- DONALDSON, L. 2003. Expert patients usher in a new era of opportunity for the NHS. *BMJ: British Medical Journal*, 326, 1279.
- DONOVAN, J. L. & BLAKE, D. R. 1992. Patient non-compliance: deviance or reasoned decision-making? *Social science & medicine*, 34, 507-513.
- DOOLITTLE, G. C., WILLIAMS, A. R., SPAULDING, A., SPAULDING, R. J. & COOK, D. J. 2004. A cost analysis of a tele-oncology practice in the United States. *Journal of telemedicine and telecare*, 10, 27-29.
- DORRIAN, C., FERGUSON, J., AH-SEE, K., BARR, C., LALLA, K., VAN DER POL, M., MCKENZIE, L. & WOOTTON, R. 2009. Head and neck cancer assessment by flexible endoscopy and telemedicine. *Journal of telemedicine and telecare*, 15, 118-121.
- DRUMMOND, M., WEATHERLY, H., CLAXTON, K., COOKSON, R., FERGUSON, B., GODFREY, C., RICE, N., SCULPHER, M. & SOWDEN, A. 2007. Assessing the challenges of applying standard methods of economic evaluation to public health interventions. *Public Health Research Consortium: Final Report*.
- EDISON, K. E., CHANCE, L., MARTIN, K., BRAUDIS, K. & WHITED, J. D. 2011. Users and nonusers of university-based dermatology services following a teledermatology encounter: a retrospective descriptive analysis. *Telemedicine and e-Health*, 17, 14-18.
- EDISON, K. E., WARD, D. S., DYER, J. A., LANE, W., CHANCE, L. & HICKS, L. L. 2008. Diagnosis, diagnostic confidence, and management concordance in live-interactive

- and store-and-forward teledermatology compared to in-person examination. *Telemedicine and e-Health*, 14, 889-895.
- EDWARDS, M. A. & PATEL, A. C. 2003. Telemedicine in the state of Maine: A model for growth driven by rural needs. *Telemedicine Journal and e-Health*, 9, 25-39.
- EKELAND, A. G., BOWES, A. & FLOTTORP, S. 2010. Effectiveness of telemedicine: a systematic review of reviews. *International journal of medical informatics*, 79, 736-771.
- EMANUEL, E. J. & FUCHS, V. R. 2008. The perfect storm of overutilization. *JAMA: the journal of the American Medical Association*, 299, 2789-2791.
- EMINOVIC, N., DE KEIZER, N. F., WYATT, J. C., TER RIET, G., PEEK, N., VAN WEERT, H. C., BRUIJNZEEL-KOOMEN, C. A. & BINDELS, P. J. 2009. Teledermatologic consultation and reduction in referrals to dermatologists: a cluster randomized controlled trial. *Archives of dermatology*, 145, 558.
- EMINOVIĆ, N., DIJKGRAAF, M. G., BERGHOUT, R. M., PRINS, A. H., BINDELS, P. J. & DE KEIZER, N. F. 2010. A cost minimisation analysis in teledermatology: model-based approach. *BMC health services research*, 10, 251.
- EMINOVIĆ, N., WITKAMP, L., RAVELLI, A., BOS, J., VAN DEN AKKER, T. W., BOUSEMA, M., HENQUET, C., KOOPMAN, R., ZEEGELAAR, J. & WYATT, J. 2003. Potential effect of patient-assisted teledermatology on outpatient referral rates. *Journal of telemedicine and telecare*, 9, 321-327.
- ENSOR, T. & COOPER, S. 2004. Overcoming barriers to health service access: influencing the demand side. *Health policy and planning*, 19, 69-79.
- EVENDEN, D., HARPER, P., BRAILSFORD, S. & HARINDRA, V. 2005a. Improving the cost-effectiveness of Chlamydia screening with targeted screening strategies. *Journal of the operational research society*, 57, 1400-1412.
- EVENDEN, D., HARPER, P. R., BRAILSFORD, S. C. & HARINDRA, V. 2005b. System Dynamics modeling of Chlamydia infection for screening intervention planning and cost-benefit estimation. *IMA Journal of Management Mathematics*, 16, 265-279.
- FATEHI, F. & WOOTTON, R. 2012. Telemedicine, telehealth or e-health? A bibliometric analysis of the trends in the use of these terms. *Journal of telemedicine and telecare*, 18, 460-464.
- FERTIG, A., ROLAND, M., KING, H. & MOORE, T. 1993. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *BMJ: British Medical Journal*, 307, 1467.
- FINKELSTEIN, S. R., LIU, N., JANI, B., ROSENTHAL, D. & POGHOSYAN, L. 2013. Appointment reminder systems and patient preferences: Patient technology usage and familiarity with other service providers as predictive variables. *Health informatics journal*, 19, 79-90.
- FOLLAND, S., GOODMAN, A. & STANO, M. 2006. Economics of health and health care.
- FONE, D., HOLLINGHURST, S., TEMPLE, M., ROUND, A., LESTER, N., WEIGHTMAN, A., ROBERTS, K., COYLE, E., BEVAN, G. & PALMER, S. 2003. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health*, 25, 325-335.
- FORREST, C. B. 2011. Pediatric Subspecialty Referral Completion. *The Journal of pediatrics*, 158, 3-4.
- FORREST, C. B., NUTTING, P. A., VON SCHRADER, S., ROHDE, C. & STARFIELD, B. 2006. Primary care physician specialty referral decision making: patient, physician, and health care system determinants. *Medical decision making*, 26, 76-85.
- FORREST, C. B., SHADMI, E., NUTTING, P. A. & STARFIELD, B. 2007. Specialty referral completion among primary care patients: results from the ASPN Referral Study. *The Annals of Family Medicine*, 5, 361-367.
- FORRESTER, J. W. 1961. *Industrial dynamics*, MIT press Cambridge, MA.
- FORRESTER, J. W. 1971. Counterintuitive behavior of social systems. *Theory and Decision*, 2, 109-140.

- FORRESTER, J. W. & SENGE, P. M. 1978. *Tests for building confidence in system dynamics models*, System Dynamics Group, Sloan School of Management, Massachusetts Institute of Technology.
- FRANKS, P., WILLIAMS, G. C., ZWANZIGER, J., MOONEY, C. & SORBERO, M. 2000. Why do physicians vary so widely in their referral rates? *Journal of General Internal Medicine*, 15, 163-168.
- FRASER, S. W. & GREENHALGH, T. 2001. Complexity science: Coping with complexity: educating for capability. *BMJ: British Medical Journal*, 323, 799.
- GARCÍA RADA, A. 2011. Virtual hospital for HIV patients halves consultation times. *BMJ*, 342.
- GEORGE, A. & RUBIN, G. 2003. Non-attendance in general practice: a systematic review and its implications for access to primary health care. *Family Practice*, 20, 178-184.
- GERARD, K. & LATTIMER, V. 2005. Preferences of patients for emergency services available during usual GP surgery hours: a discrete choice experiment. *Family practice*, 22, 28-36.
- GERARD, K. & MOONEY, G. 1993. QALY league tables: handle with care. *Health Economics*, 2, 59-64.
- GERARD, K., SALISBURY, C., STREET, D., POPE, C. & BAXTER, H. 2008. Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences. *Journal of Health Services Research & Policy*, 13, 3-10.
- GHAFFARZADEGAN, N., LYNEIS, J. & RICHARDSON, G. P. 2011. How small system dynamics models can help the public policy process. *System Dynamics Review*, 27, 22-44.
- GILBERT, N. & TROITZSCH, K. G. 2005. *Simulation for the social scientist*, McGraw-Hill International.
- GILMOUR, E., CAMPBELL, S., LOANE, M., ESMAIL, A., GRIFFITHS, C., ROLAND, M., PARRY, E., CORBETT, R., EEDY, D. & GORE, H. 1998. Comparison of teleconsultations and face-to-face consultations: preliminary results of a United Kingdom multicentre teledermatology study. *British Journal of Dermatology*, 139, 81-87.
- GODDARD, J. & TAVAKOLI, M. 1998. Referral rates and waiting lists: some empirical evidence. *Health Economics*, 7, 545-549.
- GONZÁLEZ-BUSTO, B. & GARCIA, R. 1999. Waiting lists in Spanish public hospitals: a system dynamics approach. *System Dynamics Review*, 15, 201-224.
- GORMAN, D. M., MEZIC, J., MEZIC, I. & GRUENEWALD, P. J. 2006. Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice. *American Journal of Public Health*, 96, 2055-2060.
- GRANLUND, H., THODEN, C.-J., CARLSON, C. & HARNO, K. 2003. Realtime teleconsultations versus face-to-face consultations in dermatology: immediate and six-month outcome. *Journal of telemedicine and telecare*, 9, 204-209.
- GRAVELLE, H., DUSHEIKO, M. & SUTTON, M. 2002. The demand for elective surgery in a public system: time and money prices in the UK National Health Service. *Journal of Health Economics*, 21, 423-449.
- GREEN, B. N., JOHNSON, C. D. & ADAMS, A. 2006. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine*, 5, 101-117.
- GREENE, W. 2008. Functional forms for the negative binomial model for count data. *Economics Letters*, 99, 585-590.
- GRIFFITH, S. 1990. A review of the factors associated with patient compliance and the taking of prescribed medicines. *The British Journal of General Practice*, 40, 114.
- GRIFFITHS, W. A. D. 2010. Improving melanoma diagnosis in primary care – a tele-dermatoscopy project. *Journal of Telemedicine and Telecare*, 16, 185-186.
- GRIMSMO, A. & SIEM, H. 1984. Factors affecting primary health care utilization. *Family practice*, 1, 155-161.

- GROSSMAN, M. 2000. The human capital model. *Handbook of health economics*, 1, 347-408.
- GRUPO DE TRABALHO PARA A TELEMEDICINA. 2013. *Portal da Saúde - Grupos de trabalho* [Online]. Administração Central do Sistema de Saúde Available: <http://www.min-saude.pt/portal/conteudos/a+saude+em+portugal/informatizacao/grupos+trabalho.htm>
- GÜNAL, M. M. & PIDD, M. 2010. Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4, 42-51.
- HAKANSSON, S. & GAVELIN, C. 2000. What do we really know about the cost-effectiveness of telemedicine? *Journal of Telemedicine and Telecare*, 6, 133-136.
- HALL, J., FIEBIG, D. G., KING, M. T., HOSSAIN, I. & LOUVIERE, J. J. 2006. What influences participation in genetic carrier testing?: Results from a discrete choice experiment. *Journal of Health Economics*, 25, 520-537.
- HALL, J., KENNY, P., KING, M., LOUVIERE, J., VINEY, R. & YEOH, A. 2002. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. *Health Economics*, 11, 457-465.
- HANDY, F. & SRINIVASAN, N. 2004. Valuing volunteers: An economic evaluation of the net benefits of hospital volunteers. *Nonprofit and Voluntary Sector Quarterly*, 33, 28-54.
- HANNAY, D. 1978. Symptom prevalence in the community. *The Journal of the Royal College of General Practitioners*, 28, 492.
- HANNAY, D. 1980. The 'iceberg' of illness and 'trivial' consultations. *The Journal of the Royal College of General Practitioners*, 30, 551.
- HARRISON, J. R., LIN, Z., CARROLL, G. R. & CARLEY, K. M. 2007a. Simulation modeling in organizational and management research. *Academy of Management Review*, 32, 1229-1245.
- HARRISON, M. I., KOPPEL, R. & BAR-LEV, S. 2007b. Unintended consequences of information technologies in health care—an interactive sociotechnical analysis. *Journal of the American Medical Informatics Association*, 14, 542-549.
- HAUCK, K., GODDARD, M. & SMITH, P. C. 2003. *The economics of priority setting for health care: a literature review*, International Bank for Reconstruction and Development/World Bank.
- HAUSMAN, J. A., HALL, B. H. & GRILICHES, Z. 1984. Econometric models for count data with an application to the patents-R&D relationship. National Bureau of Economic Research Cambridge, Mass., USA.
- HEATH, S. K., BUSS, A., BRAILSFORD, S. C. & MACAL, C. M. Cross-paradigm simulation modeling: challenges and successes. Proceedings of the Winter Simulation Conference, 2011. Winter Simulation Conference, 2788-2802.
- HERSH, W., HELFAND, M., WALLACE, J., KRAEMER, D., PATTERSON, P., SHAPIRO, S. & GREENLICK, M. 2002. A systematic review of the efficacy of telemedicine for making diagnostic and management decisions. *Journal of Telemedicine and Telecare*, 8, 197-209.
- HICKS, L. L., BOLES, K. E., HUDSON, S., KLING, B., TRACY, J., MITCHELL, J. & WEBB, W. 2003. Patient satisfaction with teledermatology services. *Journal of telemedicine and telecare*, 9, 42-45.
- HILBE, J. M. 1994. Generalized linear models. *The American Statistician*, 48, 255-265.
- HILL, D. 2004. *Evaluation of community-level interventions for health improvement: a review of experience in the UK*, Health Development Agency London.
- HILLMAN, B. J. & GOLDSMITH, J. C. 2010. The uncritical use of high-tech medical imaging. *The New England journal of medicine*, 363, 4.
- HJELM, N. 2005. Benefits and drawbacks of telemedicine. *Journal of telemedicine and telecare*, 11, 60-70.
- HJELMGREN, J. & ANELL, A. 2007. Population preferences and choice of primary care models: a discrete choice experiment in Sweden. *Health policy*, 83, 314-322.

- HOMER, J., HIRSCH, G., MINNITI, M. & PIERSON, M. 2004. Models for collaboration: how system dynamics helped a community organize cost-effective care for chronic illness. *System Dynamics Review*, 20, 199-222.
- HOMER, J. B. 1993. A system dynamics model of national cocaine prevalence. *System Dynamics Review*, 9, 49-78.
- HOMER, J. B. & HIRSCH, G. B. 2006. System dynamics modeling for public health: background and opportunities. *American Journal of Public Health*, 96, 452-458.
- HOUSE OF COMMONS HEALTH COMMITTEE 2013. Public expenditure on health and care services: Eleventh Report of Session 2012–13, Volume I.
- HURLEY, J. 2000. An overview of the normative economics of the health sector. *Handbook of health economics*, 1, 55-118.
- HYER, R. N. 1999. Telemedical experiences at an Antarctic station. *Journal of telemedicine and telecare*, 5, 87-89.
- INSTITUTO NACIONAL DE ESTATÍSTICA I.P. 2010. Anuário estatístico da região Alentejo 2010. Instituto Nacional de Estatística I.P.
- INSTITUTO NACIONAL DE ESTATÍSTICA I.P. 2013. População residente (N.º) por Local de residência, Sexo e Grupo etário; Anual. 14/06/2013 ed.
- INSTITUTO NACIONAL DE ESTATÍSTICA I.P. 2014. Médicas/os especialistas (N.º) por Local de residência (NUTS - 2002), Sexo e Especialidade médica ; Anual. 09/01/2014 ed.
- IVERSEN, T. 1997. The effect of a private sector on the waiting time in a national health service. *Journal of Health Economics*, 16, 381-396.
- IVERSEN, T. & LURÅS, H. 2000. The effect of capitation on GPs' referral decisions. *Health Economics*, 9, 199-210.
- JAATINEN, P. T., AARNIO, P., REMES, J., HANNUKAINEN, J. & KÖYMÄRI-SEILONEN, T. 2002. Teleconsultation as a replacement for referral to an outpatient clinic. *Journal of telemedicine and telecare*, 8, 102-106.
- JARRETT, J., WOODCOCK, J., GRIFFITHS, U. K., CHALABI, Z., EDWARDS, P., ROBERTS, I. & HAINES, A. 2012. Effect of increasing active travel in urban England and Wales on costs to the National Health Service. *The Lancet*, 379, 2198-2205.
- JARVIS-SELINGER, S., CHAN, E., PAYNE, R., PLOHMAN, K. & HO, K. 2008. Clinical telehealth across the disciplines: lessons learned. *Telemedicine and e-Health*, 14, 720-725.
- JENNETT, P., HALL, L. A., HAILEY, D., OHINMAA, A., ANDERSON, C., THOMAS, R., YOUNG, B., LORENZETTI, D. & SCOTT, R. 2003. The socio-economic impact of telehealth: a systematic review. *Journal of telemedicine and telecare*, 9, 311-320.
- JOHNSON-LAIRD, P. N. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness*, Harvard University Press.
- JOHNSTON, K., KENNEDY, C., MURDOCH, I., TAYLOR, P. & COOK, C. 2004. The cost-effectiveness of technology transfer using telemedicine. *Health policy and planning*, 19, 302-309.
- JONES, A. P., HOMER, J. B., MURPHY, D. L., ESSIEN, J. D., MILSTEIN, B. & SEVILLE, D. A. 2006. Understanding diabetes population dynamics through simulation modeling and experimentation. *American Journal of Public Health*, 96, 488-494.
- JONES, R. 2000. Self Care: Important for health services and needing more research. *BMJ: British Medical Journal*, 320, 596.
- JUN, J., JACOBSON, S. & SWISHER, J. 1999. Application of discrete-event simulation in health care clinics: a survey. *Journal of the operational research society*, 50, 109-123.
- KARNON, J. 2003. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics*, 12, 837-848.
- KAY, A. 2002. The abolition of the GP fundholding scheme: a lesson in evidence-based policy making. *The British Journal of General Practice*, 52, 141.

- KERINS, M., MCKEE, G. & BENNETT, K. 2011. Contributing factors to patient non-attendance at and non-completion of Phase III cardiac rehabilitation. *European Journal of Cardiovascular Nursing*, 10, 31-36.
- KIM, M., BLENDON, R. J. & BENSON, J. M. 2001. How interested are Americans in new medical technologies? A multicountry comparison. *Health Affairs*, 20, 194-201.
- KINCHEN, K. S., COOPER, L. A., LEVINE, D., WANG, N. Y. & POWE, N. R. 2004. Referral of patients to specialists: factors affecting choice of specialist by primary care physicians. *The Annals of Family Medicine*, 2, 245-252.
- KING, M. T., HALL, J., LANCSAR, E., FIEBIG, D., HOSSAIN, I., LOUVIERE, J., REDDEL, H. K. & JENKINS, C. R. 2007. Patient preferences for managing asthma: results from a discrete choice experiment. *Health Economics*, 16, 703-717.
- KLEIJNEN, J. P. 1995. Verification and validation of simulation models. *European Journal of Operational Research*, 82, 145-162.
- KNAUF, F. & ARONSON, P. S. 2009. ESRD as a window into America's cost crisis in health care. *Journal of the American Society of Nephrology*, 20, 2093-2097.
- KNOL, A., VAN DEN AKKER, T. W., DAMSTRA, R. & DE HAAN, J. 2006. Teledermatology reduces the number of patient referrals to a dermatologist. *Journal of telemedicine and telecare*, 12, 75-78.
- KRASNIK, A., GROENEWEGEN, P. P., PEDERSEN, P. A., VON SCHOLTEN, P., MOONEY, G., GOTTSCHAU, A., FLIERMAN, H. A. & DAMSGAARD, M. T. 1990. Changing remuneration systems: effects on activity in general practice. *BMJ: British Medical Journal*, 300, 1698.
- KRUPINSKI, E., BARKER, G., RODRIGUEZ, G., ENGSTROM, M., LEVINE, N., LOPEZ, A. M. & WEINSTEIN, R. S. 2002. Telemedicine versus in-person dermatology referrals: an analysis of case complexity. *Telemedicine Journal and e-Health*, 8, 143-147.
- KUHFELD, W. F. 2009. Marketing research methods in SAS experimental design, choice, conjoint, and graphical techniques.
- KVEDAR, J. C., MENN, E. R., BARADAGUNTA, S., SMULDERS-MEYER, O. & GONZALEZ, E. 1999. Teledermatology in a capitated delivery system using distributed information architecture: design and development. *Telemedicine Journal*, 5, 357-366.
- LABELLE, R., STODDART, G. & RICE, T. 1994. A re-examination of the meaning and importance of supplier-induced demand. *Journal of health economics*, 13, 347-368.
- LACY, N. L., PAULMAN, A., REUTER, M. D. & LOVEJOY, B. 2004. Why we don't come: patient perceptions on no-shows. *The Annals of Family Medicine*, 2, 541-545.
- LAGARDE, M. & BLAAUW, D. 2009. A review of the application and contribution of discrete choice experiments to inform human resources policy interventions. *Hum Resour Health*, 7, 62.
- LAMMINEN, H., LAMMINEN, J., RUOHONEN, K. & UUSITALO, H. 2001. A cost study of teleconsultation for primary-care ophthalmology and dermatology. *Journal of Telemedicine and Telecare*, 7, 167-173.
- LAMMINEN, J., FORSVIK, H., VOPIO, V. & RUOHONEN, K. 2011. Teleconsultation: changes in technology and costs over a 12-year period. *Journal of telemedicine and telecare*, 17, 412-416.
- LANDRY, M., BANVILLE, C. & ORAL, M. 1996. Model legitimisation in operational research. *European Journal of Operational Research*, 92, 443-457.
- LANE, D. C. 1995. The folding star: a comparative reframing and extension of validity concepts in system dynamics.
- LANE, D. C. 1999. Social theory and system dynamics practice. *European Journal of Operational Research*, 113, 501-527.
- LANE, D. C. & HUSEMANN, E. 2007. System dynamics mapping of acute patient flows. *Journal of the operational research society*, 59, 213-224.
- LANE, D. C., MONEFELDT, C. & HUSEMANN, E. 2003. Client involvement in simulation model building: hints and insights from a case study in a London hospital. *Health Care Management Science*, 6, 105-116.

- LANE, D. C., MONEFELDT, C. & ROSENHEAD, J. 2000. Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *Journal of the operational research society*, 518-531.
- LASKOWSKI, M. & MUKHI, S. 2009. Agent-based simulation of emergency departments with patient diversion. *Electronic Healthcare*. Springer.
- LATTIMER, V., BRAILSFORD, S., TURNBULL, J., TARNARAS, P., SMITH, H., GEORGE, S., GERARD, K. & MASLIN-PROTHERO, S. 2004. Reviewing emergency care systems I: insights from system dynamics modelling. *Emergency Medicine Journal*, 21, 685-691.
- LEAVITT, F. 1979. The Health Belief Model and utilization of ambulatory care services. *Social Science & Medicine. Part A: Medical Psychology & Medical Sociology*, 13, 105-112.
- LEDLEY, R. S. & LUSTED, L. B. 1959. Reasoning foundations of medical diagnosis. *Science*, 130, 9-21.
- LEGGETT, P., GRAHAM, L., STEELE, K., GILLILAND, A., STEVENSON, M., O'REILLY, D., WOOTTON, R. & TAGGART, A. 2001. Telerheumatology--diagnostic accuracy and acceptability to patient, specialist, and general practitioner. *The British Journal of General Practice*, 51, 746.
- LEÓN, A., CÁCERES, C., FERNÁNDEZ, E., CHAUSA, P., MARTIN, M., CODINA, C., ROUSAUD, A., BLANCH, J., MALLOLAS, J. & MARTINEZ, E. 2011. A new multidisciplinary home care telemedicine system to monitor stable chronic human immunodeficiency virus-infected patients: a randomized study. *PloS one*, 6, e14515.
- LEWIS, D., TRANTER, G. & AXFORD, A. T. 2009. Use of videoconferencing in Wales to reduce carbon dioxide emissions, travel costs and time. *Journal of telemedicine and telecare*, 15, 137-138.
- LIEBERMAN, J. 2003. History of the use of antidepressants in primary care. *Prim Care Companion J Clin Psychiatry*, 5, 6-10.
- LIM, A. C., SEE, A. C. & SHUMACK, S. P. 2001. Progress in Australian teledermatology. *Journal of Telemedicine and Telecare*, 7, 55-58.
- LINDSAY, C. M. & FEIGENBAUM, B. 1984. Rationing by waiting lists. *American Economic Review*, 74, 404-417.
- LITTLE, J. D. & GRAVES, S. C. 2008. Little's law. *Building Intuition*. Springer.
- LITTLE, P., DORWARD, M., WARNER, G., STEPHENS, K., SENIOR, J. & MOORE, M. 2004. Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. *Bmj*, 328, 444.
- LOANE, M., BLOOMER, S., CORBETT, R., EEDY, D., EVANS, C., HICKS, N., JACKLIN, P., LOTERY, H., MATHEWS, C. & PAISLEY, J. 2001. A randomized controlled trial assessing the health economics of realtime teledermatology compared with conventional care: an urban versus rural perspective. *Journal of Telemedicine and Telecare*, 7, 108-118.
- LOANE, M., BLOOMER, S., CORBETT, R., EEDY, D., HICKS, N., LOTERY, H., MATHEWS, C., PAISLEY, J., STEELE, K. & WOOTTON, R. 2000a. A comparison of real-time and store-and-forward teledermatology: a cost-benefit study. *British Journal of Dermatology*, 143, 1241-1247.
- LOANE, M., BLOOMER, S., CORBETT, R., EEDY, D., HICKS, N., LOTERY, H., MATHEWS, C., PAISLEY, J., STEELE, K. & WOOTTON, R. 2000b. A randomized controlled trial to assess the clinical effectiveness of both realtime and store-and-forward teledermatology compared with conventional care. *Journal of Telemedicine and Telecare*, 6, 1-3.
- LOBLEY, D. 1997. The economics of telemedicine. *Journal of Telemedicine and Telecare*, 3, 117-125.
- LORD NORMAN WARNER. 2013. How to fix the bust NHS business model. Available from: <http://www.haciric.org/news/how-to-fix-the-bust-nhs-business-model> [2013].

- LOUVIERE, J. J., FLYNN, T. N. & CARSON, R. T. 2010. Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, 3, 57-72.
- LOUVIERE, J. J., HENSHER, D. A. & SWAIT, J. D. 2000. *Stated choice methods: analysis and applications*, Cambridge University Press.
- LYDEARD, S. & JONES, R. 1989. Factors affecting the decision to consult with dyspepsia: comparison of consultants and non-consulters. *The Journal of the Royal College of General Practitioners*, 39, 495.
- MACAL, C. M. & NORTH, M. J. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4, 151-162.
- MACFARLANE, A., MURPHY, A. W. & CLERKIN, P. 2006. Telemedicine services in the Republic of Ireland: an evolving policy context. *Health policy*, 76, 245-258.
- MACKENZIE, M., O'DONNELL, C., HALLIDAY, E., SRIDHARAN, S. & PLATT, S. 2010. Evaluating complex interventions: one size does not fit all. *BMJ. British medical journal*, 340, 401-403.
- MAIR, F., HAYCOX, A. & WILLIAMS, T. 2000. A review of telemedicine cost-effectiveness studies. *Journal of telemedicine and telecare*, 6, 38-40.
- MANTOVANI, L., MONZINI, M., MANNUCCI, P., SCALONE, L., VILLA, M. & GRINGERI, A. 2005. Differences between patients', physicians' and pharmacists' preferences for treatment products in haemophilia: a discrete choice experiment. *Haemophilia*, 11, 589-597.
- MARINKER, M. & SHAW, J. 2003. Not to be taken as directed: putting concordance for taking medicines into practice. *BMJ: British Medical Journal*, 326, 348.
- MARTIN, E., RUSSELL, D., GOODWIN, S., CHAPMAN, R., NORTH, M. & SHERIDAN, P. 1991. Why patients consult and what happens when they do. *BMJ: British Medical Journal*, 303, 289.
- MARTIN, S. & SMITH, P. C. 1999. Rationing by waiting lists: an empirical investigation. *Journal of Public Economics*, 71, 141-164.
- MARTIN, S. & SMITH, P. C. 2003. Using panel methods to model waiting times for National Health Service surgery. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 166, 369-387.
- MARTÍNEZ-GARCÍA, S., DEL BOZ-GONZÁLEZ, J., MARTÍN-GONZÁLEZ, T., SAMANIEGO-GONZÁLEZ, E. & CRESPO-ERCHIGA, V. 2007. Teledermatología. Revisión de 917 teleconsultas. *Actas dermo-sifilográficas*, 98, 318-324.
- MARTINEZ-MOYANO, I. J. 2012. Documentation for model transparency. *System Dynamics Review*, 28, 199-208.
- MASINO, C., RUBINSTEIN, E., LEM, L., PURDY, B. & ROSSOS, P. G. 2010. The impact of telemedicine on greenhouse gas emissions at an academic health science center in Canada. *Telemedicine and e-Health*, 16, 973-976.
- MASS, N. J. & SENGE, P. M. 1978. Alternative tests for the selection of model variables. *Systems, Man and Cybernetics, IEEE Transactions on*, 8, 450-460.
- MASSONE, C., SOYER, H., HOFMANN-WELLENHOF, R., DI STEFANI, A., LOZZI, G., GABLER, G., DONG, H., ARGENZIANO, G., OZDEMIR, F. & FINK-PUCHES, R. 2006. Two years' experience with Web-based teleconsulting in dermatology. *Journal of telemedicine and telecare*, 12, 83-87.
- MATHIEU, J. E., HEFFNER, T. S., GOODWIN, G. F., SALAS, E. & CANNON-BOWERS, J. A. 2000. The influence of shared mental models on team process and performance. *Journal of applied psychology*, 85, 273.
- MAY, C. 2006. A rational model for assessing and evaluating complex interventions in health care. *BMC Health Serv Res*, 6, 86.
- MCCULLOCH, C. E. 2000. Generalized linear models. *Journal of the American Statistical Association*, 95, 1320-1324.
- MCCULLOCH, P., TAYLOR, I., SASAKO, M., LOVETT, B. & GRIFFIN, D. 2002. Randomised trials in surgery: problems and possible solutions. *BMJ: British Medical Journal*, 324, 1448.

- MCFADDEN, D. 1980. Econometric models for probabilistic choice among products. *Journal of Business*, S13-S29.
- MCGUIRE, T. G. 2000. Physician agency. *Handbook of health economics*, 1, 461-536.
- MCKEEARGUE, M. 2010. Budget crises, health, and social welfare programmes. *Bmj*, 341, 77.
- MCKINLAY, J. B. 1972. Some approaches and problems in the study of the use of services--an overview. *Journal of health and social behavior*, 115-152.
- MCMURRAN, M., HUBAND, N. & OVERTON, E. 2010. Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, 30, 277-287.
- MCQUEEN, D. & ANDERSON, L. 1999. What counts as evidence? Issues and debates on evidence relevant to the evaluation of community health promotion programs. *Evaluation in health promotion: principles and perspectives*. Copenhagen, Denmark. World Health Organization (Euro).
- MEADOWS, D. H. 1980. The unavoidable a priori. *Elements of the system dynamics method*, 23-57.
- MEHROTRA, A., FORREST, C. B. & LIN, C. Y. 2011. Dropping the baton: specialty referrals in the United States. *Milbank Quarterly*, 89, 39-68.
- MERRELL, R. C. & DOARN, C. R. 2009. Telemedicine is green!! *Telemedicine and e-Health*, 15, 731-732.
- MIELONEN, M.-L., OHINMAA, A., MORING, J. & ISOHANNI, M. 2000. Psychiatric inpatient care planning via telemedicine. *Journal of Telemedicine and Telecare*, 6, 152-157.
- MISTRY, H. 2012. Systematic review of studies of the cost-effectiveness of telemedicine and telecare. Changes in the economic evidence over twenty years. *Journal of telemedicine and telecare*, 18, 1-6.
- MOFID, M., NESBITT, T. & KNUTTEL, R. 2007. The other side of teledermatology: patient preferences. *Journal of Telemedicine and Telecare*, 13, 246-250.
- MOORE, C. G., WILSON-WITHERSPOON, P. & PROBST, J. C. 2001. Time and money: effects of no-shows at a family practice residency clinic. *FAMILY MEDICINE-KANSAS CITY*, 33, 522-527.
- MORECROFT, J. 2007. *Strategic modelling and business dynamics: A feedback systems approach*, Wiley. com.
- MORECROFT, J. & ROBINSON, S. Explaining puzzling dynamics: comparing the use of system dynamics and discrete-event simulation. Proceedings of the 23rd International Conference of the System Dynamics Society, 2005. 17-21.
- MORENO-RAMIREZ, D., FERRANDIZ, L., BERNAL, A. P., DURAN, R. C., MARTÍN, J. R. & CAMACHO, F. 2005. Teledermatology as a filtering system in pigmented lesion clinics. *Journal of telemedicine and telecare*, 11, 298-303.
- MORGAN, M., JENKINS, L. & RIDSDALE, L. 2007. Patient pressure for referral for headache: a qualitative study of GPs' referral behaviour. *The British Journal of General Practice*, 57, 29.
- MOSER, P. L., HAUFFE, H., LORENZ, I. H., HAGER, M., TIEFENTHALER, W., LORENZ, H. M., MIKUZ, G., SOEGNER, P. & KOLBITSCH, C. 2004. Publication output in telemedicine during the period January 1964 to July 2003. *Journal of telemedicine and telecare*, 10, 72-77.
- MUNDLAK, Y. 1978. On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69-85.
- NAIM, M. & TOWILL, D. 1993. Modelling and forecasting industrial innovations via the transfer function s-shaped learning curve. *The International Journal of Advanced Manufacturing Technology*, 8, 329-343.
- NATÁRIO, A. & AMARAL, J. C. 2011. Actuais e futuras necessidades previsionais de médicos (SNS). In: ADMINISTRAÇÃO CENTRAL DO SISTEMA DE SAÚDE I.P. (ed.).

- NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE. 2012. *Psoriasis: The assessment and management of psoriasis* [Online]. Available: <http://publications.nice.org.uk/psoriasis-cg153> [Accessed 5 September 2013].
- NEWTON, J., HAYES, V. & HUTCHINSON, A. 1991. Factors influencing general practitioners' referral decisions. *Family Practice*, 8, 308-313.
- NHS IMPROVING QUALITY 2013. Economic Evaluation of the Electronic Palliative Care Coordination System (EPaCCS) Early Implementer Sites.
- NOONE, A., GOLDACRE, M., COULTER, A. & SEAGROATT, V. 1989. Do referral rates vary widely between practices and does supply of services affect demand? A study in Milton Keynes and the Oxford region. *The Journal of the Royal College of General Practitioners*, 39, 404.
- NORDAL, E., MOSENG, D., KVAMMEN, B. & LØCHEN, M. 2001. A comparative study of teleconsultations versus face-to-face consultations. *Journal of Telemedicine and Telecare*, 7, 257-265.
- NORTHROP, R. B. 2010. *Introduction to complexity and complex systems*, CRC Press.
- O'DONNELL, C. A. 2000. Variation in GP referral rates: what can we learn from the literature? *Family Practice*, 17, 462-471.
- O'HARA, R. B. & KOTZE, D. J. 2010. Do not log-transform count data. *Methods in Ecology and Evolution*, 1, 118-122.
- OAKLEY, A., KERR, P., DUFFILL, M., RADEMAKER, M., FLEISCHL, P. & BRADFORD, N. 2000. Patient cost-benefits of realtime teledermatology-a comparison of data from Northern Ireland and New Zealand. *Journal of Telemedicine and Telecare*, 6, 97-101.
- OECD 2013a. Health at a Glance 2013: OECD Indicators. Organisation for Economic Co-operation and Development,.
- OECD. 2013b. *Major brake in health spending growth as governments cut budgets in the crisis, says OECD* [Online]. Available: <http://www.oecd.org/newsroom/major-brake-in-health-spending-growth-as-governments-cut-budgets-in-the-crisis.htm> [Accessed 28/12/2013 2013].
- OLDRIDGE, N. B. 1996. Outcomes measurement: health state preferences and economic evaluation. *Assistive Technology*, 8, 94-102.
- OSGOOD, D. W. 2000. Poisson-based regression analysis of aggregate crime rates. *Journal of quantitative criminology*, 16, 21-43.
- PAWSON, R., GREENHALGH, T., HARVEY, G. & WALSHE, K. 2005. Realist review--a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*, 10 Suppl 1, 21-34.
- PENCHEON, D. 2009. Health services and climate change: what can be done? *J Health Serv Res Policy*, 14, 2-4.
- PIDD, M. 1992. *Computer simulation in management science*, John Wiley & Sons, Inc.
- PLSEK, P. E. & GREENHALGH, T. 2001. Complexity science: The challenge of complexity in health care. *BMJ: British Medical Journal*, 323, 625.
- PLSEK, P. E. & WILSON, T. 2001. Complexity science: complexity, leadership, and management in healthcare organisations. *BMJ: British Medical Journal*, 323, 746.
- PRIMARY CARE DERMATOLOGY SOCIETY. 2012. *Atypical (dysplastic) melanocytic naevi* [Online]. Available: <http://www.pcids.org.uk/clinical-guidance/atypical-dysplastic-melanocytic-naevus> [Accessed 5 September 2013].
- PROPPER, C. 1995. The disutility of time spent on the United Kingdom's National Health Service waiting lists. *Journal of Human Resources*, 677-700.
- RADZICKI, M. J. 1990. Methodologia oeconomiae et systematis dynamis. *System Dynamics Review*, 6, 123-147.
- RESNICK, M. 1994. *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds*, Mit Press.
- RIDSDALE, L., CLARK, L. V., DOWSON, A. J., GOLDSTEIN, L. H., JENKINS, L., MCCRONE, P., MORGAN, M. & SEED, P. T. 2007. How do patients referred to neurologists for headache differ from those managed in primary care? *The British Journal of General Practice*, 57, 388.

- RIEBER, L. P. 1992. Computer-based microworlds: A bridge between constructivism and direct instruction. *Educational technology research and development*, 40, 93-106.
- RINGEL, J. S., HOSEK, S. D., VOLLAARD, B. A. & MAHNOVSKI, S. 2002. The Elasticity of Demand for Health Care. A review of the literature and its application to the military health system. DTIC Document.
- ROBERTS, A., MARSHALL, L. & CHARLESWORTH, A. 2012. A Decade of Austerity? The funding pressures facing the NHS from 2010/11 to 2021/22. *London: Nuffield Trust*.
- ROBERTS, C. & DANGERFIELD, B. 1990. Modelling the epidemiological consequences of HIV infection and AIDS: a contribution from operational research. *Journal of the operational research society*, 273-289.
- ROSENBAUM, L. & LAMAS, D. 2012. Cents and sensitivity—teaching physicians to think about costs. *New England Journal of Medicine*, 367, 99-101.
- ROSENTHAL, M. B. 2008. Beyond pay for performance—emerging models of provider-payment reform. *New England Journal of Medicine*, 359, 1197-1200.
- RUBIN, G., BATE, A., GEORGE, A., SHACKLEY, P. & HALL, N. 2006. Preferences for access to the GP: a discrete choice experiment. *The British Journal of General Practice*, 56, 743.
- RYAN, M. 1999. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social science & medicine*, 48, 535-546.
- RYAN, M. & GERARD, K. 2003. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied health economics and health policy*, 2, 55-64.
- RYAN, M., GERARD, K. & AMAYA-AMAYA, M. 2008. *Using discrete choice experiments to value health and health care*, Springer.
- RYAN, M. & SKÅTUN, D. 2004. Modelling non-demanders in choice experiments. *Health Economics*, 13, 397-402.
- SABATÉ, E. 2003. *Adherence to long-term therapies: evidence for action*, World Health Organization.
- SANTOS, S. P., BELTON, V. & HOWICK, S. 2002. Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations & Production Management*, 22, 1246-1272.
- SAPSED, J. & SALTER, A. 2004. Postcards from the edge: local communities, global programs and boundary objects. *Organization studies*, 25, 1515-1534.
- SCALVINI, S., TRIDICO, C., GLISENTI, F., GIORDANO, A., PIRINI, S., PEDUZZI, P. & AUXILIA, F. 2009. The SUMMA Project: a feasibility study on telemedicine in selected Italian areas. *Telemedicine and e-Health*, 15, 261-269.
- SCHIERITZ, N. & GROBLER, A. Emergent structures in supply chains—a study integrating agent-based and system dynamics modeling. *System Sciences*, 2003. Proceedings of the 36th Annual Hawaii International Conference on, 2003. IEEE, 9 pp.
- SCHMALZRIED, H. D. & LISZAK, J. 2012. A model program to reduce patient failure to keep scheduled medical appointments. *Journal of community health*, 37, 715-718.
- SCOTT, A. 2000. Economics of general practice. *Handbook of health economics*, 1, 1175-1200.
- SCULPHER, M. J., PANG, F. S., MANCA, A., DRUMMOND, M. F., GOLDER, S., URDAHL, H., DAVIES, L. M. & EASTWOOD, A. 2004. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess*, 8, iii-iv, 1-192.
- SEMPERE-SELVA, T., PEIRÓ, S., SENDRA-PINA, P., MARTÍNEZ-ESPÍN, C. & LÓPEZ-AGUILERA, I. 2001. Inappropriate use of an accident and emergency department: magnitude, associated factors, and reasons—an approach with explicit criteria. *Annals of emergency medicine*, 37, 568-579.
- SHANNON, G. W. & BUKER, C. M. 2010. Determining accessibility to dermatologists and teler dermatology locations in Kentucky: demonstration of an innovative geographic information systems approach. *Telemedicine and e-Health*, 16, 670-677.

- SHAPIRO, M., JAMES, W. D., KESSLER, R., LAZORIK, F. C., KATZ, K. A., TAM, J., NIEVES, D. S. & MILLER, J. J. 2004. Comparison of skin biopsy triage decisions in 49 patients with pigmented lesions and skin neoplasms: store-and-forward teledermatology vs face-to-face dermatology. *Archives of dermatology*, 140, 525.
- SHEPPERD, S., LEWIN, S., STRAUS, S., CLARKE, M., ECCLES, M. P., FITZPATRICK, R., WONG, G. & SHEIKH, A. 2009. Can we systematically review studies that evaluate complex interventions? *PLoS Medicine*, 6, e1000086.
- SHIEH, F. K., SNYDER, G., ROBERT HORSBURGH, C., BERNARDO, J., MURPHY, C. & SAUKKONEN, J. J. 2006. Predicting non-completion of treatment for latent tuberculous infection: a prospective survey. *American journal of respiratory and critical care medicine*, 174, 717-721.
- SICILIANI, L. & HURST, J. 2005. Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries. *Health policy*, 72, 201-215.
- SICOTTE, C., LEHOUX, P., VAN DOESBURG, N., CARDINAL, G. & LEBLANC, Y. 2004. A cost-effectiveness analysis of interactive paediatric telecardiology. *Journal of telemedicine and telecare*, 10, 78-83.
- SIEBERS, P.-O., MACAL, C. M., GARNETT, J., BUXTON, D. & PIDD, M. 2010. Discrete-event simulation is dead, long live agent-based simulation! *Journal of Simulation*, 4, 204-210.
- SINGH, S. 1988. Self referral to accident and emergency department: patients' perceptions. *BMJ: British Medical Journal*, 297, 1179.
- SKINNER, J. S., STAIGER, D. O. & FISHER, E. S. 2006. Is technological change in medicine always worth it? The case of acute myocardial infarction. *Health Affairs*, 25, w34-w47.
- SMITH, P. C. & VAN ACKERE, A. 2002. A note on the integration of system dynamics and economic models. *Journal of Economic Dynamics and Control*, 26, 1-10.
- SOLBERG, K. E. 2008. Telemedicine set to grow in India over the next 5 years. *The Lancet*, 371, 17-18.
- STAINSBY, H., TABOADA, M. & LUQUE, E. Towards an agent-based simulation of hospital emergency departments. Services Computing, 2009. SCC'09. IEEE International Conference on, 2009. IEEE, 536-539.
- STAR, S. L. & GRIESEMER, J. R. 1989. Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science*, 19, 387-420.
- STARFIELD, B., FORREST, C. B., NUTTING, P. A. & VON SCHRADER, S. 2002. Variability in physician referral decisions. *The Journal of the American Board of Family Practice*, 15, 473-480.
- STENSLAND, J., SPEEDIE, S. M., IDEKER, M., HOUSE, J. & THOMPSON, T. 1999. The relative cost of outpatient telemedicine services. *Telemedicine Journal*, 5, 245-256.
- STERMAN, J. 2000. *Business dynamics*, Irwin-McGraw-Hill.
- STERMAN, J. D. 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica*, 10, 51-66.
- STERMAN, J. D. 1994. Learning in and about complex systems. *System Dynamics Review*, 10, 291-330.
- STERMAN, J. D. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review*, 18, 501-531.
- STRECHER, V. J. & ROSENSTOCK, I. M. 1997. The health belief model. *Cambridge handbook of psychology, health and medicine*, 113-117.
- STUCKLER, D., BASU, S., SUHRCKE, M., COUTTS, A. & MCKEE, M. 2009. The public health effect of economic crises and alternative policy responses in Europe: an empirical analysis. *The Lancet*, 374, 315-323.
- SULLIVAN, C. O., OMAR, R. Z., AMBLER, G. & MAJEED, A. 2005. Case-mix and variation in specialist referrals in general practice. *The British Journal of General Practice*, 55, 529.

- SVEIBY, K., GRIPENBERG, P., SEGERCRANTZ, B., ERIKSSON, A. & AMINOFF, A. Unintended and undesirable consequences of innovation. XX ISPIM Conference, The Future of Innovation. Vienna, 2009.
- TAYLOR, P. 2005. Evaluating telemedicine systems and services. *Journal of telemedicine and telecare*, 11, 167-177.
- TAYLOR, P., GOLDSMITH, P., MURRAY, K., HARRIS, D. & BARKLEY, A. 2001. Evaluating a telemedicine system to assist in the management of dermatology referrals. *British Journal of Dermatology*, 144, 328-333.
- THE BRITISH ASSOCIATION OF DERMATOLOGISTS. 2008a. *Melanoma* [Online]. Available: <http://www.bad.org.uk/site/842/Default.aspx> [Accessed 5 September 2013].
- THE BRITISH ASSOCIATION OF DERMATOLOGISTS. 2008b. *Seborrhoeic Keratoses* [Online]. Available: <http://www.bad.org.uk/site/873/default.aspx> [Accessed 5 September 2013].
- THE KING'S FUND 2013. How is the health and social care system performing? Quarterly Monitoring Report.
- THORNTON, J. 2002. Estimating a health production function for the US: some new evidence. *Applied Economics*, 34, 59-62.
- TRIVEDI, P. K. 2010. Recent developments in panel model for count data. *2010 Mexican Stata Users Group Meeting*.
- TROUILLER, P., OLLIARO, P., TORREELE, E., ORBINSKI, J., LAING, R. & FORD, N. 2002. Drug development for neglected diseases: a deficient market and a public-health policy failure. *The Lancet*, 359, 2188-2194.
- TSILIMIGAKI, A., MARAKA, S., TSEKOURA, T., AGELAKOU, V., VEKIOU, A., PAPHITIS, C. & THANOPOULOS, V. 2001. Eighteen months' experience with remote diagnosis, management and education in congenital heart disease. *Journal of telemedicine and telecare*, 7, 239-243.
- TURNER, D., TARRANT, C., WINDRIDGE, K., BRYAN, S., BOULTON, M., FREEMAN, G. & BAKER, R. 2007. Do patients value continuity of care in general practice? An investigation using stated preference discrete choice experiments. *Journal of Health Services Research & Policy*, 12, 132-137.
- UBACH, C., SCOTT, A., FRENCH, F., AWRAMENKO, M. & NEEDHAM, G. 2003. What do hospital consultants value about their jobs? A discrete choice experiment. *BMJ*, 326, 1432.
- UHLENHAKE, E., BRODELL, R. & MOSTOW, E. 2009. The dermatology work force: A focus on urban versus rural wait times. *Journal of the American Academy of Dermatology*, 61, 17-22.
- VAN ACKERE, A. & SMITH, P. C. 1999. Towards a macro model of National Health Service waiting lists. *System Dynamics Review*, 15, 225-252.
- VAN DE WETERING, G., WOERTMAN, W. H. & ADANG, E. M. 2012. A model to correct for short-run inefficiencies in economic evaluations in healthcare. *Health economics*, 21, 270-281.
- VAN SCHOOR, N. M., DEVILLE, W., BOUTER, L. & LIPS, P. 2002. Acceptance and compliance with external hip protectors: a systematic review of the literature. *Osteoporosis International*, 13, 917-924.
- VERMEIRE, E., HEARNshaw, H., VAN ROYEN, P. & DENEKENS, J. 2001. Patient adherence to treatment: three decades of research. A comprehensive review. *Journal of clinical pharmacy and therapeutics*, 26, 331-342.
- VOSNIADOU, S. & BREWER, W. F. 1994. Mental models of the day/night cycle. *Cognitive science*, 18, 123-183.
- WADE, V. A., KARNON, J., ELSHAUG, A. G. & HILLER, J. E. 2010. A systematic review of economic analyses of telehealth services using real time video communication. *BMC Health Services Research*, 10, 233.
- WAKELAND, W. W., GALLAHER, E. J., MACOVSKY, L. M. & AKTIPIS, C. A. A comparison of system dynamics and agent-based simulation applied to the study of cellular

- receptor dynamics. *System Sciences*, 2004. Proceedings of the 37th Annual Hawaii International Conference on, 2004. IEEE, 10 pp.
- WALACH, H., FALKENBERG, T., FONNEBO, V., LEWITH, G. & JONAS, W. B. 2006. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol*, 6, 29.
- WALLACE, P., HAINES, A., HARRISON, R., BARBOUR, J., THOMPSON, S., JACKLIN, P., ROBERTS, J., LEWIS, L. & WAINWRIGHT, P. 2002. Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. *The lancet*, 359, 1961-1968.
- WEBB, S. & LLOYD, M. 1994. Prescribing and referral in general practice: a study of patients' expectations and doctors' actions. *The British Journal of General Practice*, 44, 165.
- WEINSTEIN, M. C., TOY, E. L., SANDBERG, E. A., NEUMANN, P. J., EVANS, J. S., KUNTZ, K. M., GRAHAM, J. D. & HAMMITT, J. K. 2001. Modeling for health care and other policy decisions: uses, roles, and validity. *Value in Health*, 4, 348-361.
- WHITED, J. D. 2006. Teledermatology research review. *International journal of dermatology*, 45, 220-229.
- WHITED, J. D., DATTA, S., HALL, R. P., FOY, M. E., MARBREY, L. E., GRAMBOW, S. C., DUDLEY, T. K., SIMEL, D. L. & ODDONE, E. Z. 2003. An economic analysis of a store and forward teledermatology consult system. *Telemedicine Journal and E-health*, 9, 351-360.
- WHITED, J. D., HALL, R. P., FOY, M. E., MARBREY, L. E., GRAMBOW, S. C., DUDLEY, T. K., DATTA, S., SIMEL, D. L. & ODDONE, E. Z. 2002. Teledermatology's impact on time to intervention among referrals to a dermatology consult service. *Telemedicine Journal and E-Health*, 8, 313-321.
- WHITTEN, P., JOHANNESSEN, L. K., SOERENSEN, T., GAMMON, D. & MACKERT, M. 2007. A systematic review of research methodology in telemedicine studies. *Journal of telemedicine and telecare*, 13, 230-235.
- WHITTEN, P. S., MAIR, F. S., HAYCOX, A., MAY, C. R., WILLIAMS, T. L. & HELLMICH, S. 2002. Systematic review of cost effectiveness studies of telemedicine interventions. *BMJ: British Medical Journal*, 324, 1434.
- WILKIN, D. & SMITH, A. 1987. Explaining variation in general practitioner referrals to hospital. *Family Practice*, 4, 160-169.
- WILSON, T. & HOLT, T. 2001. Complexity science: complexity and clinical care. *BMJ: British Medical Journal*, 323, 685.
- WINDMEIJER, F., GRAVELLE, H. & HOONHOUT, P. 2005. Waiting lists, waiting times and admissions: an empirical analysis at hospital and general practice level. *Health Economics*, 14, 971-985.
- WISE, P. 2013. IMF warns Portugal of tough years ahead. *Financial Times*, 19/12/2013.
- WOLSTENHOLME, E. 1999. A patient flow perspective of UK Health Services: exploring the case for new "intermediate care" initiatives. *System Dynamics Review*, 15, 253-271.
- WOLSTENHOLME, E. F. 1993. A case study in community care using systems thinking. *Journal of the operational research society*, 925-934.
- WOOLDRIDGE, J. M. 1997. Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory*, 13, 667-678.
- WOOTTON, R., BLOOMER, S., CORBETT, R., EEDY, D., HICKS, N., LOTERY, H., MATHEWS, C., PAISLEY, J., STEELE, K. & LOANE, M. 2000. Multicentre randomised control trial comparing real time teledermatology with conventional outpatient dermatological care: societal cost-benefit analysis. *Bmj*, 320, 1252-1256.
- YIP, W. C. 1998. Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. *Journal of Health Economics*, 17, 675-699.
- YOGESAN, K., KUMAR, S., TAY-KEARNEY, M.-L. & CONSTABLE, I. 2006. Evaluation of Internet-based eye care delivery. *Journal of telemedicine and telecare*, 12, 103-104.

- ZAGONEL, A. A. Model conceptualization in group model building: A review of the literature exploring the tension between representing reality and negotiating a social order. Proceedings of the 20th International System Dynamics Conference, 2002.
- ZUCKERMAN, K. E., CAI, X., PERRIN, J. M. & DONELAN, K. 2011a. Incomplete specialty referral among children in community health centers. *The Journal of pediatrics*, 158, 24-30.
- ZUCKERMAN, K. E., NELSON, K., BRYANT, T. K., HOBRECKER, K., PERRIN, J. M. & DONELAN, K. 2011b. Specialty referral communication and completion in the community health center setting. *Academic Pediatrics*, 11, 288-296.
- ZWEIFEL, P. & MANNING, W. G. 2000. Moral hazard and consumer incentives in health care. *Handbook of health economics*, 1, 409-459.

Appendix A

INTERVIEW SCHEDULE AND GUIDE

As mentioned in Section 5.1.1, I conducted twelve semi-structured exploratory interviews in Alentejo resulting in a total of 8 hours and 50 minutes of recordings, which I transcribed and analysed. The study was limited to teleconsultations performed in 2010 involving consultants from the region's main hospital in Évora (65 percent of all teleconsultations in 2010 in Alentejo were performed by specialists from this hospital). Referring units included seven primary care practices and two hospitals across three primary care sub-regions. Potential participants included physicians and managers who had been involved with teleconsultations during 2010: twelve consultants, nine referring physicians (seven general practitioners and two consultants), and the managing directors of the three primary care sub-regions.

I contacted each of the twenty four candidates by telephone to arrange a date and place for the interview. Twelve candidates were not included in the study as I was either unable to reach them, or unable to set a date for the interview within the period of time allocated to the study. I interviewed the remaining twelve candidates, who were responsible for the majority of teleconsultations performed in the region: six consultants from dermatology, cardiology, psychiatry, physical and rehabilitation medicine, neurology and

gastroenterology (participating in 66 percent of the hospital's teleconsultations in all disciplines), three general practitioners and one referring consultant (85 percent of teleconsultations), and two managing directors of primary care sub-regions (78 percent of teleconsultations involved practices from the sub-regions under their management).

Questions covered a broad range of topics around the use of telemedicine in the region, including both closed-response (e.g. when did you start performing teleconsultations?) and open-ended questions (e.g. why did you start performing teleconsultations?). Participants were first asked about a regular workday, the activities involved and the difficulties they faced. This was followed by questions on the use of teleconsultations: motivation for using teleconsultations; the main differences relative to a traditional face-to-face appointment; the reasons for subsequent face-to-face appointments following teleconsultations, and how frequently these occur; differences in prescribing between teleconsultations and face-to-face appointments; waiting time for a teleconsultation and what affects it; percentage of teleconsultations relative to all appointments; time dedicated to teleconsultations and impact on workload; and the perceived advantages and drawbacks from the interviewee's perspective and from that of the patients. Other topics included how teleconsultations are reimbursed, and whether there are financial incentives to perform teleconsultations.

The interviews took place at the participants' places of work in November 2010. Copies of the interview scripts, information sheets and consent forms (all in Portuguese) are available below (Figures A-1 through A-6). The interviews took between 30 and 60 minutes. All participants gave their consent to participate. Questions covered a broad range of topics, including both closed-response and open-ended questions. All interviews were recorded and transcribed.

FOLHETO DE INFORMAÇÃO PARA PARTICIPANTES

Título do projecto: Telemedicina no Alentejo

Nome do Investigador principal: Tiago Cravo Oliveira

Nomes dos orientadores: Prof. James Barlow e Dr. Steffen Bayer

O seu nome foi indicado pelo Coordenador do Programa de Telemedicina do Alentejo Dr. Luis Gonçalves como uma referência para este estudo de investigação, o qual procura compreender melhor os efeitos do uso de telemedicina na prestação de cuidados de saúde. Numa altura em que estas tecnologias estão a ser cada vez mais utilizadas, há ainda muitas dúvidas em relação aos custos, benefícios e impactos para o sistema de saúde em geral. Com este estudo procuramos perceber melhor de que forma a telemedicina altera o planeamento, financiamento e a prestação de cuidados de saúde.

Convidamo-lo(a) a participar neste estudo, o qual tem a aprovação da Administração Regional de Saúde do Alentejo. Antes de tomar uma decisão, por favor leia a informação que se segue e não hesite em contactar o investigador principal ou os seus orientadores se tiver alguma questão.

Qual o objectivo deste estudo?

Este estudo procura investigar de que forma o uso de telemedicina alterou o planeamento, financiamento e prestação de cuidados de saúde no Alentejo, e de que forma poderia ser adoptado a nível nacional e com que impacto.

Porque fui escolhido(a) para participar?

Estamos nesta fase a entrevistar médicos de família e médicos especialistas que estejam ou tenham estado directamente relacionados com serviços de telemedicina na região do Alentejo. O seu nome foi-nos indicado pelo Dr. Luis Gonçalves como uma referência.

Sou obrigado(a) a participar?

Não. A decisão é sua. Se decidir participar, pode no futuro suspender a sua participação a qualquer momento sem que seja necessária uma justificação. Caso decida colaborar por favor guarde este folheto de informação para futura referência.

Figure A-1 Information sheet (page 1 of 2)

O que é esperado de mim?

O estudo será baseado em entrevistas semi-estruturadas entre os participantes e o investigador principal. As entrevistas terão lugar no local de trabalho dos participantes ou qualquer outro local que lhes seja conveniente. As entrevistas terão uma duração inferior a uma hora podendo ser gravadas para futura transcrição e análise. Caso não queira que as entrevistas sejam gravadas, por favor comunique essa decisão ao investigador principal.

A minha participação neste estudo é confidencial?

Toda e qualquer informação que advir da sua participação é anónima e confidencial. A informação será guardada num computador do Imperial College Business School, em Londres, protegido por password e identificada através de números codificados específicos ao estudo. O seu nome não será guardado com a informação. Nenhuma informação relativa a si será entregue a outra pessoa. Qualquer publicação que resulte deste estudo mencionará apenas informação relativa ao grupo de participantes, nunca mencionando dados relativos a qualquer participante em específico. Quando o estudo estiver completo, as transcrições e gravações serão guardadas por um período adequado para garantir que os resultados do estudo possam ser comunicados e validados.

Quais são os potenciais benefícios de participar?

A sua participação será essencial para entender melhor de que forma a telemedicina altera a prestação de cuidados de saúde e de que forma se poderá tirar o melhor partido destas tecnologias no futuro.

O que acontecerá aos resultados do estudo?

Os resultados do estudo serão comunicados na tese de doutoramento do investigador principal e em publicações científicas. Os resultados serão também comunicados ao Dr. Luis Gonçalves e poderão ser entregues aos participantes, caso estes o desejem.

Quem está a organizar e a financiar este estudo?

Este estudo faz parte do doutoramento do investigador principal, em colaboração com o Prof. James Barlow e com o Dr. Steffen Bayer do Imperial College Business School. O financiamento para este estudo é garantido pela Fundação para a Ciência e Tecnologia, através de uma bolsa de doutoramento.

Há alguém responsável por revêr a qualidade do estudo?

Sim. Este estudo está sujeito a revisões anuais por parte do comité de doutoramento, cujo director é o Prof. Erkki Autio. Este estudo tem ainda a aprovação da Administração Regional de Saúde do Alentejo através do Coordenador do Programa de Telemedicina do Alentejo.

Obrigado por ter lido este folheto de informação. Se tiver dúvidas ou questões, por favor, não hesite em contactar o investigador principal ou um dos seus orientadores:

Investigador principal

Tiago Cravo Oliveira MSc

tiago.oliveira@imperial.ac.uk

+44 (0)7807 347 839

+351 96 484 34 24

Orientador principal

Prof. James Barlow

j.barlow@imperial.ac.uk

Co-orientador

Dr. Steffen Bayer

s.bayer@imperial.ac.uk

Versão 1.0 (11 de Novembro de 2010)

Imperial College of Science, Technology and Medicine



Figure A-2 Information sheet (page 2 of 2)

FORMULÁRIO DE CONSENTIMENTO INFORMADO

Título do projecto: Telemedicina no Alentejo

Nome do Investigador principal: Tiago Cravo Oliveira

Nomes dos orientadores: Prof. James Barlow e Dr. Steffen Bayer

Por favor marque com uma cruz

1. Confirmando que li e compreendo o folheto de informação relativo ao estudo acima indicado e que tive oportunidade de fazer perguntas, as quais foram integralmente respondidas. ☐
2. Compreendo que a minha participação é voluntária e que tenho liberdade de suspende-la a qualquer momento, sem necessidade de justificação. ☐
3. Compreendo que as entrevistas nas quais participarei poderão ser gravadas e transcritas, e que será assegurada total confidencialidade e anonimidade. ☐
4. Aceito participar no estudo acima indicado. ☐

Nome do participante

Assinatura

Data

Tiago Cravo Oliveira

Nome do Investigador principal

Assinatura

Data

1 cópia para o participante; 1 cópia para o Investigador principal

Figure A-3 Consent form (page 1 of 1)

GUIÃO PARA ENTREVISTAS

Título do projecto: Telemedicina no Alentejo

Nome do Investigador principal: Tiago Cravo Oliveira

Nomes dos orientadores: Prof. James Barlow e Dr. Steffen Bayer

Antes de mais, muito obrigado por aceitar participar neste projecto de investigação. Deixe-me começar por me apresentar e falar-lhe um pouco sobre este projecto de investigação.

O meu nome é Tiago Cravo Oliveira e sou aluno de doutoramento no Imperial College Business School em Londres. Com este projecto de investigação procuramos perceber melhor de que forma a telemedicina altera o planeamento, financiamento e a prestação de cuidados de saúde no contexto do Alentejo.

A entrevista não deverá durar mais de uma hora e com a sua permissão será gravada. Desta forma será mais fácil recordar aquilo que foi dito nesta entrevista. No entanto, quero assegurar-lhe que toda a informação obtida através desta entrevista será confidencial e quaisquer nomes ou características identificadoras serão alteradas ou omitidas para assegurar anonimidade.

Antes de começarmos, tem alguma questão? Ok. Vou começar por fazer-lhe algumas questões sobre o seu trabalho e como este está ligado à telemedicina. De seguida, perguntar-lhe de que forma a telemedicina alterou a forma como trabalhava e quais as suas reacções em relação a essa mudança.

Figure A-4 Interview script (page 1 of 3)

PERGUNTAS SOBRE TRABALHO

1. Pode-me falar um pouco sobre o seu trabalho?

Qual a sua especialidade?

Que tipo de actividades?

2. Pode-me falar um pouco sobre como começou a utilizar telemedicina?

Por que motivo começou?

3. Quanto tempo por semana dedica a actividades relacionadas com telemedicina?

Que tipo de actividades estamos a falar?

O que fazia nesse espaço de tempo antes da telemedicina?

Considera que trabalha mais ou menos desde a introdução da telemedicina?

PERGUNTAS SOBRE CONSULTAS

4. Pode-me falar um pouco sobre as diferenças entre uma consulta normal e uma teleconsulta?

Quanto tempo dura uma e outra?

Fale-me um pouco sobre o processo tanto numa como noutra?

O que muda na forma como diagnostica?

Sente o mesmo nível de confiança numa e noutra?

5. Com que frequência necessita de uma consulta presencial após uma teleconsulta?

Quais as razões para tal?

6. Quais os factores que tem em conta quando referencia ou prescreve tratamentos?

Tem em conta as preferências do paciente? De que forma?

Alguma vez um paciente lhe disse abertamente quais as suas preferências?

Alguma vez um paciente o pressionou para referenciar/prescrever? Como reagiu?

7. Existe alguma lista de espera para consultas na sua especialidade?

Qual é o tempo de espera nessa lista?

De que forma é que esta lista de espera afecta a sua referenciação?

Como encaram os pacientes esta lista de espera?

Como encaixam na lista de espera as consultas presenciais após teleconsultas?

8. Existe alguma lista de espera para tratamentos na sua especialidade?

Qual é o tempo de espera nessa lista?

De que forma é que esta lista de espera afecta a sua prescrição?

Como encaram os pacientes esta lista de espera?

Como encaixam na lista de espera os tratamentos prescritos em teleconsultas?

9. Que percentagem das suas consultas poderia ser substituídas por teleconsultas?

O que mudaria para si caso isso acontecesse?

Considera que isso seria vantajoso ou não? Porque?

Figure A-5 Interview script (page 2 of 3)

PERGUNTAS SOBRE RESPONSABILIDADE

10. Quem e que assina as prescricoes?

O medico especialista ou o medico de familia?

11. Quem e que e medicamente responsavel pelo doente?

O medico especialista ou o medico de familia?

PERGUNTAS SOBRE VANTAGENS/DESVANTAGENS

12. Quais sao as vantagens para si?

Que tipo de beneficios ve no uso de telemedicina para si como medico?

13. Quais sao as desvantagens para si?

Que tipo de desvantagens ve no uso de telemedicina para si como medico?

14. Quais sao as vantagens para o paciente?

Que tipo de vantagens ve no uso de telemedicina para o paciente?

15. Quais sao as desvantagens para o paciente?

Que tipo de desvantagens ve no uso de telemedicina para o paciente?

16. Aconselharia outro medico da sua especialidade a usar telemedicina?

PERGUNTAS SOBRE GESTAO

17. As teleconsultas sao consideradas consultas de especialidade ou de familia?

Sao contabilizadas como consultas hospitalares? Entram na producao do hospital?
Como sao pagas? De que forma este esquema difere do usado nas consultas presenciais?
Porque e que e necessario um esquema especifico para telemedicina?

Figure A-6 Interview script (page 3 of 3)

Appendix B

SURVEY INSTRUMENT

As mentioned in Section 5.1.2, I conducted a survey of 100 teleconsultation patients and 100 face-to-face outpatients who had appointments in the second half of 2011 with consultants from the *Hospital do Espírito Santo de Évora*. Respondents were asked about distances travelled and modes of transport used. Patients who had had consultations in neurology, dermatology, physical and rehabilitation medicine, and general surgery were included in the survey. Lists of all patients who had an appointment during the study period were produced for each specialty and each pathway (i.e. teleconsultation or face-to-face). Each specialty/pathway combination was given a weighting factor according to the proportion of all appointments in the study population. The hundred patients were randomly selected from each pathway, and within these specialties the number of patients reflected the weighting factors calculated above.

The survey was conducted in February 2012 by personnel from the hospital's Patient Management Department, who regularly conduct patient satisfaction surveys. Patients were contacted by telephone. If a patient failed to answer the call, the next patient on the specific specialty/pathway list would be contacted. None of the patients who were successfully

contacted declined to participate. Copies of the survey instruments (in Portuguese) are available below (Figures B-1 and B-2).

INQUÉRITO CONSULTA PRESENCIAL

Por favor indique apenas uma opção por pergunta

1. Que meio de transporte utilizou para chegar ao hospital?	☐ ₁ Viatura própria (ou familiar, amigo) ☐ ₂ Autocarro ☐ ₃ Táxi ☐ ₄ Ambulância ☐ ₅ Outro: _____
2. A que distância vive do hospital?	_____ metros / quilómetros (indique a unidade)
3. Quanto tempo esperou na sala de espera do hospital pela consulta?	_____ minutos / horas (indique a unidade)
4. Quanto tempo passou desde que foi para o hospital até que voltou a casa?	_____ minutos / horas (indique a unidade)
5. Quanto tempo esperou pela consulta?	_____ dias / semanas / meses (indique a unidade)
6. Resultou nova consulta no HESE?	☐ ₁ Sim ☐ ₂ Não
7. Qual a sua ocupação?	☐ ₁ Activo (empregado, estudante activo) ☐ ₂ Inactivo (reformado, desempregado) ☐ ₃ Estudante inactivo
8. Trouxe acompanhante?	☐ ₁ Sim ☐ ₂ Não ➔ Passe para a pergunta 9.
9. Qual a ocupação do seu acompanhante?	☐ ₁ Activo (empregado, estudante trabalhador) ☐ ₂ Inactivo (reformado, desempregado) ☐ ₃ Estudante inactivo
10. Se o seu centro de saúde tivesse teleconsultas preferia uma teleconsulta a esta consulta?	☐ ₁ Sim ☐ ₂ Não
Explique porquê:	

O questionário continua no verso ➔

Figure B-1 Survey instrument for face-to-face patients (page 1 of 2)

11. Qual o nível de satisfação geral com esta consulta?	☐ ₁ Muito satisfeito
	☐ ₂ Satisfeito
	☐ ₃ Indiferente
	☐ ₄ Insatisfeito
	☐ ₅ Muito insatisfeito
12. Quanto ganhou em 2010? (todo o tipo de rendimentos)	☐ ₁ < €7,000
	☐ ₂ €7,001 - €15,000
	☐ ₃ €15,001 - €30,000
	☐ ₄ €30,001 - €60,000
	☐ ₅ > €60,000
13. Como descreve o seu estado de saúde?	☐ ₁ Muito bom
	☐ ₂ Bom
	☐ ₃ Razoável
	☐ ₄ Mau
	☐ ₅ Muito mau

A PREENCHER DEPOIS DO TELEFONEMA:

14. Centro de saúde referenciador: _____

15. Data do pedido de consulta (dd-mm-aaaa): _____

16. Data de marcação da consulta (dd-mm-aaaa): _____

17. Data de realização da consulta (dd-mm-aaaa): _____

Ano em que nasceu o utente (aaaa):

--	--	--	--

Sexo do utente: ☐₁ Masculino ☐₂ Feminino

Por favor confirme que respondeu a todas as questões.

Figure B-2 Survey instrument for face-to-face patients (page 2 of 2)

INQUÉRITO TELECONSULTA

Por favor indique apenas uma opção por pergunta

1. Que meio de transporte utilizou para chegar ao centro de saúde?	☐ ₁ Viatura própria (ou familiar, amigo) ☐ ₂ Autocarro ☐ ₃ Taxi ☐ ₄ Ambulância ☐ ₅ Outro: _____
2. A que distância vive do centro de saúde?	_____ metros / quilómetros (indique a unidade)
3. Quanto tempo esperou na sala de espera do centro de saúde pela teleconsulta?	_____ minutos / horas (indique a unidade)
4. Quanto tempo passou desde que foi para o centro de saúde até que voltou a casa?	_____ minutos / horas (indique a unidade)
5. Quanto tempo esperou pela teleconsulta?	_____ dias / semanas / meses (indique a unidade)
6. Resultou consulta/tratamento no hospital?	☐ ₁ Não ☐ ₂ Sim, no HESE. ☐ ₃ Sim no hospital de _____
7. Qual a sua ocupação?	☐ ₁ Activo (empregado, estudante activo) ☐ ₂ Inactivo (reformado, desempregado) ☐ ₃ Estudante inactivo
8. Trouxe acompanhante?	☐ ₁ Sim ☐ ₂ Não → Passe para a pergunta 9.
9. Qual a ocupação do seu acompanhante?	☐ ₁ Activo (empregado, estudante trabalhador) ☐ ₂ Inactivo (reformado, desempregado) ☐ ₃ Estudante inactivo
10. Se esta teleconsulta não fosse possível iria ao HESE?	☐ ₁ Sim ☐ ₂ Não ↓

Explique porquê:

O questionário continua no verso →

Figure B-3 Survey instrument for teleconsultation patients (page 1 of 2)

11. Qual o nível de satisfação geral com esta teleconsulta?	☐ ₁ Muito satisfeito
	☐ ₂ Satisfeito
	☐ ₃ Indiferente
	☐ ₄ Insatisfeito
	☐ ₅ Muito Insatisfeito
12. Quanto ganhou em 2010? (todo o tipo de rendimentos)	☐ ₁ < €7,000
	☐ ₂ €7,001 - €15,000
	☐ ₃ €15,001 - €30,000
	☐ ₄ €30,001 - €60,000
	☐ ₅ > €60,000
13. Como descreve o seu estado de saúde?	☐ ₁ Muito bom
	☐ ₂ Bom
	☐ ₃ Razoável
	☐ ₄ Mau
	☐ ₅ Muito mau

A PREENCHER DEPOIS DO TELEFONEMA:

14. Centro de saúde referenciador: _____

15. Data do pedido de consulta (dd-mm-aaaa): _____

16. Data de marcação da consulta (dd-mm-aaaa): _____

17. Data de realização da consulta (dd-mm-aaaa): _____

Ano em que nasceu o utente (aaaa):

--	--	--	--

Sexo do utente: ☐₁ Masculino ☐₂ Feminino

Por favor confirme que respondeu a todas as questões.

Figure B-4 Survey instrument for teleconsultation patients (page 2 of 2)

Appendix C

THE DISCRETE CHOICE EXPERIMENT

Each participant in the discrete choice experiment was given a questionnaire which included an information sheet, an initial page of general questions (age, gender, etc.) and 16 (or 17 depending on the version) choice sets. Copies of the information sheet, general questions and an example of a choice set are available below (Figures C-1 to C-3).

ESTUDO SOBRE REFERENCIAÇÃO PARA CONSULTA EXTERNA

Qual o objectivo deste estudo?

Com este estudo procuramos investigar o Impacto de vários factores na decisão de referenciação.

Sou obrigado(a) a participar?

Não. A decisão é sua. Se decidir participar, pode no futuro suspender a sua participação a qualquer momento sem que seja necessária uma justificação. A sua participação é muito importante.

O que é esperado de mim?

O estudo envolve o preenchimento de um questionário (anexo), o qual demora, em média, cerca de 10 minutos a preencher.

A minha participação neste estudo é confidencial?

Toda a informação que advier deste estudo é anónima e confidencial. Os resultados serão guardados num computador da Imperial College Business School, em Londres, protegido por password e sem quaisquer elementos identificadores (não é possível saber quem preencheu cada questionário).

Quais são os potenciais benefícios de participar?

A sua participação é essencial para perceber qual o Impacto de vários factores na referenciação. A validade do estudo depende do número de respostas, pelo que a sua participação é indispensável.

O que acontecerá aos resultados do estudo?

Os resultados do estudo serão incluídos num relatório e potencialmente em publicações científicas. Serão também disponibilizados aos participantes (e APMGF), caso estes o desejem.

Quem está a organizar e a financiar este estudo?

Este estudo está inserido num projecto de investigação que conta com o apoio da Administração Regional de Saúde do Alentejo, do Professor James Barlow e do Dr. Steffen Bayer da Imperial College Business School. O projecto é co-financiado pela Fundação para a Ciência e Tecnologia e pelo Health and Care Infrastructure Research and Innovation Centre do UK EPSRC.

Há alguém responsável por rever a qualidade do estudo?

Sim. O projecto em que este estudo está inserido é sujeito a revisões anuais por parte de um comité científico da Imperial College Business School cujo director é o Professor Enrico Aulio. O questionário foi submetido a um pré-teste na região de saúde de Lisboa e Vale do Tejo, no ano passado, e a versão final foi distribuída, já este ano, no 30º Encontro Nacional de Medicina Geral e Familiar, em Aveiro. A distribuição do questionário neste encontro foi autorizada pela APMGF.

Se tiver dúvidas, questões ou comentários, por favor utilize os contactos disponibilizados acima.

Figure C-1 Information sheet used in the discrete choice experiment

Neste questionário, ser-lhe-ão apresentados vários casos de pacientes, baseados em casos reais. Tendo em conta a informação disponibilizada em cada caso, deverá escolher a opção que acha mais adequada, de entre as que lhe são apresentadas.

Por favor, responda a todas as perguntas e indique apenas uma opção por pergunta.

O questionário demora, na sua totalidade, cerca de 10 minutos a preencher. Na medida do possível, uma vez que comece o questionário, preencha-o até ao fim, com o mínimo de pausas.

A sua participação é confidencial e anónima. Não é possível identificar os participantes neste estudo.

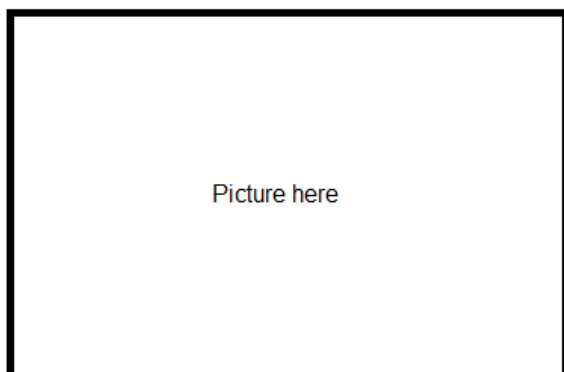
Por favor comece por responder às seguintes questões:

Idade: _____	Sexo: ☐ ₁ Masculino ☐ ₂ Feminino
Há quantos anos trabalha como médico de família: _____	
Aproximadamente a que distância do Hospital de Évora fica o seu centro de saúde (em km): _____	
Aproximadamente quantos utentes estão inscritos no seu centro de saúde: _____	
Aproximadamente a que distância fica o seu centro de saúde do consultório de dermatologia privado mais próximo (em km; se não souber, deixe em branco): _____	
Pratica actualmente, ou já praticou, telemedicina:	☐ ₁ Sim ☐ ₂ Não
Tem um interesse especial em dermatologia:	☐ ₁ Sim ☐ ₂ Não
Em média, como descreve o estado de saúde dos utentes inscritos no seu centro de saúde?	☐ ₁ Muito bom
	☐ ₂ Bom
	☐ ₃ Razoável
	☐ ₄ Mau
	☐ ₅ Muito mau

Por favor vir a página para começar o questionário ➡

Figure C-2 General questions used in the discrete choice experiment

5. Um doente se apresenta-se na consulta com uma lesão com 3 a 4 anos de evolução com mudança de aspecto nos últimos meses, por vezes com prurido, castanha, heterogénea, bordo irregular, com 5 mm.



Tendo em conta que:

- Em média, o tempo de espera para uma consulta com prioridade normal é de 365 dias.
- A consulta de dermatologia tem lugar no seu centro de saúde.
- O doente, e/ou a família, parecem neutros relativamente à referenciação.

Escolhe:

☐₁ Não referenciar o doente.

☐₂ Referenciar o doente com prioridade 'normal'.

☐₃ Referenciar o doente com prioridade 'urgente'.

Figure C-3 Example of a choice set used in the discrete choice experiment

Appendix D

THE WAITING TIME DATA SET

As mentioned in Chapter 6, the data set used to estimate the elasticities of referrals and non-completion with respect to waiting time is based on a database of outpatient activity from the *Hospital do Espírito Santo de Évora*, Alentejo's central hospital. The database contains information on outpatient appointments since July 2001, when the information system that collects and aggregates the data was first introduced. The data set includes all records of outpatient referrals for first appointments, completed first appointments and missed first appointments for the period between July 2001 and August 2012, for practices in health sub-regions *Alentejo Central I*, *Alentejo Central II* and *Baixo Alentejo*.

The original data set contains a total of 9,165 records of referrals which led to completed appointments and 3,490 records of referrals which were not completed. 39 records of referrals which were completed had to be excluded as the scheduling date (i.e., the day the appointment was scheduled) was prior to the triage date (i.e., the date the referral was returned to the scheduling personnel by the triaging physician). This is simply not possible in reality and is likely the result of a database error. Likewise, 16 records of non-completed referrals were excluded for the same reason. No other errors were detected, so

that the final data set included 9,126 valid completed referrals and 3,474 valid non-completed referrals.

Ideally, I would have taken advantage of the fact the data set is at the individual-level. However, this was found to be unfeasible as there were very few covariates available at the individual-level. The hospital database does not include any information on the patients besides a 4-digit post code, the practice from which the patient was referred, age, gender, identification of sub-system (i.e., special public and private insurance schemes for certain professions, e.g., teachers), exemption from co-payment (e.g., due to disability, long-term chronic condition, children aged 12 years old or younger, etc.), and these last two only for the completed referrals (i.e., subsystem and exemption were not available for non-completed referrals).

As mentioned above, the data set included a 4-digit post code for the patient's residence. In Portugal, complete post codes are of the form xxxx-xxx. The first four digits identify one of nine national postal regions and the last three digits identify building blocks and designated addresses (i.e., addresses with high volumes of mail). The 4-digit post code is thus insufficient to accurately estimate location. As such, I used the location of the practice as a proxy for patient residence. I then had to devise a way to estimate the distance and travelling time between each of the 27 practices in the data set and the supply of dermatological care: the two hospitals in Évora and Beja and the closest private provider.

First, I contacted the Portuguese Directorate-General of Health which provided a list of addresses for private providers in Alentejo with outpatient dermatological services. Second, I retrieved the addresses for all 27 practices and the hospitals in Évora and Beja from the website of the Central Administration of the Health System. Using Google Maps™, I estimated the coordinates for each provider. I then developed a routine in Matlab™ to estimate the distances and travelling times between combinations of coordinates using the Google Maps™ application programming interface (API). The routine reads the coordinates

of the 27 practices and the 2 hospitals plus 22 private providers from an Excel file. It then estimates distances and travelling times between each of the 27 practices and each of the 24 secondary care providers, using Google Maps™ API. The resulting estimates are stored in an XML file which is parsed (i.e., read) and distances and travelling times copied to a second Excel file. Based on the distances and travelling times estimated between each practice and the 22 private providers, I am able to select the closest private provider for each practice. I include in the data set the distance and travelling times for the closest private provider, the central hospital in Évora and the hospital in Beja. The code for the routine is available on request.

Appendix E

SIMULATION MODEL DOCUMENTATION

The model documentation was produced according to good practices (Martinez-Moyano, 2012). It includes general information such as the number of variables and the time unit used in simulations, as well as variable-specific information such as equation, units and brief description. As shown in Table C-1, the model has 374 variables, 66 stocks and 13 lookup functions. These numbers provide a misleading picture of true model size. As explained in Chapter 8, the model includes secondary dermatological care (i.e., the hospital), face-to-face primary care (i.e., a face-to-face practice) and teleconsultation primary care (i.e., three potential teleconsultation practices). The three teleconsultation practices are essentially reproductions of a teleconsultation practice. All the behaviours and mechanisms involved are replicated three times, thus increasing the number of variables in the model, despite there being no new behaviour.

Large models can be presented more clearly using multiple views. Each view contains a part of the model (e.g., primary care flows) which is connected to other views/parts of the model through shadow variables (i.e., variables which are defined in one view and used in another; shadow variables appear in the model in the format <variable>). The model

presented in this thesis is divided into 14 views. In the remainder of this appendix, I will provide figures of each view followed by information on the variables defined in that view (information on shadow variables is not provided; for information on a specific shadow variable consult the view in which the variable is defined).

It is important to note the terminology used in Chapter 8 was chosen to ease communication of the model. The actual simulation model (as implemented in Vensim™) makes use of simpler variable names. A useful naming rule is that variables related to face-to-face care end with the letter *f* and variables related to teleconsultation end with the letter *t*. As a number of variables related to teleconsultation practices are copies, I provide only information for one of the practices (practice A); variable names for other practices are the same but use the letters B and C. For example, the variable *GP Waiting list t A* is a stock equivalent to the *Primary care waiting list* in Figure 8-3 for practice A. For practices B and C, the respective waiting lists for GP appointments are defined as *GP Waiting list t B* and *GP Waiting list t C*. This naming approach will be clear throughout the figures in this appendix.

Table C-1 Model assessment results

Model information	
Total number of variables	374
Total number of stocks	66
Data lookup functions	13
Time unit	Month
Initial time	0
Final time	132
Time step	0.03125

Complete model documentation is available on request.

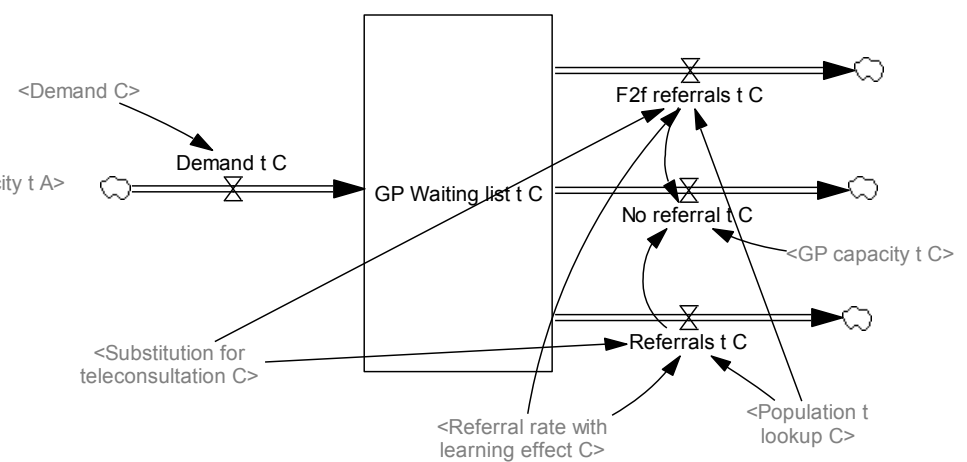
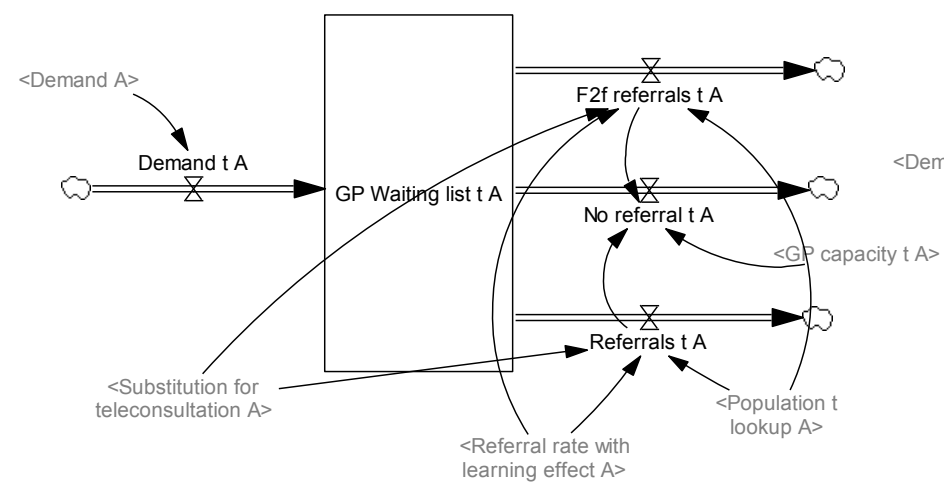
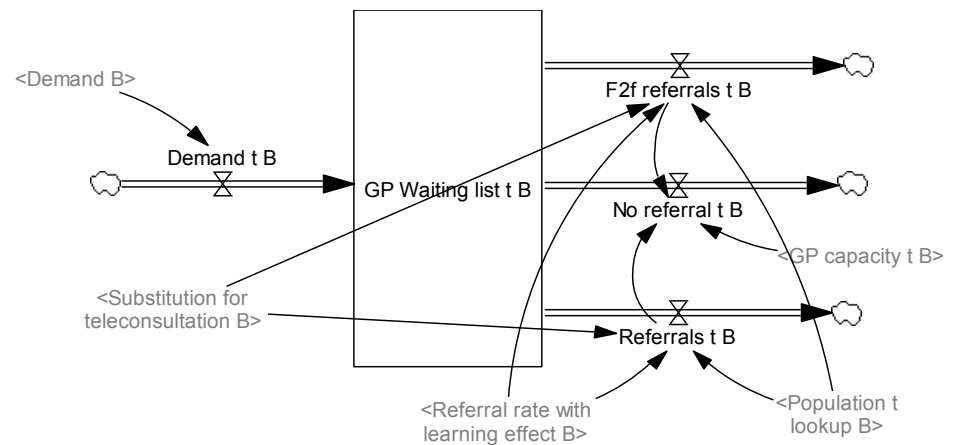
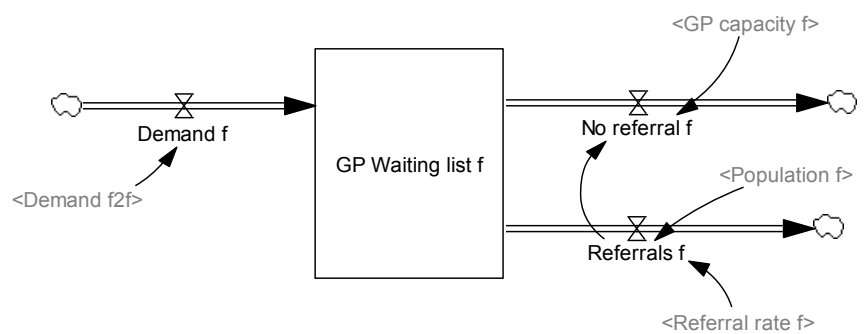


Figure C-1 View 1: Primary care flows

Demand f (Patient/Month)

= Demand f2f

Description: *Demand for GP appointments in face-to-face practices.*

Demand t A (Patient/Month)

= Demand A

Description: *Demand for GP appointments in teleconsultation practice A.*

F2f referrals t A (Patient/Month)

= Population t lookup A/1000*Referral rate with learning effect A*Substitution for teleconsultation A

Description: *Referrals to face-to-face appointments from teleconsultation practice A.*

GP Waiting list f (Patients)

= [Demand f-No referral f-Referrals f dt + [Demand f2f/4]

Description: *Waiting list for primary care appointments in face-to-face practices.*

GP Waiting list t A (Patients)

= [Demand t A-F2f referrals t A-No referral t A-Referrals t A dt + [Demand A/4]

Description: *Waiting list for GP appointments in teleconsultation practice A.*

No referral f (Patient/Month)

= GP capacity f-Referrals f

Description: *GP appointments which do not lead to a referral in face-to-face practices.*

No referral t A (Patient/Month)

= GP capacity t A-F2f referrals t A-Referrals t A

Description: *GP appointments which do not lead to referrals in teleconsultation practice A.*

Referrals f (Patient/Month)

= Population f/1000*Referral rate f

Description: *Referrals from face-to-face practices.*

Referrals t A (Patients/Month)

= Population t lookup A/1000*Referral rate with learning effect A*(1-Substitution for teleconsultation A)

Description: *Referrals to teleconsultation, practice A.*

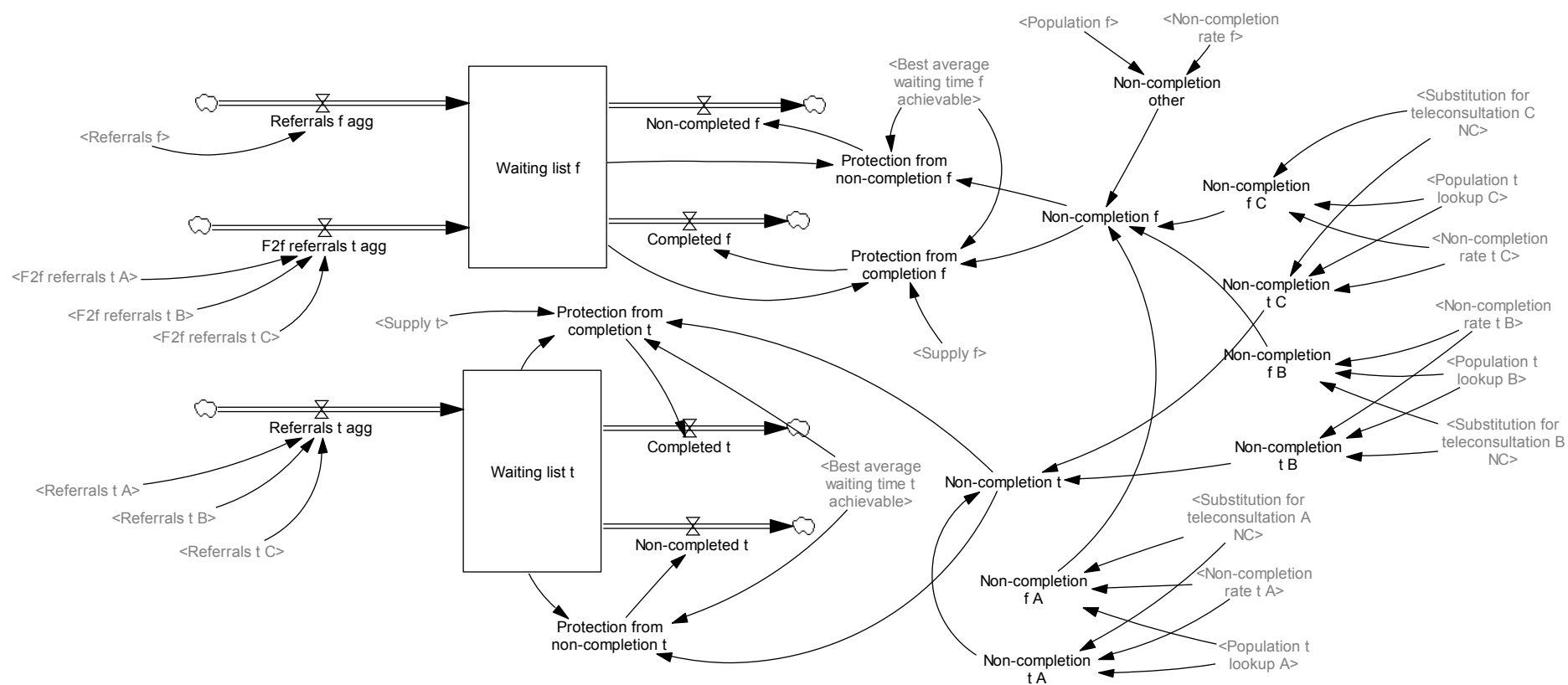


Figure C-2 View 2: Secondary care flows

Completed f (Patients/Month)

= Protection from completion f

Description: *Completed face-to-face appointments.*

Completed t (Patients/Month)

= Protection from completion t

Description: *Completed teleconsultations.*

F2f referrals t agg (Patient/Month)

= F2f referrals t A+F2f referrals t B+F2f referrals t C

Description: *Referrals to face-to-face appointments from all teleconsultation practices.*

Non-completed f (Patients/Month)

= Protection from non-completion f

Description: *Missed face-to-face appointments.*

Non-completed t (Patient/Month)

= Protection from non-completion t

Description: *Missed teleconsultations.*

Non-completion f (Patients)

= Non-completion f A+Non-completion f B+Non-completion f C+Non-completion other

Description: *Face-to-face non-completion.*

Non-completion f A (Patients)

= Substitution for teleconsultation A NC*Non-completion rate t A*Population t lookup A/1000

Description: *Face-to-face non-completion in practice A.*

Non-completion other (Patients)

= Non-completion rate f*(Population f/1000)

Description: *Non-completion in face-to-face practices.*

Non-completion t (Patients)

= Non-completion t A+Non-completion t B+Non-completion t C

Description: *Total teleconsultation non-completion.*

Non-completion t A (Patients)

= (1-Substitution for teleconsultation A NC)*Non-completion rate t A*Population t lookup A/1000

Description: *Teleconsultation non-completion in practice A.*

Protection from completion f (Patients/Month)

= IF THEN ELSE(Waiting list f/Best average waiting time f achievable>=(Supply f-Non-completion f), Supply f-Non-completion f, IF THEN ELSE(Waiting list f/Best average waiting time f achievable>=Non-completion f, Waiting list f/Best average waiting time f achievable-Non-completion f, 0))

Description: *Auxiliary variable protecting the waiting list for teleconsultations from going below zero, by limiting the number of face-to-face appointments.*

Protection from completion t (Patients/Month)

= IF THEN ELSE(Waiting list t/Best average waiting time t achievable>=(Supply t-Non-completion t), Supply t-Non-completion t, IF THEN ELSE(Waiting list t/Best average waiting time t achievable>=Non-completion t, Waiting list t/Best average waiting time t achievable-Non-completion t, 0))

Description: *Auxiliary variable protecting the waiting list for teleconsultations from going below zero, by limiting the number of teleconsultations.*

Protection from non-completion f (Patients/Month)

= IF THEN ELSE(Waiting list f/Best average waiting time f achievable>=Non-completion f, Non-completion f, Waiting list f/Best average waiting time f achievable)

Description: *Auxiliary variable protecting the waiting list for teleconsultations from going below zero, by limiting the number of missed face-to-face appointments.*

Protection from non-completion t (Patients/Month)

= IF THEN ELSE(Waiting list t/Best average waiting time t achievable>=Non-completion t, Non-completion t, Waiting list t/Best average waiting time t achievable)

Description: *Auxiliary variable protecting the waiting list for teleconsultations from going below zero, by limiting the number of missed teleconsultations.*

Referrals f agg (Patient/Month)

= Referrals f

Description: *Referrals from face-to-face practices per month.*

Referrals t agg (Patients/Month)

= Referrals t A+Referrals t B+Referrals t C

Description: *Referrals to teleconsultation, all practices.*

Waiting list f (Patients)

$$= \int \text{F2f referrals t agg} + \text{Referrals f agg} - \text{Completed f} - \text{Non-completed f} dt + [81.5202 * (186.5 * 12 / 365)]$$

Description: *Waiting list for face-to-face appointments (initial value: 186.5 days)*

Waiting list t (Patients)

$$= \int \text{Referrals t agg} - \text{Completed t} - \text{Non-completed t} dt + [0]$$

Description: *Waiting list for teleconsultations.*

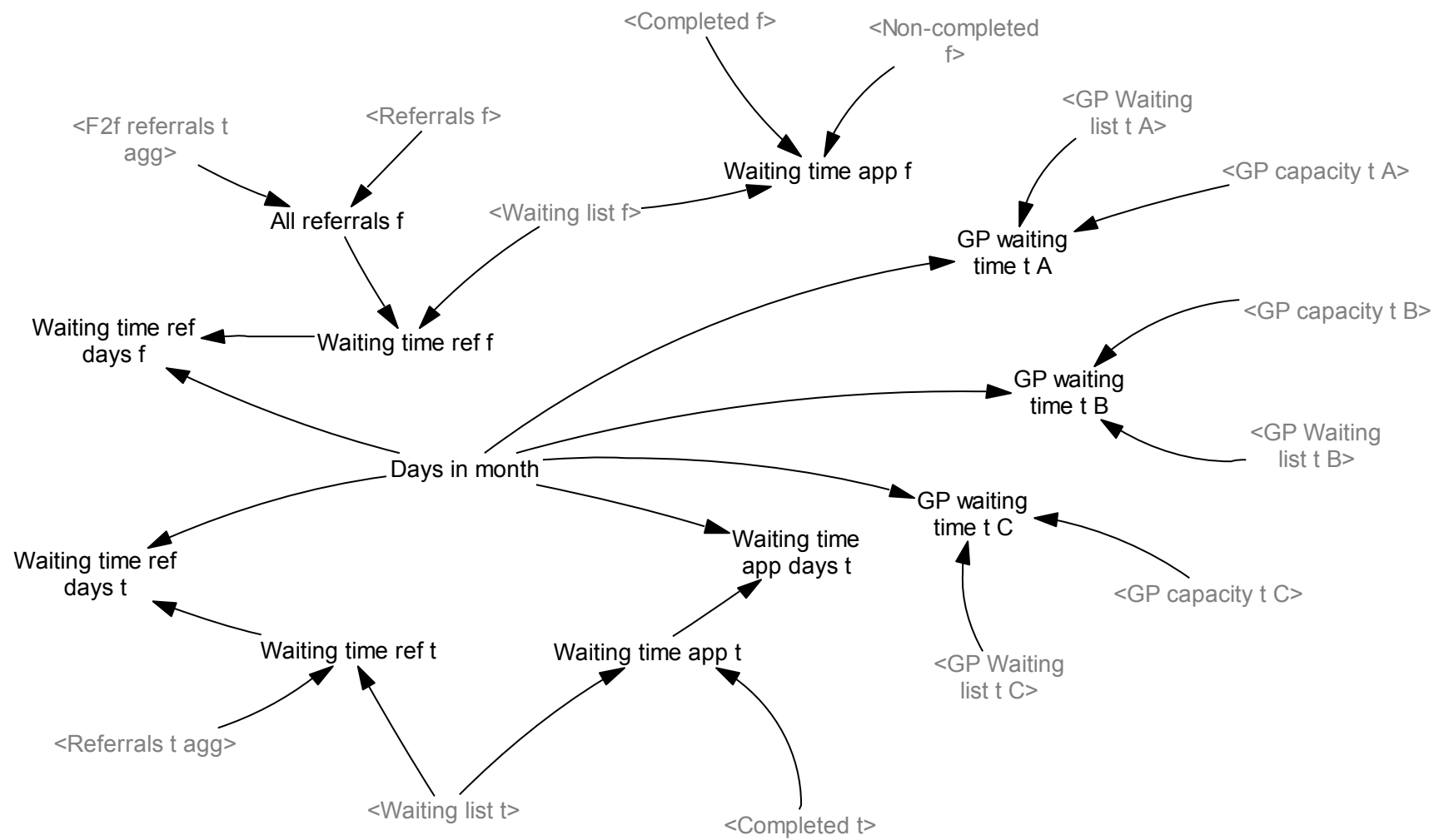


Figure C-3 View 3: Waiting times

Days in month (Day/Month)

= 365/12

Description: *Average number of days in a month.*

GP waiting time t A (Day)

= GP Waiting list t A/GP capacity t A*Days in month

Description: *Average waiting time for a GP appointment in practice A.*

Waiting time app days t (Day)

= Waiting time app t*Days in month

Description: *Waiting time for teleconsultations in days (calculated using appointments).*

Waiting time app f (Month)

= Waiting list f/(Completed f+Non-completed f)

Description: *Waiting time for face-to-face appointments in months (calculated using appointments).*

Waiting time app t (Month)

= ZIDZ(Waiting list t,Completed t)

Description: *Waiting time for teleconsultations in months (calculated using appointments).*

Waiting time ref days f (Day)

= Waiting time ref f*Days in month

Description: *Waiting time for face-to-face appointments in days (calculated using referrals).*

Waiting time ref days t (Day)

= Waiting time ref t*Days in month

Description: *Waiting time for teleconsultations in days (calculated using referrals).*

Waiting time ref f (Month)

= Waiting list f/All referrals f

Description: *Waiting time for face-to-face appointments in months (calculated using referrals).*

Waiting time ref t (Month)

= ZIDZ(Waiting list t, Referrals t agg)

Description: *Waiting time for teleconsultations in months (calculated using referrals).*

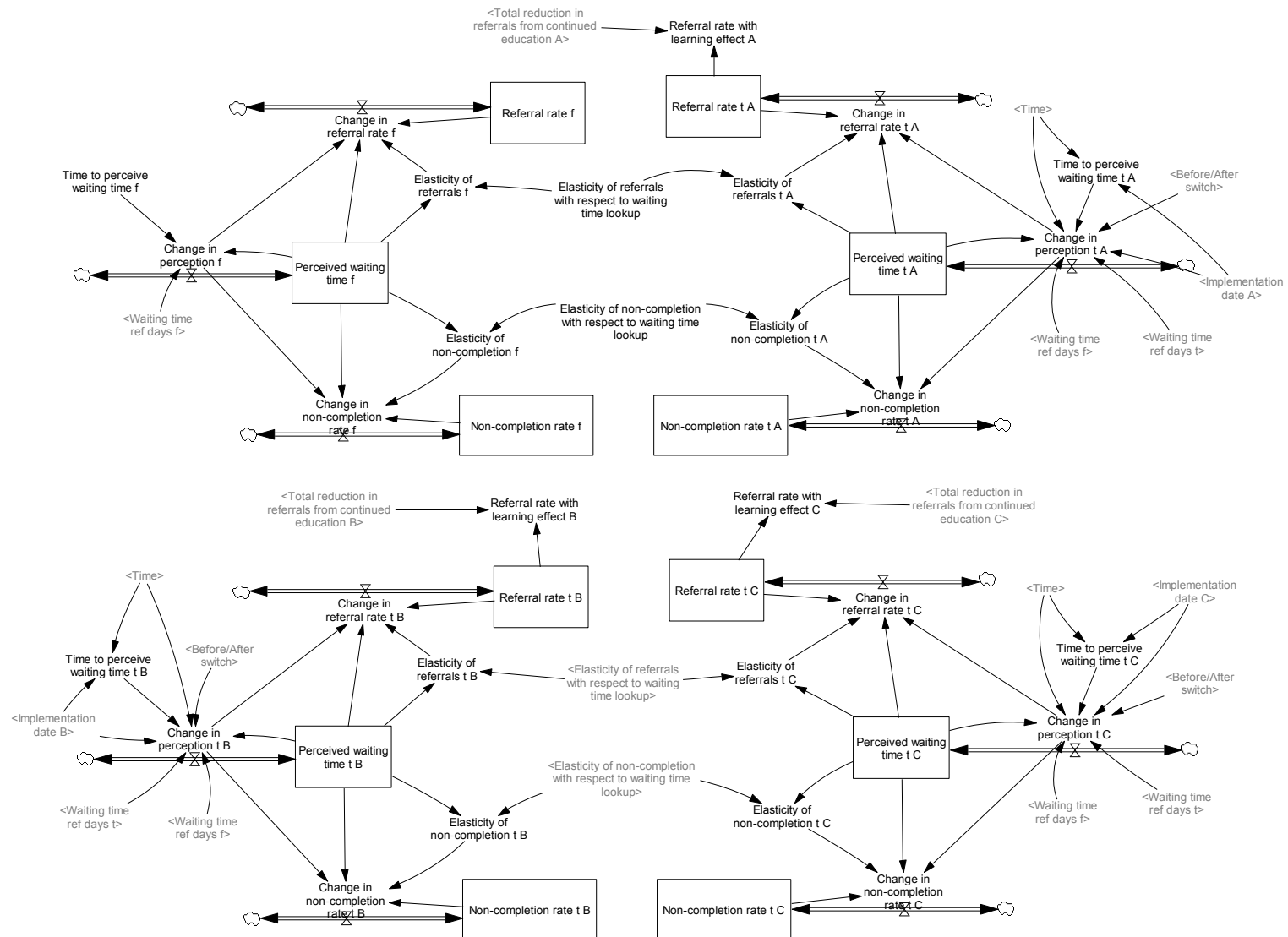


Figure C-4 View 4: Utilisation rates

Change in non-completion rate f (Patients/(Month*People))

= Elasticity of non-completion f*Non-completion rate f*Change in perception f/Perceived waiting time f

Description: *Change in non-completion rate for face-to-face appointments.*

Change in non-completion rate t A (Patients/(Month*People))

= Elasticity of non-completion t A*Non-completion rate t A*Change in perception t A/Perceived waiting time t A

Description: *Change in non-completion rate for teleconsultations in practice A.*

Change in perception f (Days/Month)

= (Waiting time ref days f-Perceived waiting time f)/Time to perceive waiting time f

Description: *Change in perceived waiting time for a face-to-face appointment.*

Change in perception t A (Days/Month)

= IF THEN ELSE(Before/After switch=1, IF THEN ELSE(Time>=Implementation date A,
(Waiting time ref days t A-Perceived waiting time t A)/Time to perceive waiting time t A,
(Waiting time ref days f-Perceived waiting time t A)/Time to perceive waiting time t A),(Waiting time ref days f-Perceived waiting time t A)/Time to perceive waiting time t A)

Description: *Change in perceived waiting time for a teleconsultation in practice A.*

Change in referral rate f (Patients/(Month*People))

= Elasticity of referrals f*Referral rate f*Change in perception f/Perceived waiting time f

Description: *Change in referral rate for face-to-face appointments.*

Change in referral rate t A (Patients/(Month*People))

= Elasticity of referrals t A*Referral rate t A*Change in perception t A/Perceived waiting time t A

Description: *Change in referral rate for teleconsultation in practice A.*

Elasticity of non-completion f (dmnl)

= Elasticity of non-completion with respect to waiting time lookup(Perceived waiting time f)

Description: *Elasticity of non-completion with respect to waiting time for face-to-face appointments.*

Elasticity of non-completion t A (dmnl)

= Elasticity of non-completion with respect to waiting time lookup(Perceived waiting time t A)

Description: *Elasticity of non-completion with respect to waiting time for teleconsultations in practice A.*

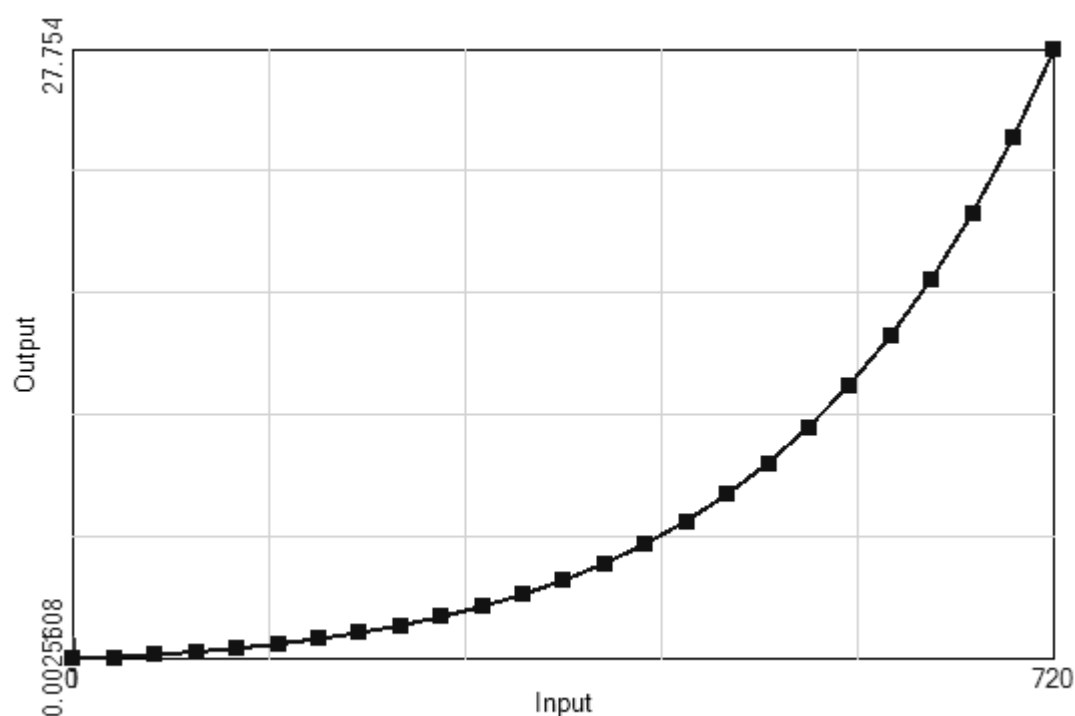
Elasticity of non-completion with respect to waiting time lookup (dmnl)

= [(0,0.0025608)-

(720,27.7549)],(0,0.0025608),(30,0.0888736),(60,0.195787),(90,0.326993),(120,0.486772),(150,0.680081),(180,0.912653),(210,1.19111),(240,1.52311),(270,1.91746),(300,2.38433),(330,2.93542),(360,3.58421),(390,4.34616),(420,5.23908),(450,6.28336),(480,7.50246),(510,8.92325),(540,10.5765),(570,12.4975),(600,14.7267),(630,17.3103),(660,20.3011),(690,23.7596),(720,27.7549)

Description: *Elasticity of non-completion with respect to waiting time, as a function of waiting time (assumed to be the same for face-to-face and teleconsultations).*

Source: *Econometric analysis, Chapter 6*

**Elasticity of referrals f (dmnl)**

= Elasticity of referrals with respect to waiting time lookup(Perceived waiting time f)

Description: *Elasticity of referrals with respect to waiting time for face-to-face appointments.*

Elasticity of referrals t A (dmnl)

= Elasticity of referrals with respect to waiting time lookup(Perceived waiting time t A)

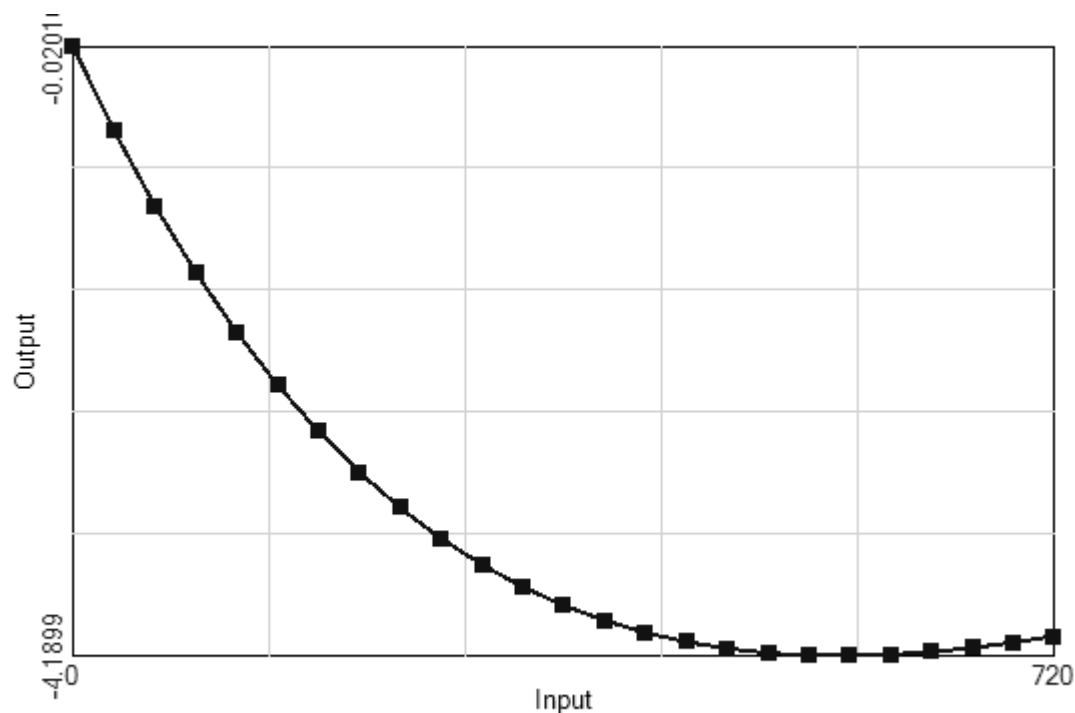
Description: *Elasticity of referrals with respect to waiting time for teleconsultations in practice A.*

Elasticity of referrals with respect to waiting time lookup (dmnl)

= [(0,-4.1899)-(720,-0.020164)],(0,-0.020164),(30,-0.592696),(60,-1.10584),(90,-1.56423),(120,-1.97214),(150,-2.33359),(180,-2.65229),(210,-2.93169),(240,-3.17502),(270,-3.38527),(300,-3.56521),(330,-3.71742),(360,-3.84427),(390,-3.94801),(420,-4.03068),(450,-4.09418),(480,-4.14028),(510,-4.17061),(540,-4.18669),(570,-4.1899),(600,-4.18154),(630,-4.1628),(660,-4.13477),(690,-4.09848),(720,-4.05484)

Description: *Elasticity of referrals with respect to waiting time, as a function of waiting time (assumed to be the same for face-to-face and teleconsultations).*

Source: *Econometric analysis, Chapter 6*

**Non-completion rate f (Patients/People)**

= $\int \text{Change in non-completion rate f } dt + [0.0827]$

Description: *Non-completion rate (missed appointments per 1000 inhabitants) for face-to-face appointments.*

Non-completion rate t A (Patients/People)

= $\int \text{Change in non-completion rate t A } dt + [0.0827]$

Description: *Non-completion rate (missed appointments per 1000 inhabitants) for teleconsultations, in practice A.*

Perceived waiting time f (Days)

$$= \text{[Change in perception f]} dt + \text{[Waiting time ref days f]}$$

Description: *Perceived waiting time for a face-to-face appointment.*

Perceived waiting time t A (Days)

$$= \text{[Change in perception t A]} dt + \text{[Waiting time ref days f]}$$

Description: *Perceived waiting time for teleconsultation in practice A.*

Referral rate f (Patients/People)

$$= \text{[Change in referral rate f]} dt + [0.273817]$$

Description: *Referral rate (referrals per 1000 inhabitants) for face-to-face appointments.*

Referral rate t A (Patients/People)

$$= \text{[Change in referral rate t A]} dt + [0.273817]$$

Description: *Referral rate (referrals per 1000 inhabitants) for teleconsultation in practice A.*

Referral rate with learning effect A (Patients/People)

$$= \text{Referral rate t A} - \text{Total reduction in referrals from continued education A}$$

Description: *Referral rate accounting for continued GP education in practice A.*

Time to perceive waiting time f (Days)

$$= 365$$

Description: *Time patients and GPs take to perceive the actual waiting time for face-to-face appointments.*

Source: *Van Ackere and Smith(1999)*

Time to perceive waiting time t A (Days)

$$= \text{IF THEN ELSE}(\text{Time} \geq \text{Implementation date A}, 365/12, 365)$$

Description: *Time patients and GPs take to perceive the actual waiting time in practice A.*

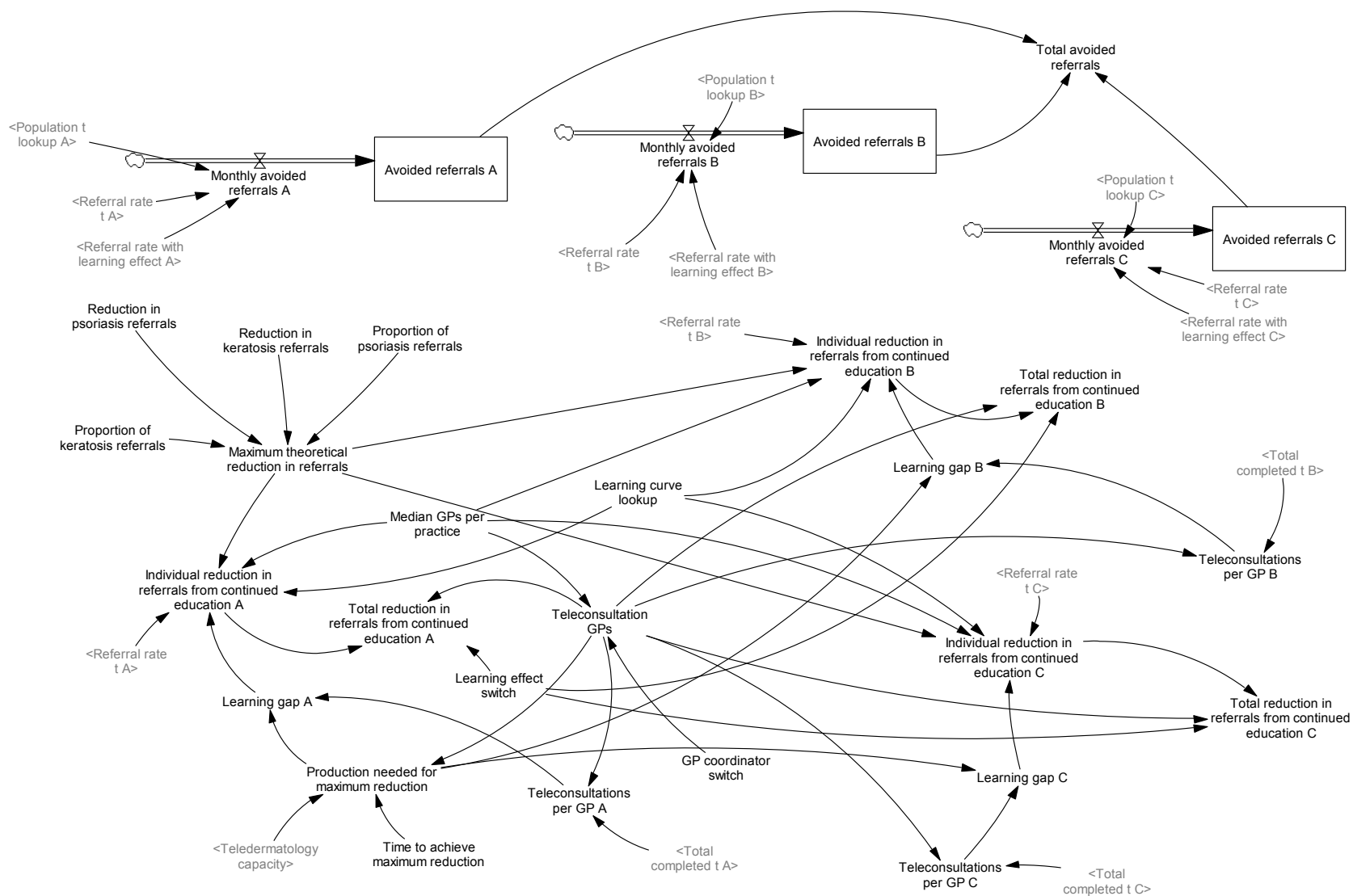


Figure C-5 View 5: GP education and learning

Avoided referrals A (Patients)

= $\int \text{Monthly avoided referrals A } dt + [0]$

Description: *Cumulative number of referrals avoided as a consequence of continued GP education and learning, in practice A.*

GP coordinator switch (dmnl)

= 0

Description: *Auxiliary variable indicating whether there is a GP coordinator which is responsible for all teleconsultations (1), or whether each GP is present in his/her patients' teleconsultations (0).*

Individual reduction in referrals from continued education

A (Patients/(People*Physician))

= $\frac{\text{Learning curve lookup}(\text{Learning gap A}) * \text{Maximum theoretical reduction in referrals} * \text{Referral rate t A}}{\text{Median GPs per practice}}$

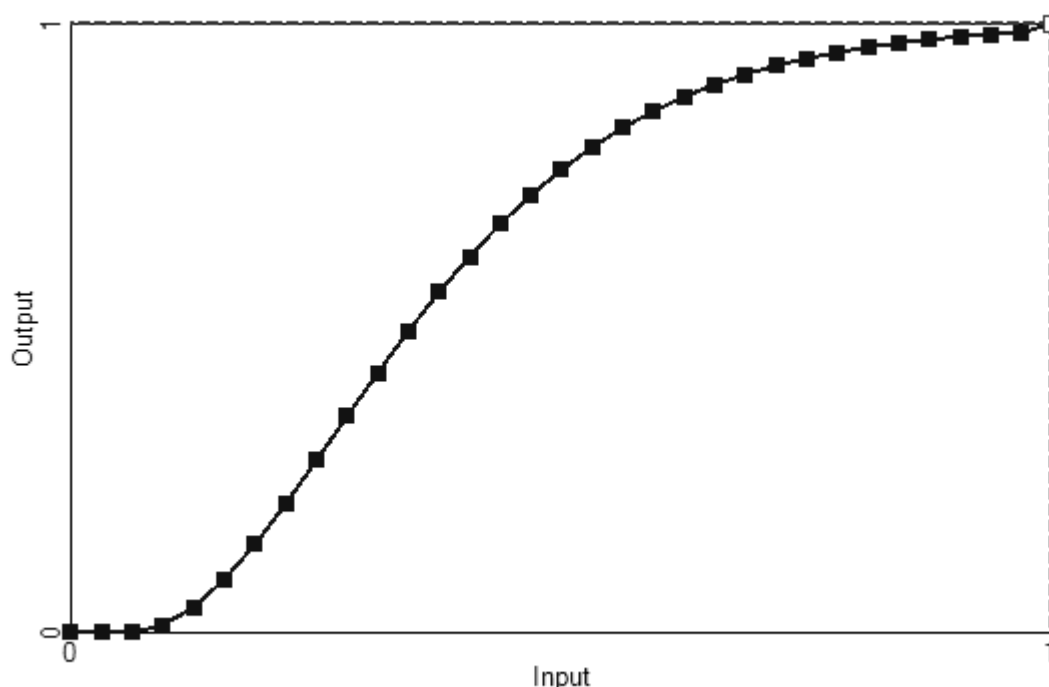
Description: *Actual individual reduction in referral rate from continued GP education and learning, in practice A.*

Learning curve lookup (dmnl)

= [(0,0)-

(1,1)],(0,0),(0.03125,0),(0.0625,0),(0.09375,0.0128746),(0.125,0.0424459),(0.15625,0.0877271),(0.1875,0.145508),(0.21875,0.211865),(0.25,0.282992),(0.28125,0.355601),(0.3125,0.427075),(0.34375,0.495478),(0.375,0.559487),(0.40625,0.618295),(0.4375,0.671507),(0.46875,0.719037),(0.5,0.761026),(0.53125,0.797766),(0.5625,0.829645),(0.59375,0.857104),(0.625,0.880601),(0.65625,0.900589),(0.6875,0.917504),(0.71875,0.931749),(0.75,0.943694),(0.78125,0.953671),(0.8125,0.961974),(0.84375,0.96886),(0.875,0.974555),(0.90625,0.97925),(0.9375,0.983111),(0.96875,0.986278),(1,1),(1e+011,1)

Description: Lookup function detailing the learning curve (3rd order smooth, smoothing time = 0.4).

**Learning effect switch (dmnl)**

= 1

Description: Auxiliary variable indicating whether to include learning effects in simulations (0 - No, 1 - Yes).

Learning gap A (dmnl)

= ZIDZ(Teleconsultations per GP A, Production needed for maximum reduction)

Description: Learning gap expressed as the percentage of teleconsultations per GP relative to the production needed to achieve maximum learning, in practice A.

Maximum theoretical reduction in referrals (dmnl)

= Proportion of keratosis referrals*Reduction in keratosis referrals+Proportion of psoriasis referrals*Reduction in psoriasis referrals

Description: *Maximum reduction possible accounting for: 1) number of GPs that participate in teleconsultations (i.e., potential learning); 2) the proportion of patients those GPs see that can benefit (i.e., potential impact); 3) the average reduction in referrals for those patients.*

Source: *Discrete choice experiment; HESE, E.P.E.*

Median GPs per practice (Physician)

= 6

Description: *Median GPs per practice.*

Source: *INE, I.P.*

Monthly avoided referrals A (Patients/Month)

= (Referral rate t A-Referral rate with learning effect A)*Population t lookup A/1000

Description: *Avoided referrals in practice A.*

Production needed for maximum reduction (Patients/Physician)

= Teledermatology capacity*Time to achieve maximum reduction/Teleconsultation GPs

Description: *Number of teleconsultations needed for each physician to reach maximum reduction in referrals from continued education and learning.*

Proportion of keratosis referrals (dmnl)

= 0.18

Description: *Proportion of patients referred with keratosis relative to all referrals.*

Source: *HESE, I.P.*

Proportion of psoriasis referrals (dmnl)

= 0.16

Description: *Proportion of patients referred with psoriasis relative to all referrals.*

Source: *HESE, I.P.*

Reduction in keratosis referrals (dmnl/Physician)

= 0.18

Description: *Percentage reduction probability of referring a patient with keratosis due to continued GP education and learning from participation in teleconsultations.*

Source: *Discrete choice experiment*

Reduction in psoriasis referrals (dmnl/Physician)

= 0.42

Description: *Percentage reduction probability of referring a patient with psoriasis due to continued GP education and learning from participation in teleconsultations.*

Source: *Discrete choice experiment*

Referral rate with learning effect A (Patients/People)

= Referral rate t A-Total reduction in referrals from continued education A

Description: *Referral rate accounting for continued GP education in practice A.*

Teleconsultation GPs (Physician)

= IF THEN ELSE(GP coordinator switch=1, 1, Median GPs per practice)

Description: *Number of GPs practicing teleconsultation per practice.*

Teleconsultations per GP A (Patients/Physician)

= Total completed t A/Teleconsultation GPs

Description: *Number of cumulative teleconsultations per GP in practice A.*

Time to achieve maximum reduction (Months)

= 6

Description: *Time needed to achieve maximum learning effect.*

Source: *Exploratory interviews*

Total avoided referrals (Patients)

= Avoided referrals A+Avoided referrals B+Avoided referrals C

Description: *Cumulative avoided referrals, all teleconsultation practices.*

Total reduction in referrals from continued education A (Patients/People)

= IF THEN ELSE(Learning effect switch=1, Individual reduction in referrals from continued education A*Teleconsultation GPs, 0)

Description: *Actual aggregate reduction in referral rate from continued GP education and learning, in practice A.*

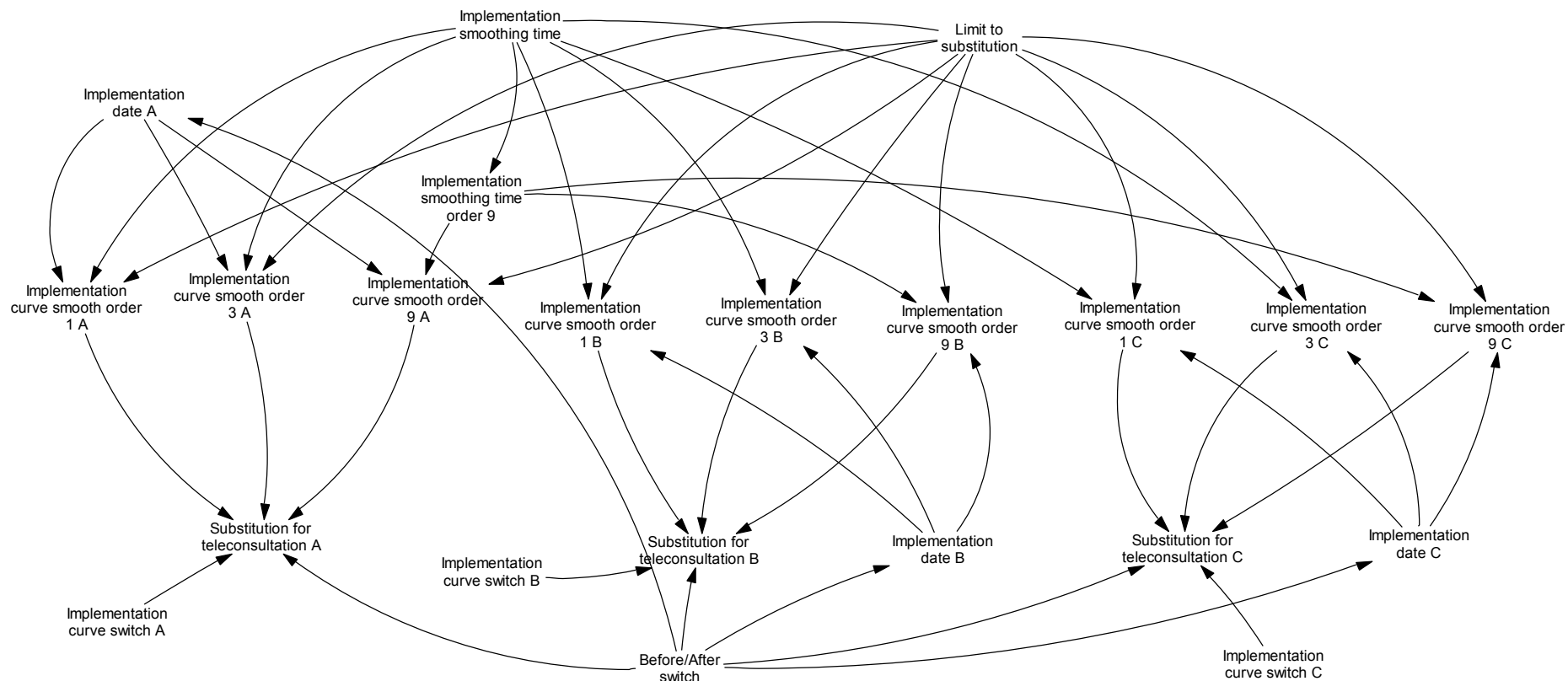


Figure C-6 View 6a: Implementation and substitution (part 1/2)

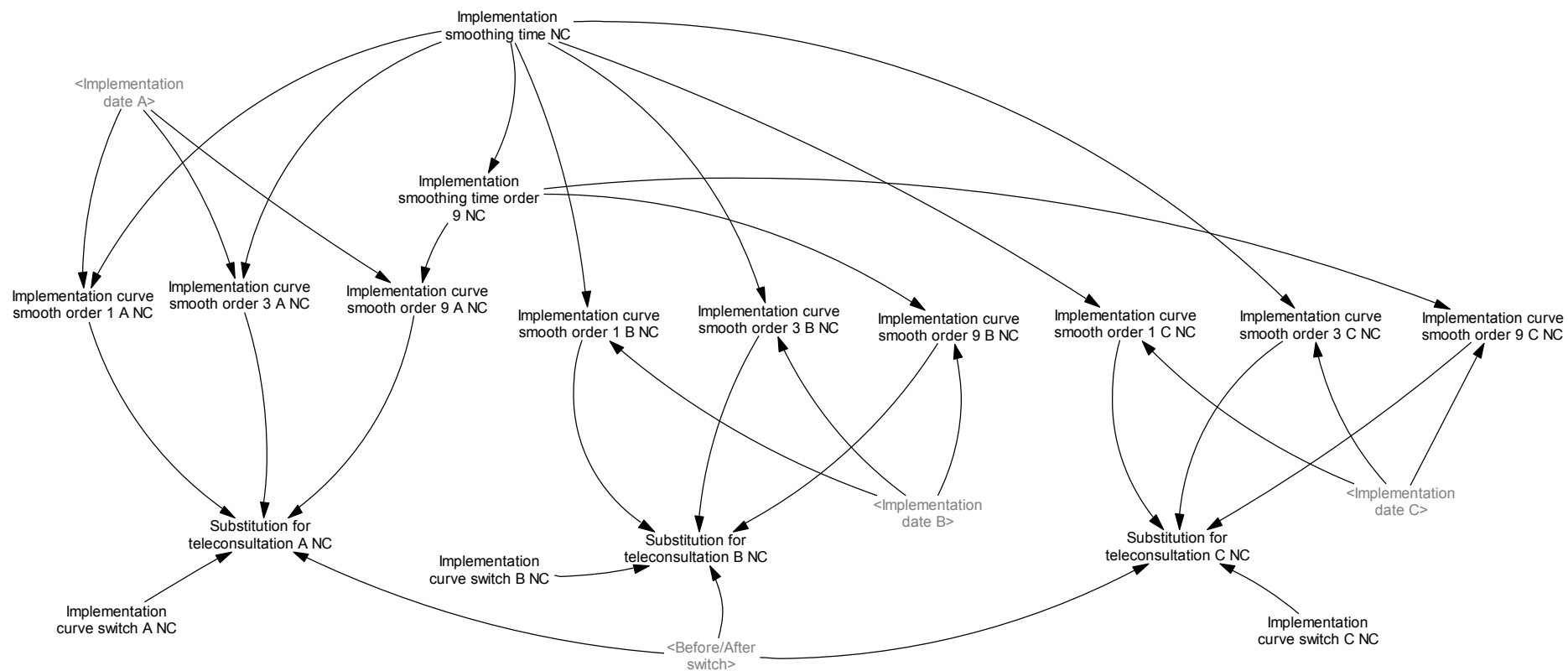


Figure C-7 View 6b: Implementation and substitution (part 2/2)

Before/After switch (dmnl)

= 1

Description: *Auxiliary variable indicating whether model is simulating situation before or after the introduction of teleconsultation (1 = after, 0 = before).*

Implementation curve smooth order 1 A (dmnl)

= DELAY FIXED(SMOOTHI(Limit to substitution, Implementation smoothing time, 1),Implementation date A, 1)

Description: *Teleconsultation implementation curve in practice A, for referrals (represents shift from face-to-face appointments to teleconsultations), 1st order smooth.*

Implementation curve smooth order 1 A NC (dmnl)

= DELAY FIXED(SMOOTHI(0, Implementation smoothing time NC, 1),Implementation date A, 1)

Description: *Teleconsultation implementation curve in practice A, for non-completion (represents shift from face-to-face appointments to teleconsultations), 1st order smooth.*

Implementation curve smooth order 3 A (dmnl)

= DELAY FIXED(SMOOTH3I(Limit to substitution, Implementation smoothing time, 1),Implementation date A, 1)

Description: *Teleconsultation implementation curve in practice A, for referrals (represents shift from face-to-face appointments to teleconsultations), 3rd order smooth.*

Implementation curve smooth order 3 A NC (dmnl)

= DELAY FIXED(SMOOTH3I(0, Implementation smoothing time NC, 1),Implementation date A, 1)

Description: *Teleconsultation implementation curve in practice A, for non-completion (represents shift from face-to-face appointments to teleconsultations), 3rd order smooth.*

Implementation curve smooth order 9 A (dmnl)

= DELAY FIXED(SMOOTH3I(SMOOTH3I(SMOOTH3I(Limit to substitution, Implementation smoothing time order 9, 0), Implementation smoothing time order 9, 0), Implementation smoothing time order 9, 1),Implementation date A, 1)

Description: *Teleconsultation implementation curve in practice A, for referrals (represents shift from face-to-face appointments to teleconsultations), 9th order smooth*

Implementation curve smooth order 9 A NC (dmnl)

= DELAY FIXED(SMOOTH3I(SMOOTH3I(SMOOTH3I(0, Implementation smoothing time order 9 NC, 0), Implementation smoothing time order 9 NC, 0), Implementation smoothing time order 9 NC, 1),Implementation date A,1)

Description: *Teleconsultation implementation curve in practice A, for non-completion (represents shift from face-to-face appointments to teleconsultations), 9th order smooth.*

Implementation curve switch A (dmnl)

= 3

Description: *Auxiliary variable indicating which implementation curve is being used in practice A (1 for 1st order smooth 3 for 3rd order smooth, 9 for 9th order smooth).*

Implementation curve switch A NC (dmnl)

= 3

Description: *Auxiliary variable indicating which implementation curve is being used in practice A, for non-completion (1 for 1st order smooth 3 for 3rd order smooth, 9 for 9th order smooth).*

Implementation date A (Month)

= IF THEN ELSE(Before/After switch=1, 12, 10000)

Description: *Simulation month at which practice A begins using teleconsultations.*

Implementation smoothing time (Months)

= 1

Description: *Time needed to achieve full implementation of teleconsultations.*

Source: *ARS Alentejo, I.P.*

Implementation smoothing time NC (Months)

= 1

Description: *Time needed to achieve full implementation of teleconsultations.*

Source: *ARS Alentejo, I.P.*

Implementation smoothing time order 9 (Month)

= Implementation smoothing time/3

Description: *Transformation of implementation smoothing time to be used in 9th order smooth.*

Implementation smoothing time order 9 NC (Month)

= Implementation smoothing time NC/3

Description: *Transformation of implementation smoothing time to be used in 9th order smooth.*

Limit to substitution (dmnl)

= 0.03

Description: *Proportion of referrals from teleconsultation practices that will always need face-to-face care (cannot be substituted for with teleconsultations).*

Source: *ARS Alentejo, I.P.*

Substitution for teleconsultation A (dmnl/Physician)

= IF THEN ELSE(Before/After switch=1, IF THEN ELSE(Implementation curve switch A=1, Implementation curve smooth order 1 A, IF THEN ELSE(Implementation curve switch A=3,Implementation curve smooth order 3 A, Implementation curve smooth order 9 A)),0)

Description: *Implementation of teleconsultation, substituting for face-to-face referrals, in practice A.*

Substitution for teleconsultation A NC (dmnl)

= IF THEN ELSE(Before/After switch=1, IF THEN ELSE(Implementation curve switch A NC=1, Implementation curve smooth order 1 A NC, IF THEN ELSE(Implementation curve switch A NC=3,Implementation curve smooth order 3 A NC, Implementation curve smooth order 9 A NC)),1)

Description: *Implementation of teleconsultation in practice A, for non-completion.*

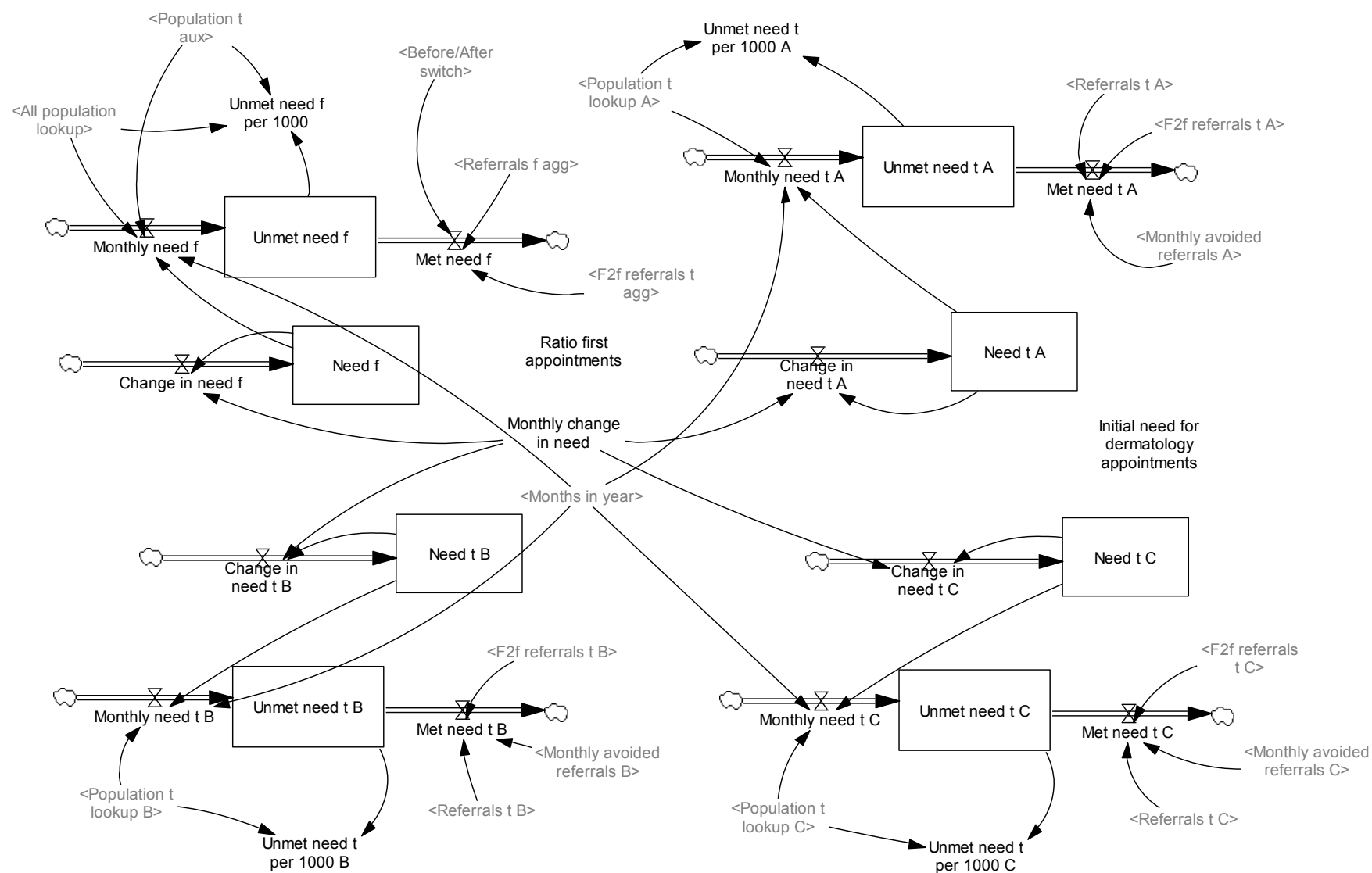


Figure C-8 View 7: Need for dermatological outpatient care

Change in need f (Patients/(People*Year*Month))

= Need f*Monthly change in need

Description: *Change in need for dermatology first appointments in face-to-face practices.*

Change in need t A (Patients/(People*Year*Month))

= Need t A*Monthly change in need

Description: *Change in need for dermatology first appointments in practice A.*

Initial need for dermatology appointments (Patients/(People*Year))

= 50/1000

Description: *Number of dermatology appointments which covers population needs in a given year.*

Source: ACSS

Met need f (Patients/Month)

= IF THEN ELSE(Before/After switch=0, Referrals f agg+F2f referrals t agg, Referrals f agg)

Description: *Need for dermatology appointments which is being met, in face-to-face practices.*

Met need t A (Patients/Month)

= F2f referrals t A+Monthly avoided referrals A+Referrals t A

Description: *Need for dermatology appointments which is being met, in practice A.*

Monthly change in need (dmnl)

= 0.0568402/12*0

Description: *Monthly growth rate in need for dermatological appointments.*

Source: Statistics Portugal, DGS

Monthly need f (Patients/Month)

= (All population lookup-Population t aux)*Need f/Months in year

Description: *Monthly need for dermatology first face-to-face appointments.*

Monthly need t A (Patients/Month)

= Population t lookup A*Need t A/Months in year

Description: *Monthly need for dermatology first teleconsultations, in practice A.*

Need f (Patients/(People*Year))

= [Change in need f dt + [Initial need for dermatology appointments*Ratio first appointments]

Description: *Actual need for dermatology first appointments in face-to-face practices.*

Need t A (Patients/(People*Year))

= $\int \text{Change in need t A } dt + [\text{Initial need for dermatology appointments} * \text{Ratio first appointments}]$

Description: *Actual need for dermatology first appointments in practice A.*

Ratio first appointments (dmnl)

= 0.4

Description: *Percentage of all appointments which are first appointments.*

Source: ACSS

Unmet need f (Patients)

= $\int \text{Monthly need f} - \text{Met need f } dt + [0]$

Description: *Unmet need for dermatology first appointments in face-to-face practices.*

Unmet need f per 1000 (Patients/People)

= $\text{Unmet need f} / ((\text{All population lookup} - \text{Population t aux}) / 1000)$

Description: *Unmet need for dermatology first appointments in face-to-face practices per 1000 inhabitants.*

Unmet need t A (Patients)

= $\int \text{Monthly need t A} - \text{Met need t A } dt + [0]$

Description: *Unmet need for dermatology first appointments in practice A.*

Unmet need t per 1000 A (Patients/People)

= $\text{Unmet need t A} / (\text{Population t lookup A} / 1000)$

Description: *Unmet need for dermatology first appointments in practice A, per 1000 inhabitants.*

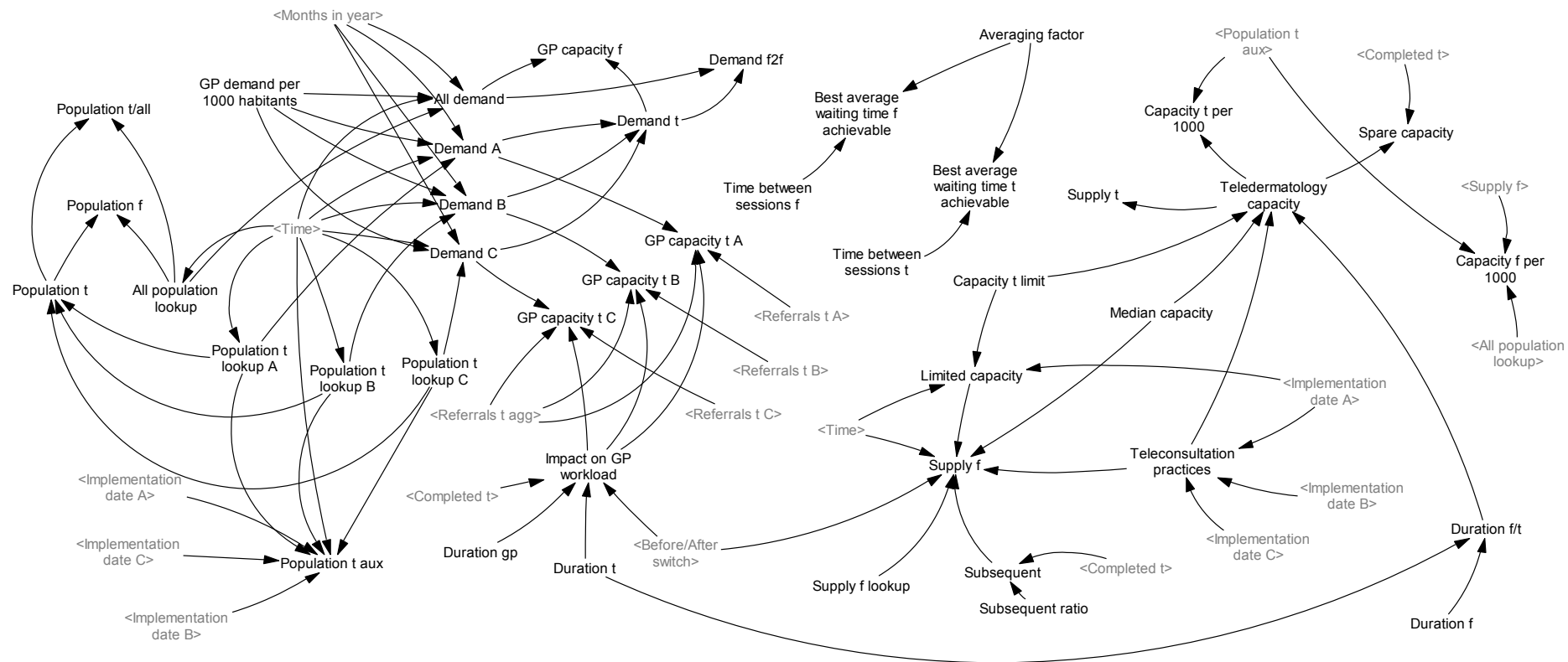


Figure C-9 View 8: Supply and demand

All demand (Patients/Month)

= GP demand per 1000 inhabitants(Time)*(All population lookup/1000)/Months in year

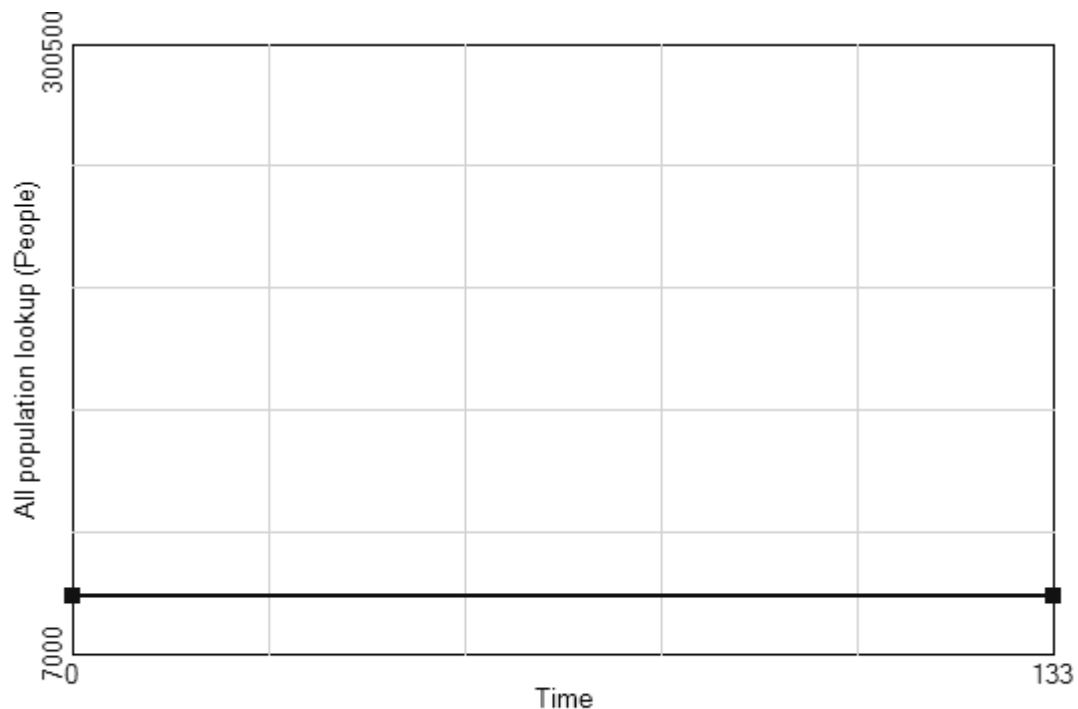
Description: Demand for GP appointments in face-to-face practices, per month.

All population lookup (People)

= WITH LOOKUP (Time,([(0,7000)-(133,3.005e+006)],(0,297718),(133,297718)))

Description: Population in face-to-face practices lookup function.

Source: Statistics Portugal



Averaging factor (dmnl)

= 0.5

Description: Averaging factor, defined as one half; when multiplied by the time between sessions gives the best achievable average waiting time per patient.

Best average waiting time f achievable (Month)

= Averaging factor*Time between sessions f

Description: Defined as half of the inverse of number of face-to-face sessions per month.

Best average waiting time t achievable (Month)

= Averaging factor*Time between sessions t

Description: Defined as half of the inverse of number teleconsultation sessions per month.

Capacity f per 1000 (Patients/People)

= $\frac{\text{Supply f}}{((\text{All population lookup} - \text{Population t aux})/1000)}$

Description: *Number of face-to-face appointments available per 1000 inhabitants.*

Capacity t limit (Patients/Month)

= 0

Description: *Limit to teleconsultation capacity. If zero, then limitless. Other options include 5 (1 hour of F2f appointments), 10 (2 hours of F2f appointments), etc.*

Capacity t per 1000 (Patients/People)

= $\text{ZIDZ}(\text{Teledermatology capacity}, (\text{Population t aux}/1000))$

Description: *Number of teleconsultations available per 1000 inhabitants.*

Demand A (Patients/Month)

= $\frac{\text{GP demand per 1000 inhabitants}(\text{Time}) * (\text{Population t lookup A}/1000)}{\text{Months in year}}$

Description: *Demand for GP appointments in teleconsultation practice A lookup function.*

This Lookup Table is used as a Data Repository.

Source: *Statistics Portugal*

Demand f2f (Patients/Month)

= $\text{All demand} - \text{Demand t}$

Description: *Demand for GP appointments in face-to-face practices.*

Demand t (Patients/Month)

= $\text{Demand A} + \text{Demand B} + \text{Demand C}$

Description: *Total demand for GP appointments in teleconsultation practices.*

Duration f (Minutes/Patient)

= 12

Description: *Average duration of face-to-face appointment.*

Source: *Interviews*

Duration f/t (dmnl)

= $\frac{\text{Duration f}}{\text{Duration t}}$

Description: *Ratio of duration of face-to-face appointments and teleconsultations.*

Duration gp (Minutes/Patient)

= 15

Description: *Average duration of GP appointment.*

Source: *Interviews*

Duration t (Minutes/Patient)

= 6

Description: *Average duration of teleconsultation.*

Source: *Interviews*

GP capacity f (Patients/Month)

= All demand-Demand t

Description: *GP capacity in face-to-face practices.*

GP capacity t A (Patient/Month)

= Demand A-ZIDZ(Impact on GP workload*Referrals t A, Referrals t agg)

Description: *GP capacity in teleconsultation practice A.*

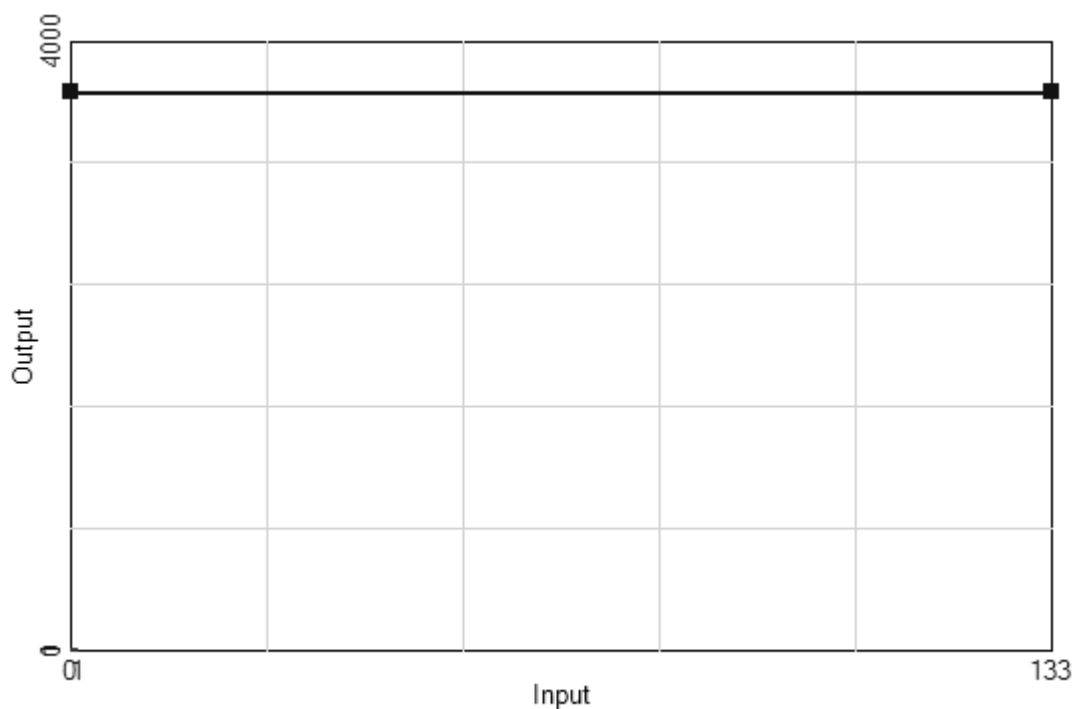
GP demand per 1000 inhabitants (Patients/(People*Year))

= [(0,0)-(133,4000)],(0,3671),(133,3671)

Description: *Median number of GP appointments per practice, per 1000 inhabitants.*

This Lookup Table is used as a Data Repository.

Source: *Statistics Portugal*



Impact on GP workload (Patients/Month)

= IF THEN ELSE(Before/After switch=1, Completed t*Duration t/Duration gp, 0)

Description: *Impact on GP workload from participation in teleconsultation.*

Limited capacity (Patients/Month)

= IF THEN ELSE(Time>=Implementation date A, IF THEN ELSE(Capacity t limit>0, Capacity t limit, 0), 0)

Description: *Teleconsultation capacity, when a limit is used.*

Median capacity (Patient/Month)

= 5

Description: *Median outpatient dermatology capacity.*

Source: *HESE, I.P.*

Population f (People)

= All population lookup-Population t

Description: *Population served by face-to-face practices.*

Population t (People)

= Population t lookup A+Population t lookup B+Population t lookup C

Description: *Population served by teleconsultation practices.*

Population t aux (People)

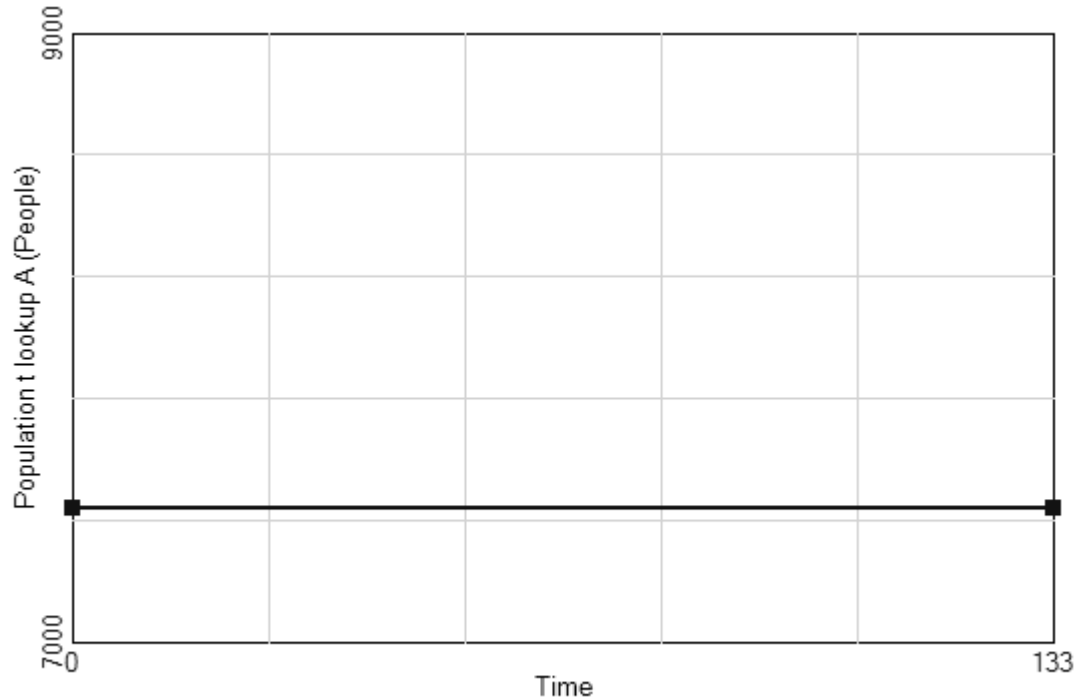
= IF THEN ELSE(Time>=Implementation date A, IF THEN ELSE(Time>=Implementation date B, IF THEN ELSE(Time>=Implementation date C, Population t lookup A+Population t lookup B+Population t lookup C, Population t lookup B+Population t lookup A), Population t lookup A), 0)

Description: *Auxiliary variable taking into account different dates of implementation of teleconsultation in different practices.*

Population t lookup A (People)

= WITH LOOKUP (Time,([(0,7000)-(133,9000)],(0,7442),(133,7442)))

Description: *Population in teleconsultation practice A lookup function.*



Population t/all (dmnl)

= Population t/All population lookup

Description: *Ratio of teleconsultation population and all population.*

Spare capacity

= Teledermatology capacity-Completed t

Description: *Unfilled teleconsultation capacity.*

Subsequent (Patient/Month)

= Completed t*Subsequent ratio

Description: *Number of teleconsultation patients that get a subsequent face-to-face appointment.*

Subsequent ratio (dmnl)

= 0.1

Description: *Ratio of teleconsultations that lead to a subsequent face-to-face appointment.*

Source: *Literature review*

Supply f (Patient/Month)

= IF THEN ELSE(Before/After switch=1, IF THEN ELSE(Limited capacity>0, Supply f lookup(Time)-Subsequent-Limited capacity, Supply f lookup(Time)-Subsequent-Median capacity*Teleconsultation practices), Supply f lookup(Time))

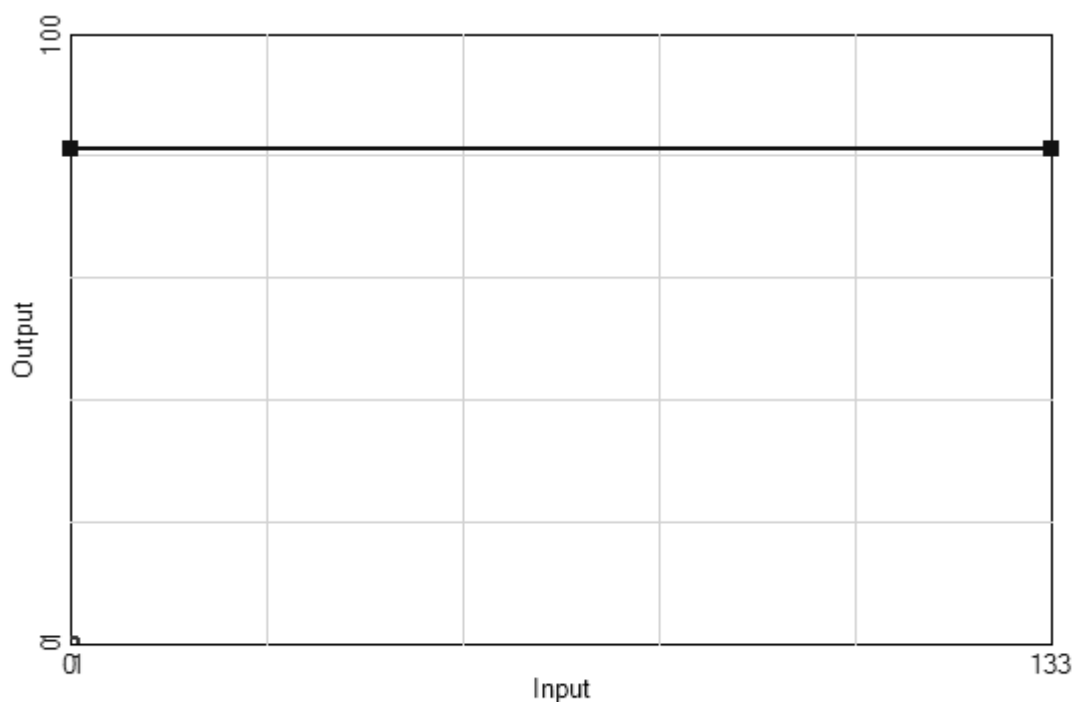
Description: *Face-to-face dermatology supply.*

Supply f lookup (Patients/Month)

= [(0,0)-(133,100)],(0,81.5202),(133,81.5202)

Description: *Face-to-face dermatology appointments lookup function.*

Source: *Hospital do Espírito Santo, E.P.E.*

**Supply t** (Patient/Month)

= Teledermatology capacity

Description: *Dermatology teleconsultations supply.*

Teleconsultation practices (Practice)

= STEP(1, Implementation date A)+STEP(1, Implementation date B)+STEP(1, Implementation date C)

Description: *Number of practices using teleconsultation at a given point in time.*

Teledermatology capacity (Patients)

= IF THEN ELSE(Capacity t limit>0, Capacity t limit*Duration f/t, Median capacity*Teleconsultation practices*Duration f/t)

Description: *Auxiliary variable determining whether a limited, median or unlimited teleconsultation capacity should be used.*

Time between sessions f (Month)

= 0.125

Description: *Number of face-to-face sessions per month.*

Time between sessions t (Month)

= 1

Description: *Number of teleconsultation sessions per month.*

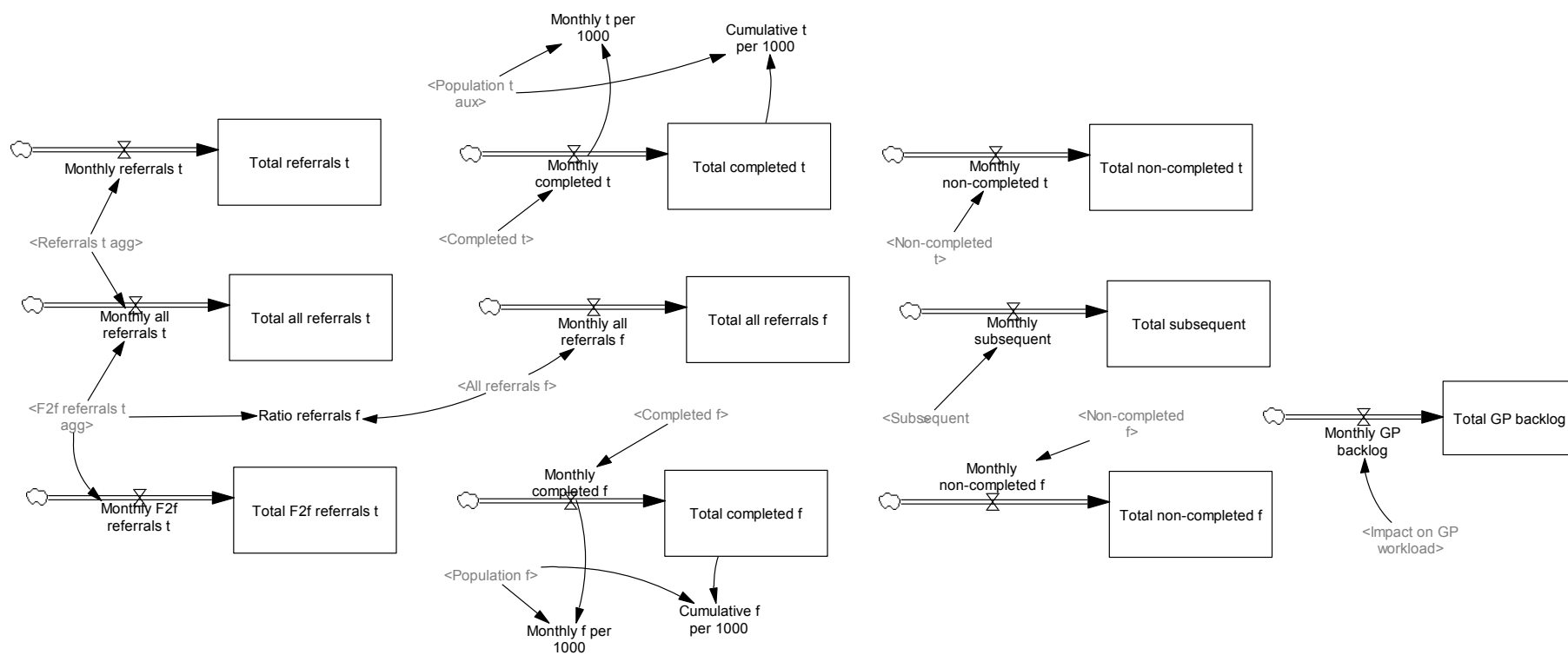


Figure C-10 View 9a: Production (part 1/2)

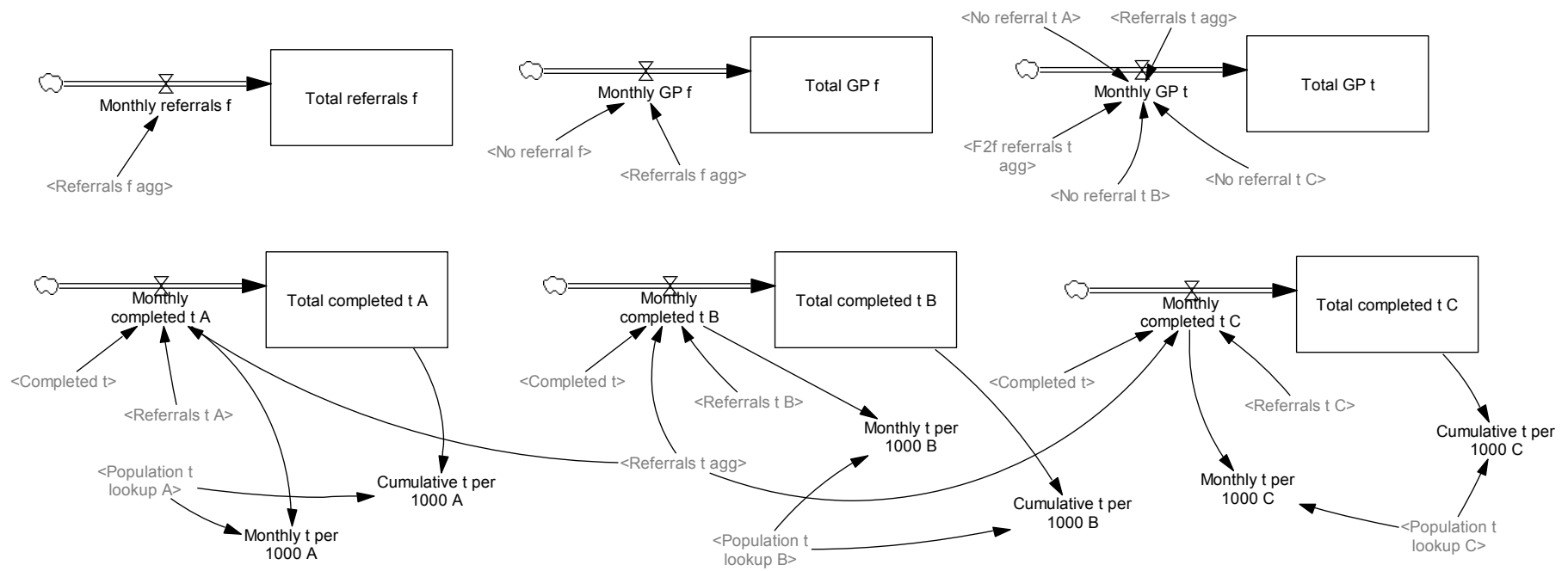


Figure C-11 View 9b: Production (part 2/2)

Cumulative f per 1000 (Patient/People)

= Total completed f/(Population f/1000)

Description: *Total face-to-face appointments per 1000 inhabitants.*

Cumulative t per 1000 (Patient/People)

= ZIDZ(Total completed t, Population t aux/1000)

Description: *Total teleconsultations per 1000 inhabitants.*

Cumulative t per 1000 A (Patient/People)

= Total completed t A/(Population t lookup A/1000)

Description: *Total teleconsultations per 1000 inhabitants, in practice A.*

Monthly all referrals f (Patients/Month)

= All referrals f

Description: *All referrals to face-to-face appointments per month (from both teleconsultation and face-to-face practices).*

Monthly all referrals t (Patients/Month)

= F2f referrals t agg+Referrals t agg

Description: *All referrals from teleconsultation practices per week (including to face-to-face appointments).*

Monthly completed f (Patients/Month)

= Completed f

Description: *Completed face-to-face appointments per month.*

Monthly completed t (Patients/Month)

= Completed t

Description: *Completed teleconsultations per month, all practices.*

Monthly completed t A (Patients/Month)

= Completed t*ZIDZ(Referrals t A, Referrals t agg)

Description: *Completed teleconsultations per month in practice A.*

Monthly f per 1000 (Patients/(People*Month))

= Monthly completed f/(Population f/1000)

Description: *Number of completed face-to-face appointments per month per 1000 inhabitants.*

Monthly F2f referrals t (Patients/Month)

= F2f referrals t agg

Description: *Referrals for face-to-face from teleconsultation practices per month.*

Monthly GP backlog (Patients/Month)

= Impact on GP workload

Description: *Number of GP appointments not performed due to increased workload from participation in teleconsultation.*

Monthly GP f (Patients/Month)

= No referral f+Referrals f agg

Description: *GP appointments in face-to-face practices per month.*

Monthly GP t (Patients/Month)

= F2f referrals t agg+Referrals t agg+No referral t A+No referral t B+No referral t C

Description: *GP appointments in teleconsultation practices per month.*

Monthly non-completed f (Patients/Month)

= Non-completed f

Description: *Missed face-to-face appointments per month.*

Monthly non-completed t (Patients/Month)

= Non-completed t

Description: *Missed teleconsultations per month.*

Monthly referrals f (Patients/Month)

= Referrals f agg

Description: *Referrals from face-to-face practices per month.*

Monthly referrals t (Patients/Month)

= Referrals t agg

Description: *Referrals to teleconsultation per month.*

Monthly subsequent (Patients/Month)

= Subsequent

Description: *Subsequent face-to-face appointments per month.*

Monthly t per 1000 (Patients/(People*Month))

= ZIDZ(Monthly completed t, Population t aux/1000)

Description: *Monthly teleconsultations per 1000 inhabitants, all practices.*

Monthly t per 1000 A (Patients/(People*Month))

= $\frac{\text{Monthly completed t A}}{(\text{Population t lookup A}/1000)}$

Description: *Monthly teleconsultations per 1000 inhabitants, in practice A.*

Ratio referrals f (dmnl)

= $\frac{\text{F2f referrals t agg}}{\text{All referrals f}}$

Description: *Ratio of referrals to face-to-face appointments from teleconsultation practices to all referrals to face-to-face appointments.*

Total all referrals f (Patients)

= $\int \text{Monthly all referrals f } dt + [0]$

Description: *Cumulative referrals to face-to-face appointments.*

Total all referrals t (Patients)

= $\int \text{Monthly all referrals t } dt + [0]$

Description: *Cumulative referrals from teleconsultation practices.*

Total completed f (Patients)

= $\int \text{Monthly completed f } dt + [0]$

Description: *Cumulative completed face-to-face appointments.*

Total completed t (Patients)

= $\int \text{Monthly completed t } dt + [0]$

Description: *Cumulative completed teleconsultations, all practices.*

Total completed t A (Patients)

= $\int \text{Monthly completed t A } dt + [0]$

Description: *Cumulative completed teleconsultations, practice A.*

Total F2f referrals t (Patients)

= $\int \text{Monthly F2f referrals t } dt + [0]$

Description: *Cumulative referrals to face-to-face from teleconsultations practices.*

Total GP backlog (Patients)

= $\int \text{Monthly GP backlog } dt + [0]$

Description: *Cumulative GP backlog from increased workload.*

Total GP f (Patients)

= $\int \text{Monthly GP f } dt + [0]$

Description: *Cumulative GP appointments in face-to-face practices.*

Total GP t (Patients)

$$= \int \text{Monthly GP t } dt + [0]$$

Description: *Cumulative GP appointments in teleconsultation practices.*

Total non-completed f (Patients)

$$= \int \text{Monthly non-completed f } dt + [0]$$

Description: *Cumulative missed face-to-face appointments.*

Total non-completed t (Patients)

$$= \int \text{Monthly non-completed t } dt + [0]$$

Description: *Cumulative missed teleconsultations.*

Total referrals f (Patients)

$$= \int \text{Monthly referrals f } dt + [0]$$

Description: *Cumulative referrals from face-to-face practices.*

Total referrals t (Patients)

$$= \int \text{Monthly referrals t } dt + [0]$$

Description: *Cumulative referrals to teleconsultation.*

Total subsequent (Patients)

$$= \int \text{Monthly subsequent } dt + [0]$$

Description: *Cumulative subsequent face-to-face appointments, following a teleconsultation.*

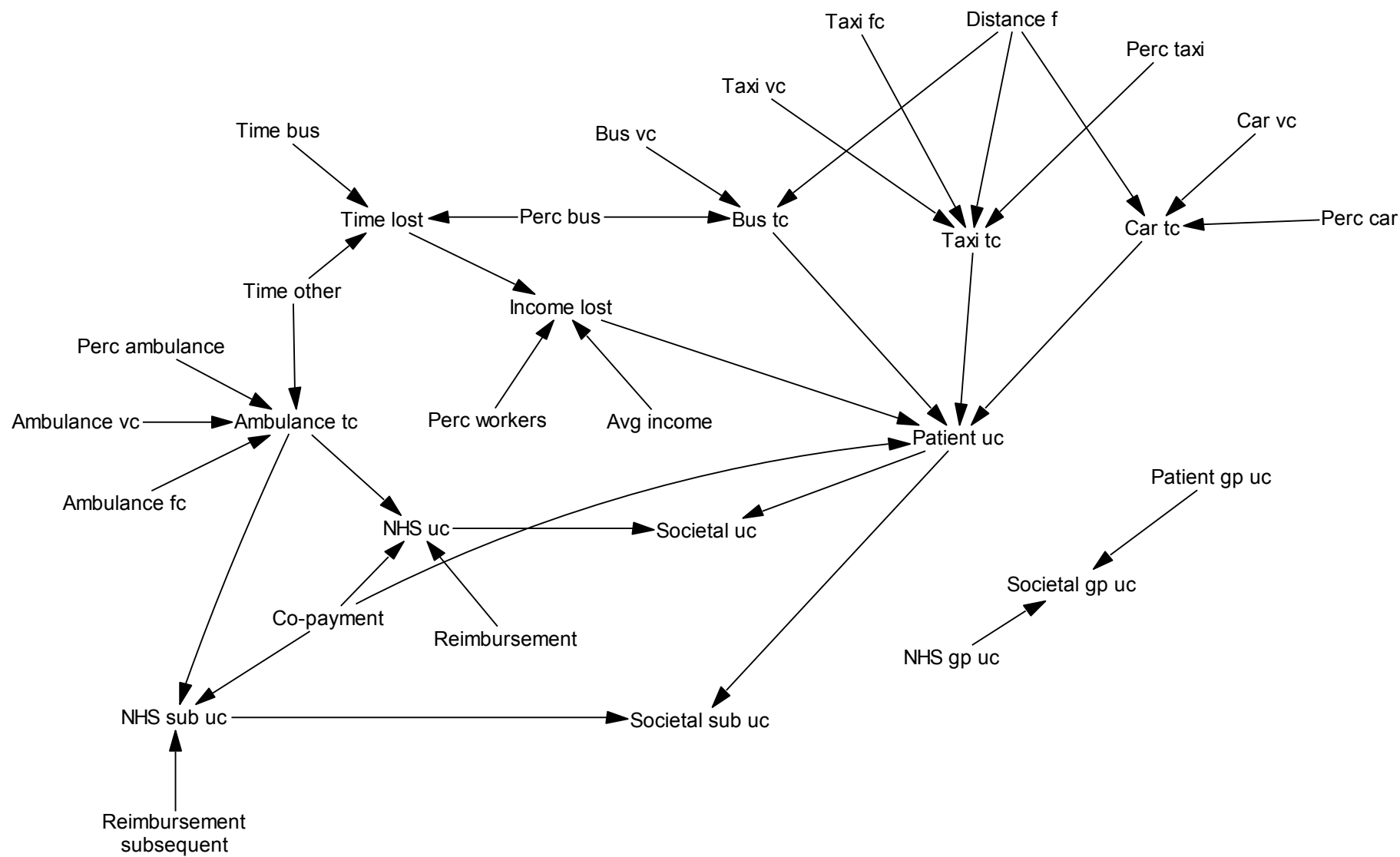


Figure C-12 View 10: Unit costs face-to-face care

Ambulance fc (Euro/Patient)

= 18

Description: *Fixed costs of ambulance travel.*

Source: *Bombeiros voluntarios*

Ambulance tc (Euro/Patient)

= Perc ambulance*(Time other*Ambulance vc+Ambulance fc)

Description: *Averaged patient expenses associated with ambulance travel.*

Ambulance vc (Euro/(Hour*Patient))

= 12

Description: *Variable cost for ambulance travel.*

Source: *Bombeiros voluntarios*

Avg income (Euro/(Hour*Patient))

= 5.139

Description: *Average income per hour.*

Source: *Statistics Portugal*

Bus tc (Euro/Patient)

= Bus vc*Distance f*Perc bus

Description: *Averaged patient expenses associated with bus travel.*

Bus vc (Euro/Km)

= 0.11

Description: *Bus allowance per km.*

Source: *Legislation: Decreto-Lei 137/2010*

Car tc (Euro/Patient)

= Perc car*Car vc*Distance f

Description: *Averaged patient expenses associated with car travel.*

Car vc (Euro/Km)

= 0.36

Description: *Car allowance per km.*

Source: *Legislation: Decreto-Lei 137/2010*

Co-payment (Euro/Patient)

= 7.5

Description: *Co-payment for a specialist appointment.*

Source: *Legislation: Portaria n. 306-A/2011*

Distance f (Km/Patient)

= 59.54

Description: *Average distance to primary care practice.*

Source: *Survey*

Income lost (Euro/Patient)

= Avg income*Perc workers*Time lost

Description: *Averaged income lost per face-to-face patient.*

NHS gp uc (Euro/Patient)

= 4.08

Description: *Total cost per patient incurred by the NHS in face-to-face practices.*

NHS sub uc (Euro/Patient)

= Ambulance tc+Reimbursement subsequent-Co-payment

Description: *Total cost per patient per subsequent appointment incurred by the NHS.*

NHS uc (Euro/Patient)

= Ambulance tc+Reimbursement-Co-payment

Description: *Total cost per patient incurred by the NHS, face-to-face pathway.*

Patient gp uc (Euro/Patient)

= 3

Description: *Total cost per patient incurred by the patients in primary care.*

Patient uc (Euro/Patient)

= Bus tc+Car tc+Income lost+Taxi tc+Co-payment

Description: *Total cost per patient incurred by patients.*

Perc ambulance (dmnl)

= 0.11

Description: *Percentage of patients using an ambulance to get to appointments.*

Source: *Survey*

Perc bus (dmnl)

= 0.08

Description: *Percentage of patients taking a bus to appointments.*

Source: *Survey*

Perc car (dmnl)

= 0.8

Description: *Percentage of patients using car to get to appointments.*

Source: *Survey*

Perc taxi (dmnl)

= 0.01

Description: *Percentage of patients taking a taxi to appointments.*

Source: *Survey*

Perc workers (dmnl)

= 0.3

Description: *Percentage of patients that work.*

Source: *Survey*

Reimbursement (Euro/Patient)

= 83.79

Description: *Reimbursement for a face-to-face appointment.*

Source: *ACSS*

Reimbursement subsequent (Euro/Patient)

= 76.17

Description: *Reimbursement for subsequent appointment.*

Source: *ACSS*

Societal gp uc (Euro/Patient)

= NHS gp uc+Patient gp uc

Description: *Total cost per patient incurred by society in primary care.*

Societal sub uc (Euro/Patient)

= NHS sub uc+Patient uc

Description: *Total cost per patient per subsequent appointment incurred by society.*

Societal uc (Euro/Patient)

= NHS uc+Patient uc

Description: *Total cost per patient incurred by society.*

Taxi fc (Euro/Patient)

= 3.25*2

Description: *Fixed costs associated with travelling by taxi.*

Source: *Antral*

Taxi tc (Euro/Patient)

= Perc taxi*(Taxi fc+Taxi vc*Distance f)

Description: *Averaged patient expenses associated with taxi travel.*

Taxi vc (Euro/Km)

= 0.45

Description: *Taxi cost per km.*

Source: *Antral*

Time bus (Hour)

= 8

Description: *Time lost for patients taking a bus.*

Source: *Ferrovial do Alentejo*

Time lost (Hour)

= Time bus*Perc bus+Time other*(1-Perc bus)

Description: *Averaged total time lost per patient.*

Time other (Hour)

= 3

Description: *Total time lost in face-to-face.*

Source: *Survey*

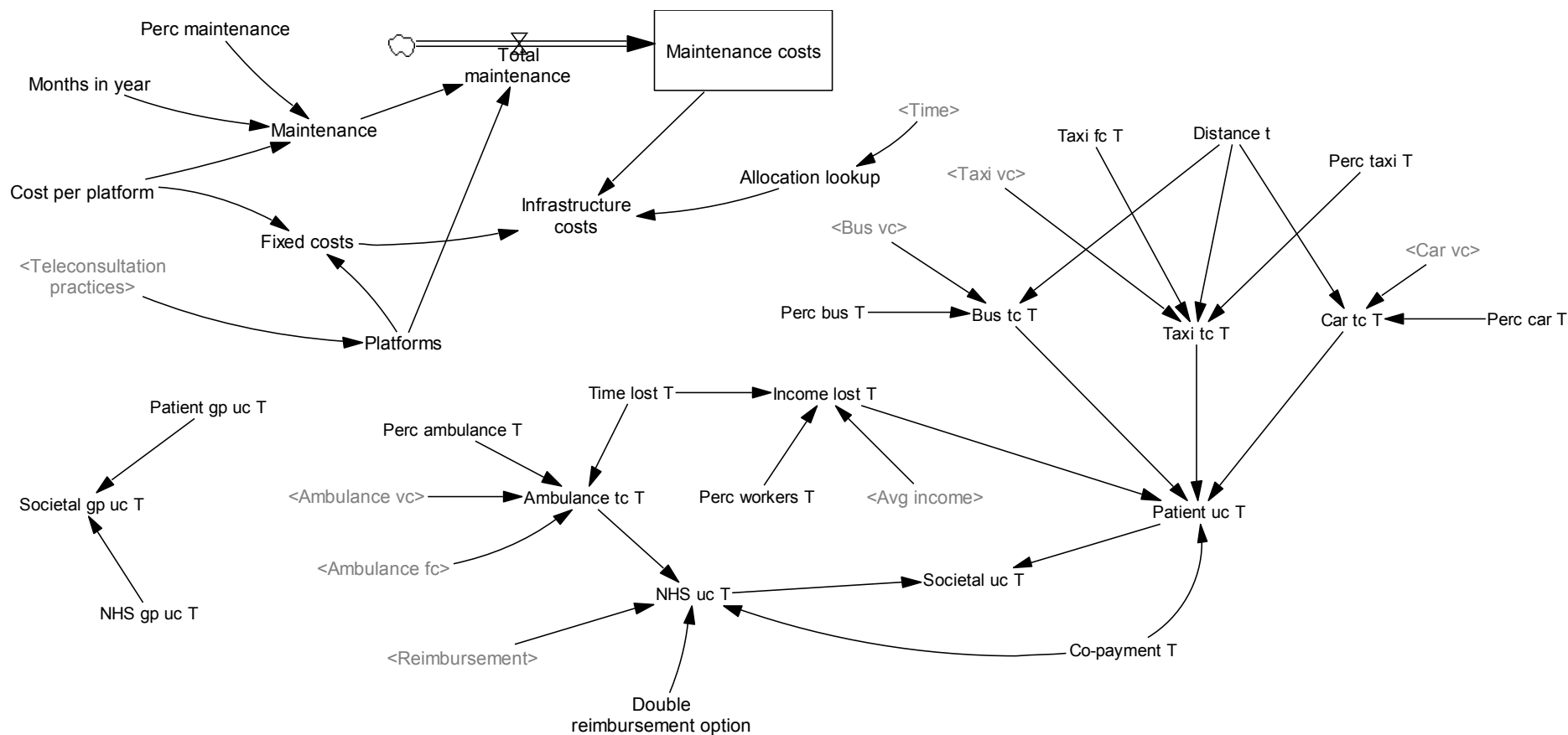


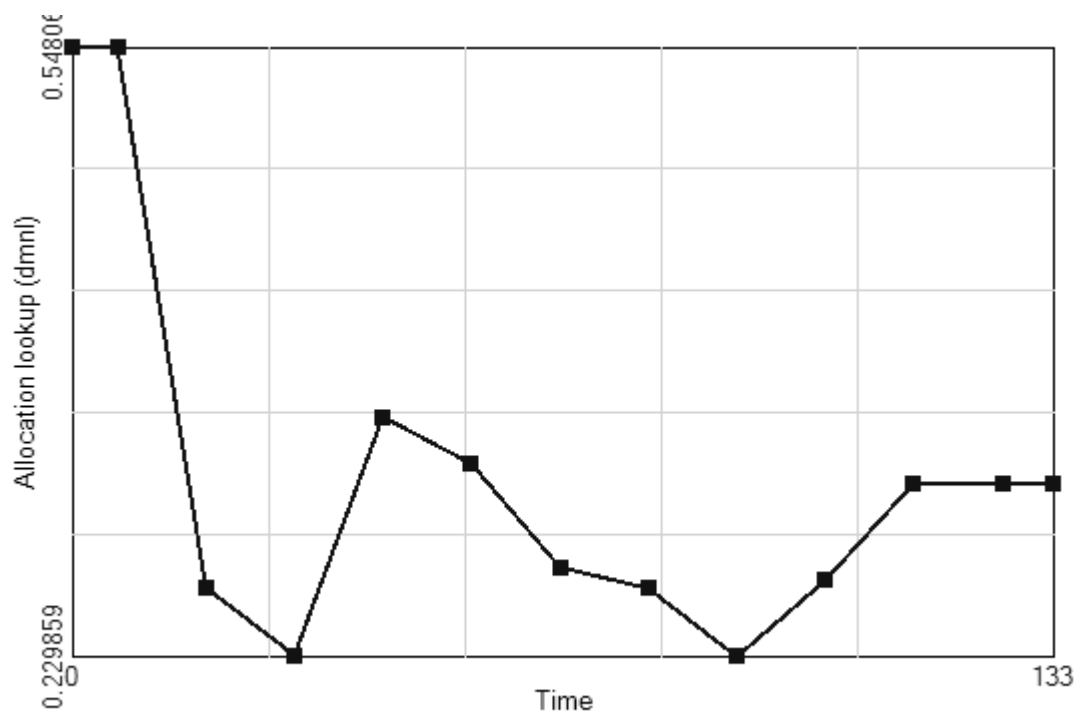
Figure C-13 View 11: Unit costs teleconsultation

Allocation lookup (dmnl)

= WITH LOOKUP (Time,([(0,0.229859)-
(133,0.548067)],(0,0.548067),(6,0.548067),(18,0.266042),(30,0.230682),(42,0.355389),(54,
0.330745),(66,0.276665),(78,0.265616),(90,0.229859),(102,0.269883),(114,0.320671),(126,
0.320671),(133,0.320671)))

Description: Allocation of equipment to specialty of interest as a function of simulated time.

Source: Calculated as the quotient of teleconsultations in dermatology and teleconsultations in all other specialties.

**Ambulance tc T (Euro/Patient)**

= Perc ambulance T*(Time lost T*Ambulance vc+Ambulance fc)

Description: Averaged patient expenses associated with ambulance travel.

Bus tc T (Euro/Patient)

= Bus vc*Distance t*Perc bus T

Description: Averaged patient expenses associated with bus travel.

Car tc T (Euro/Patient)

= Perc car T*Car vc*Distance t

Description: Averaged patient expenses associated with car travel.

Co-payment T (Euro/Patient)

$$= 3 \cdot 0 + 7.5 \cdot 1$$

Description: *Co-payment for a specialist appointment.*

Source: *Legislation: Portaria n. 306-A/2011*

Cost per platform (Euro/Platform)

$$= 9723.08$$

Description: *Cost of a video-conferencing platform.*

Source: *ARS Alentejo, I.P.*

Distance t (Km/Patient)

$$= 5.87$$

Description: *Average distance to primary care practice.*

Source: *Survey*

Double reimbursement option (dmnl)

$$= 2$$

Description: *If double reimbursement is allowed (i.e., both practice and hospital get reimbursed) then equals 2, otherwise 1.*

Fixed costs (Euro)

$$= \text{Cost per platform} \cdot \text{Platforms}$$

Description: *Total fixed costs of teleconsultation equipment.*

Income lost T (Euro/Patient)

$$= \text{Avg income} \cdot \text{Perc workers T} \cdot \text{Time lost T}$$

Description: *Averaged income lost per teleconsultation patient.*

Infrastructure costs (Euro)

$$= (\text{Fixed costs} + \text{Maintenance costs}) \cdot \text{Allocation lookup} \cdot 0 + (\text{Fixed costs} + \text{Maintenance costs}) \cdot 0.3$$

Description: *Total costs associated with teleconsultation equipment.*

Maintenance (Euro/(Month*Platform))

$$= \text{Cost per platform} \cdot \text{Perc maintenance} / \text{Months in year}$$

Description: *Maintenance expenses per unit of equipment.*

Maintenance costs (Euro)

$$= \int \text{Total maintenance } dt + [0]$$

Description: *Cumulative costs with video-conferencing platforms.*

Months in year (Months/Year)

= 12

Description: *Months in one year.*

NHS gp uc T (Euro/Patient)

= 4.08

Description: *Total cost per patient incurred by the NHS in teleconsultation practices.*

NHS uc T (Euro/Patient)

= Ambulance tc T+Reimbursement*Double reimbursement option-Co-payment T

Description: *Total cost per patient incurred by the NHS, teleconsultation pathway.*

Patient gp uc T (Euro/Patient)

= 3

Description: *Total cost per patient incurred by the patients in primary care.*

Patient uc T (Euro/Patient)

= Bus tc T+Car tc T+Income lost T+Taxi tc T+Co-payment T

Description: *Total cost per patient incurred by patients.*

Perc ambulance T (dmnl)

= 0.02

Description: *Percentage of patients using an ambulance to get to appointments.*

Source: *Survey*

Perc bus T (dmnl)

= 0.02

Description: *Percentage of patients taking a bus to appointments.*

Source: *Survey*

Perc car T (dmnl)

= 0.57

Description: *Percentage of patients using car to get to appointments.*

Source: *Survey*

Perc maintenance (dmnl/Year)

= 0.1

Description: *Maintenance costs as a percentage of initial cost of platform.*

Perc taxi T (dmnl)

= 0

Description: *Percentage of patients taking a taxi to appointments.*

Source: *Survey*

Perc workers T (dmnl)

= 0.3

Description: *Percentage of patients that work.*

Source: *Survey*

Platforms (Platform)

= IF THEN ELSE(Teleconsultation practices=0, 0, 1+Teleconsultation practices)

Description: *Number of video-conferencing platforms in use (equal to the number of practices plus on for the hospital).*

Societal gp uc T (Euro/Patient)

= NHS gp uc T+Patient gp uc T

Description: *Total cost per patient incurred by society in primary care.*

Societal uc T (Euro/Patient)

= NHS uc T+Patient uc T

Description: *Total cost per patient incurred by society.*

Taxi fc T (Euro/Patient)

= 2*2

Description: *Fixed costs associated with travelling by taxi.*

Source: *Antral*

Taxi tc T (Euro/Patient)

= Perc taxi T*(Taxi fc T+Taxi vc*Distance t)

Description: *Averaged patient expenses associated with taxi travel.*

Time lost T (Hour)

= 1.5

Description: *Averaged total time lost per patient.*

Source: *Survey*

Total maintenance (Euro/Month)

= Maintenance*Platforms

Description: *Monthly maintenance costs of teleconsultation platforms.*

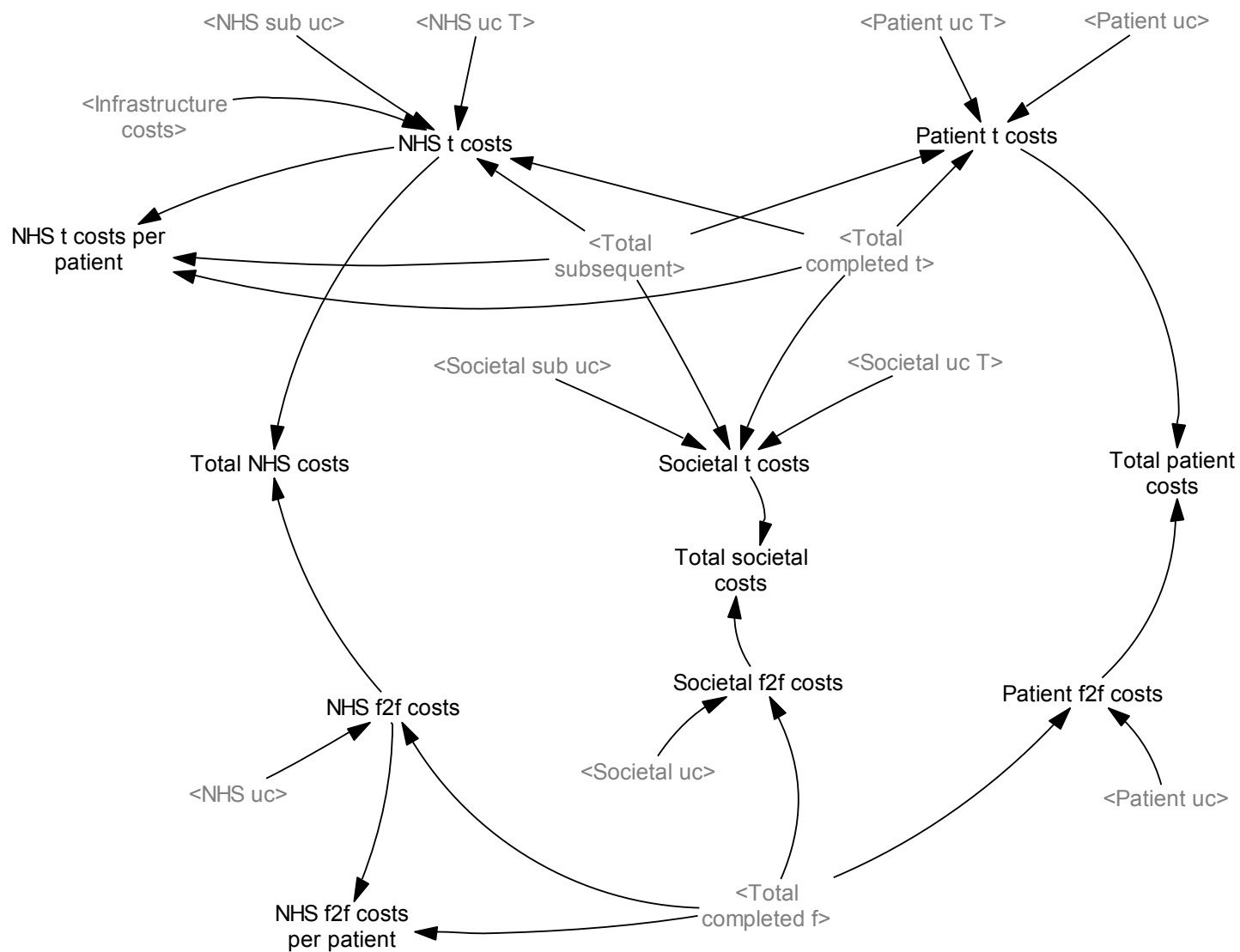


Figure C-14 View 12: Total costs and savings

NHS f2f costs (Euro)

$$= \text{NHS uc} * \text{Total completed f}$$

Description: *Costs to the NHS of face-to-face care.*

NHS f2f costs per patient (Euro/Patients)

$$= \text{ZIDZ}(\text{NHS f2f costs}, \text{Total completed f})$$

Description: *Costs to the NHS of each face-to-face patient.*

NHS t costs (Euro)

$$= \text{Infrastructure costs} + \text{NHS sub uc} * \text{Total subsequent} + \text{Total completed t} * \text{NHS uc T}$$

Description: *Costs to the NHS of teleconsultation care.*

NHS t costs per patient (Euro/Patients)

$$= \text{ZIDZ}(\text{NHS t costs}, \text{Total completed t} + \text{Total subsequent})$$

Description: *Costs to the NHS of each teleconsultation patient (including equipment costs).*

Patient f2f costs (Euro)

$$= \text{Patient uc} * \text{Total completed f}$$

Description: *Costs to patients of face-to-face care.*

Patient t costs (Euro)

$$= \text{Patient uc} * \text{Total completed t} + \text{Patient uc T} * \text{Total subsequent}$$

Description: *Costs to patients of teleconsultation care.*

Societal f2f costs (Euro)

$$= \text{Societal uc} * \text{Total completed f}$$

Description: *Costs to society of face-to-face care.*

Societal t costs (Euro)

$$= \text{Societal sub uc} * \text{Total subsequent} + \text{Societal uc T} * \text{Total completed t}$$

Description: *Costs to society of teleconsultation care.*

Total NHS costs (Euro)

$$= \text{NHS f2f costs} + \text{NHS t costs}$$

Description: *Costs to the NHS of all dermatological care.*

Total patient costs (Euro)

$$= \text{Patient t costs} + \text{Patient f2f costs}$$

Description: *Costs to patients of all dermatological care.*

Total societal costs

$$= \text{Societal f2f costs} + \text{Societal t costs}$$

Description: *Costs to society of all dermatological care.*

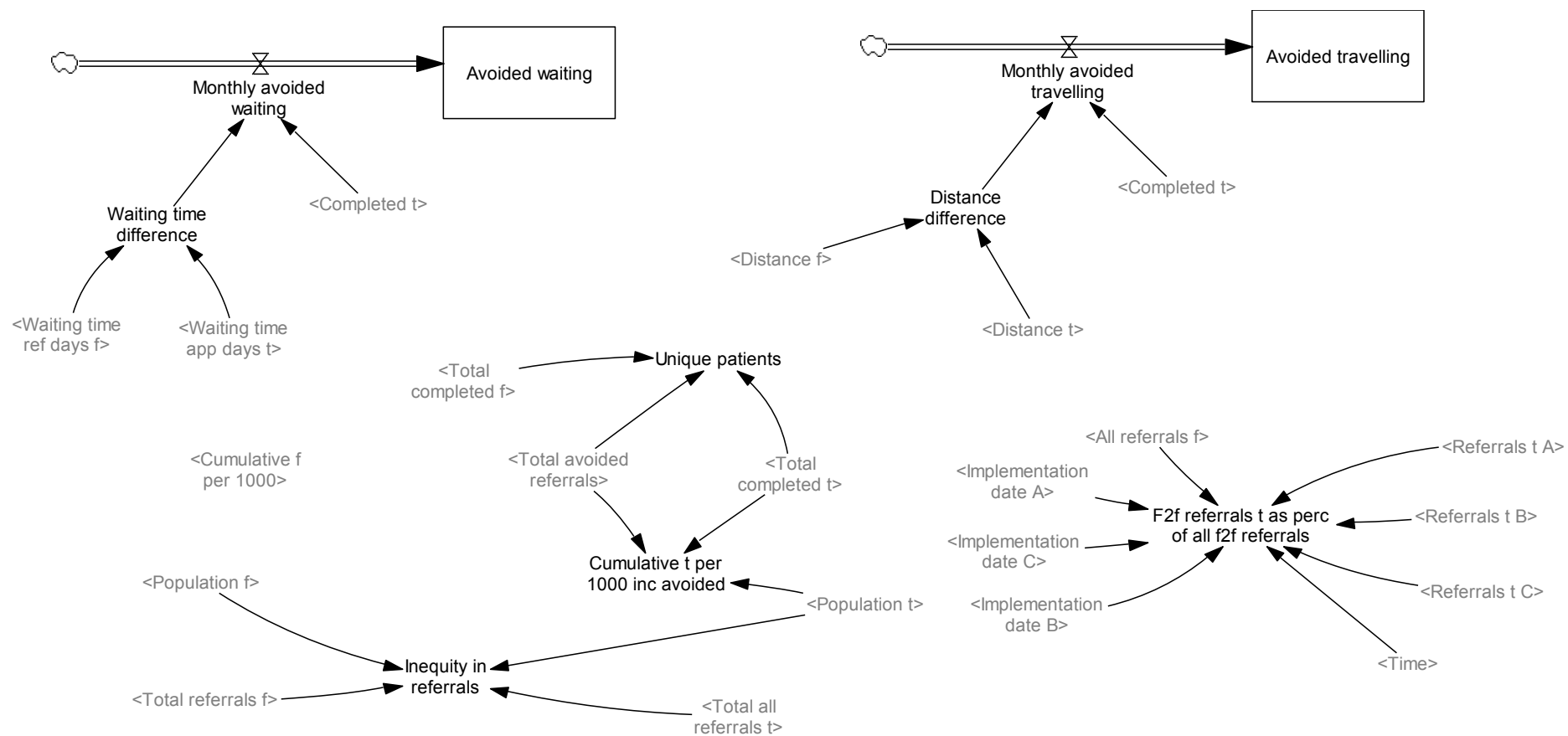


Figure C-15 View 13: Consequences

All referrals f (Patients/Month)

= F2f referrals t agg+Referrals f

Description: *All referrals to face-to-face appointments (from both teleconsultation and face-to-face practices).*

Avoided travelling (Km)

= Monthly avoided travelling dt + [0]

Description: *Travelling distance avoided.*

Avoided waiting (Patient*Day)

= Monthly avoided waiting dt + [0]

Description: *Cumulative patient days of waiting time avoided.*

Cumulative t per 1000 inc avoided (Patients/People)

= (Total avoided referrals+Total completed t)/(Population t/1000)

Description: *Total teleconsultations per 1000 inhabitants, including avoided referrals.*

Distance difference (Km/Patient)

= Distance f-Distance t

Description: *Difference between average travelling distance to face-to-face appointments and average travelling distance to teleconsultations.*

F2f referrals t as perc of all f2f referrals (dmnl)

= IF THEN ELSE(Time>=Implementation date A, IF THEN ELSE(Time>=Implementation date B, IF THEN ELSE(Time>=Implementation date C, (Referrals t A+Referrals t C+Referrals t B)/All referrals f, (Referrals t A+Referrals t B)/All referrals f), (Referrals t A)/All referrals f), 0)

Description: *Ratio of referrals from all teleconsultation practices to face-to-face appointments as a percentage of total referrals.*

Inequity in referrals (dmnl)

= ZIDZ((Total all referrals t/(Population t/1000)),(Total referrals f/(Population f/1000)))

Description: *Ratio of referrals per 1000 inhabitants from face-to-face practices and referrals per 1000 inhabitants from teleconsultation practices.*

Monthly avoided travelling (Km/Month)

= Completed t*Distance difference

Description: *Travelling distances avoided per week.*

Monthly avoided waiting (Days*Patient/Month)

= Completed t*Waiting time difference

Description: *Patient days of waiting time avoided per month.*

Unique patients (Patients)

= Total avoided referrals+Total completed f+Total completed t

Description: *Total patients seen.*

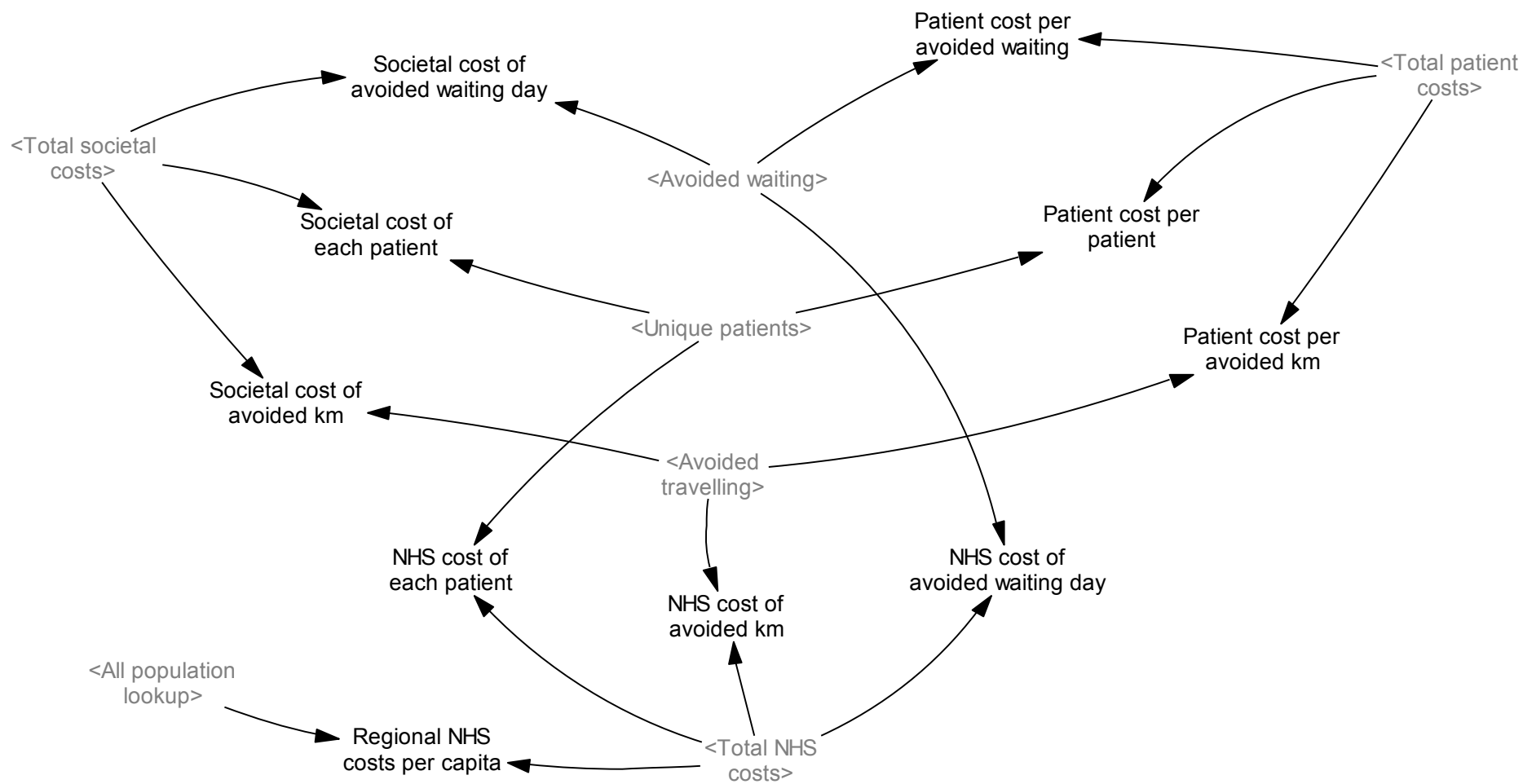


Figure C-16 View 14: Cost per consequence ratios

NHS cost of avoided km (Euro/Km)

= ZIDZ(Total NHS costs, Avoided travelling)

Description: *Cost to the NHS of each kilometre of travelling avoided.*

NHS cost of avoided waiting day (Euro/(Patients*Day))

= ZIDZ(Total NHS costs, Avoided waiting)

Description: *Cost to the NHS of each patient waiting day avoided.*

NHS cost of each patient (Euro/Patients)

= ZIDZ(Total NHS costs, Unique patients)

Description: *Cost to the NHS of each extra patient.*

Patient cost per avoided km (Euro/Km)

= ZIDZ(Total patient costs, Avoided travelling)

Description: *Costs to patients of each kilometre of travelling avoided.*

Patient cost per avoided waiting (Euro/(Patients*Day))

= ZIDZ(Total patient costs, Avoided waiting)

Description: *Costs to patients of each patient waiting day avoided.*

Patient cost per patient (Euro/Patients)

= ZIDZ(Total patient costs, Unique patients)

Description: *Costs to patients of each extra patient seen in teleconsultation.*

Regional NHS costs per capita (Euro/People)

= Total NHS costs/All population lookup

Description: *Costs to the NHS per Alentjo inhabitant.*

Societal cost of avoided km (Euro/Km)

= ZIDZ(Total societal costs, Avoided travelling)

Description: *Costs to society of each kilometre of travelling avoided.*

Societal cost of avoided waiting day (Euro/(Patients*Day))

= ZIDZ(Total societal costs, Avoided waiting)

Description: *Costs to society of each patient waiting day avoided.*

Societal cost of each patient (Euro/Patients)

= ZIDZ(Total societal costs, Unique patients)

Description: *Costs to society of each extra patient seen in teleconsultation.*