Maastricht University

Optimising delays in access to specialist outpatient clinics

Citation for published version (APA):

Rouppe van der Voort, M. B. V. (2014). *Optimising delays in access to specialist outpatient clinics*. [Doctoral Thesis, Maastricht University]. Maastricht University. https://doi.org/10.26481/dis.20140612mr

Document status and date:
Published: 01/01/2014

DOI:
10.26481/dis.20140612mr

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

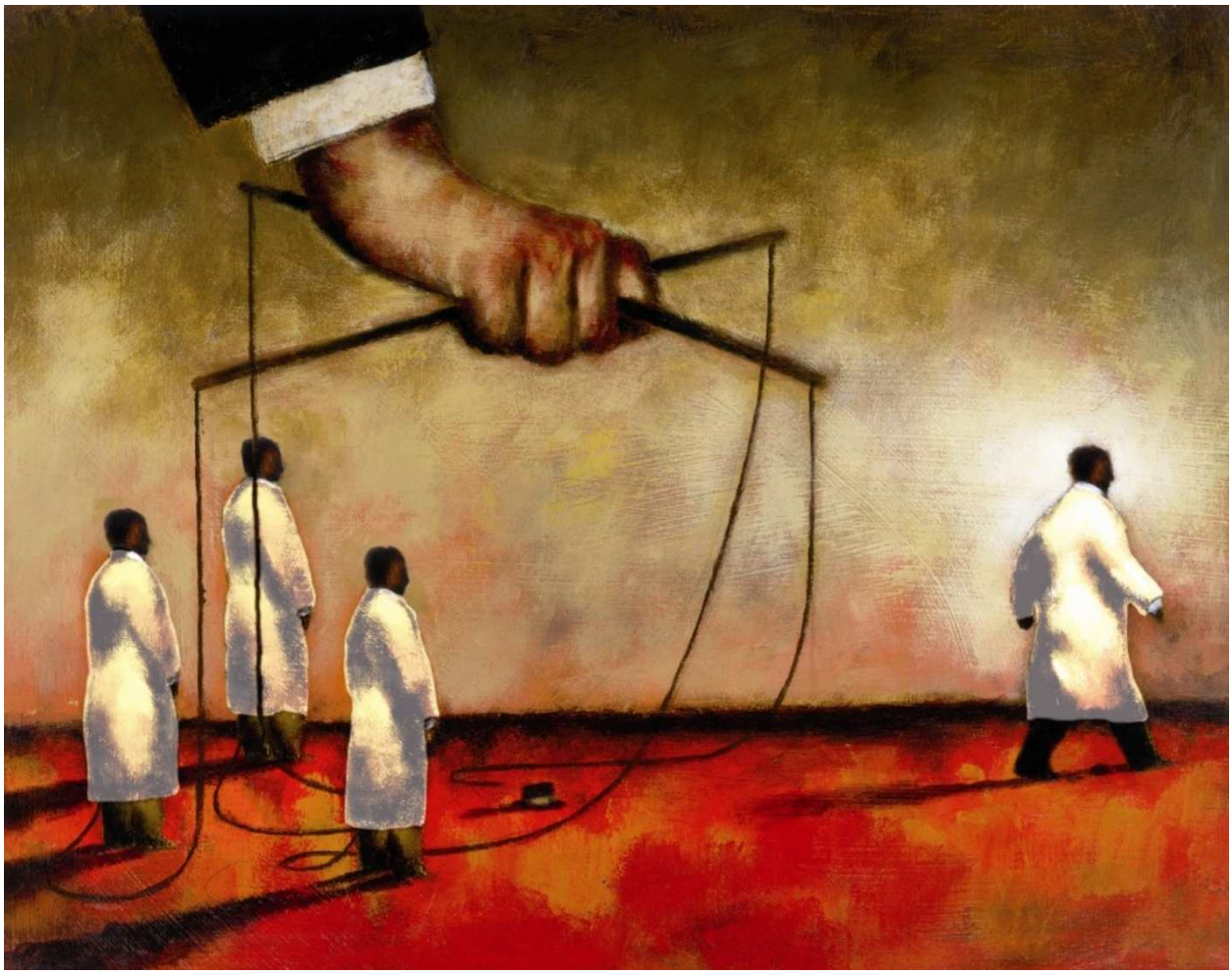
Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 20 Apr. 2024

Optimising delays in access to specialist outpatient clinics



Marc Rouppe van der Voort

The research for this thesis was performed at the department of Health Services Research, within the School for Public Health and Primary Care (CAPHRI), Faculty of Health, Medicine and Life Sciences, Maastricht University.

Cover illustration: Hollandse Hoogte, Corbis, Zefa
Printing: Gildeprint Drukkerijen
ISBN/EAN 978-94-6108-695-2

© M.B.V. Rouppe van der Voort, Tilburg 2014

Optimising delays in access to specialist outpatient clinics

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,

op gezag van Rector Magnificus, Prof. Dr. L.L.G. Soete,

volgens het besluit van het College van Decanen,

in het openbaar te verdedigen

op donderdag 12 juni 2014 om 14:00 uur

door

Marc Boudewijn Victor Rouppe van der Voort

Promotores

Prof. dr. G.G. van Merode
Prof. dr. H.J.J.M. Berden (Universiteit van Tilburg)

Beoordelingscommissie

Prof. dr. J.A.M. Maarse (Voorzitter)
Prof. dr. P.J.M. Bakker (Universiteit van Amsterdam)
Prof. dr. S.M.A.A. Evers
Prof. dr. W.N.J. Groot
Prof. dr. ing. T.W. Hardjono (Erasmus Universiteit Rotterdam)

Financieel bijgedragen aan de totstandkoming van dit proefschrift

St. Elisabeth Ziekenhuis in Tilburg, financiering van de aanstelling van de promovendus gedurende de eerste zeven jaar van het onderzoek.

St. Antonius Ziekenhuis in Utrecht/Nieuwegein, financiering van de aanstelling van de promovendus gedurende het laatste jaar van het onderzoek.

"We can't solve problems

by using the same kind of thinking

that was used when we created them"

Albert Einstein

BRIEF CONTENTS

Chapter 1 Introduction 13

Part I THEORY

Chapter 2 Theoretical framework 23

Chapter 3 Literature review: what is known about delays in access to specialist outpatient clinics? 53

Chapter 4 Literature review: what is known about improving processes in hospitals with the 'lean'-philosophy? 73
Ned Tijdschr Geneeskd. 2013;157:A5541

Part II RESEARCH

Chapter 5 Effectiveness of Working Without a Waiting List 83
Health Policy, August 2010;97;44-52

Chapter 6 Sustainability of Working Without a Waiting List 95
Journal of Health Services Research & Policy, 2014, Vol 19(2) 94-101

Chapter 7 How do specialist outpatient clinics match supply and demand? 107

Part III SYNTHESIS AND DISCUSSION

Chapter 8 Production Planning and Control framework and typology for specialist outpatient clinics 137
Submitted

Chapter 9 General discussion 155

Summary 172

Samenvatting 180

Dankwoord 188

Curriculum Vitae 190

Publications 191

C O N T E N T S

Chapter 1 Introduction **13**

1. Introduction 14
2. Problem statement 16
3. Goals and research questions 16
4. Thesis outline 17
5. Dutch context of specialist outpatient clinics 18

References 18

Part I THEORY **21**

Chapter 2 Theoretical framework **23**

1. Introduction 24
2. The science of waiting 24
   - 2.1 Introduction 24
   - 2.2 Queuing Theory 24
   - 2.3 Causes of delays 25
   - 2.4 Managing uncertainty 27
   - 2.5 Managing complexity 28
   - 2.6 Single and shared queues 29
   - 2.7 The role of information 30
   - 2.8 Optimisation questions 31
3. Pull 32
   - 3.1 Introduction 32
   - 3.2 What preceded pull: MRP, MRP II and ERP 32
   - 3.3 General definition and application of pull 33
   - 3.4 Strategic and tactical level of pull 34
   - 3.5 Example of a pull system 35
   - 3.6 Benefits of pull 37
   - 3.7 Related concepts 38
   - 3.8 Cycle mix planning 38
   - 3.9 Combination of push and pull 39
   - 3.10 First and second order change 40
4. The science of waiting and pull applied to specialist outpatient clinics 40
   - 4.1 Introduction 40
   - 4.2 Characteristics of specialist outpatient clinics 40
   - 4.3 Uncertainty, variability and buffering 42
   - 4.4 Definition of pull for specialist outpatient clinics 44
   - 4.5 Cycle-mix planning in specialist outpatient clinics 45
   - 4.6 Potential benefits of pull 45
   - 4.7 The larger hospital system 46
5. Production Planning and Control framework 47

5.1 Introduction 47
5.2 A pull based Production Planning and Control (PPC) framework 47
5.3 Variables for specialist outpatient clinics 47
5.4 A PPC framework for specialist outpatient clinics 49
References 50

Chapter 3 Literature review: optimising delays in access to specialist outpatient clinics 53
1. Introduction 54
2. Methods 54
3. Results 54
3.1 Introduction 54
3.2 What is known about the extent of the problem of delays? 54
3.3 What is known about how outpatient clinics match supply and demand? 56
3.4 What is known about the causes of delays in access? 56
3.5 What is known about possible interventions to optimise delays in access? 57
3.6 What is known about the effects of interventions to optimise delays? 60
4. Conclusions 64
References 65

Chapter 4 Literature review: improving processes in hospitals with the 'lean'-philosophy 73
Ned Tijdschr Geneeskd. 2013;157:A5541
Abstract 74
1. Introduction 74
2. What is the lean-philosophy? 74
3. Application of the lean-philosophy 75
4. Is the lean-philosophy effective? 77
5. Reflection 78
6. Conclusion 79
References 79

Part II RESEARCH 81

Chapter 5 Effectiveness of Working Without a Waiting List 83
Health Policy, August 2010;97;44-52
Abstract 84
1. Introduction 84
2. Theoretical framework 84
3. The application of Advanced Access 85
4. Materials and methods 85
5. Results 86
5.1 Applied interventions 86

5.2 Effects on access 87
5.3 Relation between interventions and effects on access 89
5.4 Other effects 89
6. Discussion 90
7. Conclusion 92
References 92

Chapter 6 Sustainability of Working Without a Waiting List 95
Journal of Health Services Research & Policy, 2014, Vol 19(2) 94-101
Abstract 95
1. Background 96
2. Methods 96
3. Results 97
3.1 Access three years after the project 97
3.2 Sustaining work practices or continuous improvement 99
3.3 Increased responsiveness to better match supply and demand 99
3.4 Clinical leadership and incentives 101
3.5 A shared belief that they can and should control access together 102
4. Discussion 103
4.1 Main findings 103
4.2 Limitations 104
4.3 Recommendations 104
References 104

Chapter 7 How specialist outpatient clinics match supply and demand 107
1. Introduction 108
2. Methods 109
2.1 Research design 109
2.2 Data collection methods 109
2.3 Data analysis 110
3. Results 110
3.1 Introduction 110
3.2 General characteristics of the studied specialist outpatient clinics 110
3.3 Characteristics of demand 111
3.4 Characteristics of supply 111
3.5 The planning of sessions 112
3.6 The planning of appointments 113
3.7 Reconstruction how the way demand and supply are matched evolved 115
3.8 Perspectives of the involved actors on optimal access 116
3.9 System visualisation how demand and supply are matched 117
4. Reflection on how demand and supply are matched 118
4.1 The aggregate level: the planning of sessions 118

4.2 The operational level: the planning of appointments 120
4.3 Push and pull methods: what triggers the release of capacity? 121
4.4 How the matching of demand and supply evolved over time 121
4.5 General observations 121
5. Variables to optimise delays in access to specialist outpatient clinics 122
6. What drives the decisions how supply and demand are matched? 124
7. Conclusion 126
References 128

Part III SYNTHESIS AND DISCUSSION 135

Chapter 8 Production Planning and Control framework and typology for specialist outpatient clinics 137
Submitted
1. Introduction 138
2. Identified causes of delays 138
3. Identified methods to address the causes of delays 140
4. Production Planning and Control framework for specialist outpatient clinics 142
4.1 Introduction 142
4.2 Strategic level 142
4.3 Tactical level 144
4.4 Control level 145
4.5 The PPC framework and drivers how supply and demand are matched 146
5. Push-pull typology for specialist outpatient clinics 146
6. Push-pull analysis of the WWWL results 147
7. Discussion 150
7.1 Introduction 150
7.2 Reflection on the hypotheses 150
7.3 Comparison of the applied WWWL methods and the PPC framework 151
7.4 Push-pull applications 152
7.5 The drivers to optimise the way demand and supply are matched 153
7.6 How to determine the optimal delay in access 153
References 154

Chapter 9 General discussion 155
1. Introduction 156
2. Part I – Theory 156
3. Part II – Research 157
3.1 Introduction 157
3.2 The context of the problem of delays in access 157
3.3 The causes of delays in access 158
3.4 What is required to optimise delays in access? 159
3.5 The effectiveness and sustainability of WWWL 159

3.6 Methodological considerations 163
4. Part III – synthesis 164
5. Conclusion 166
5.1 Introduction 166
5.2 Applicability of our findings 166
5.3 Implications for practice 167
5.4 Recommendations for further research 168
5.5 Closing remarks 169
References 169

Summary 172

Samenvatting 180

Dankwoord 188

Curriculum Vitae 190

Publications 191

CHAPTER 1

Introduction

1. Introduction

Accessibility to care providers is a vital aspect of the quality of care.[1] In access to outpatient specialty care, however, delays have been a common problem for many years.[2] When we started our research in 2005 25% of Dutch specialist outpatient clinics could not offer an appointment within 28 days for any of their specialists.[3] In the 2009 European Health Consumer Index the Netherlands score very well on all sub-disciplines, except 'Waiting times/Access for specialist treatment', where the score is mediocre.[4] This is not unique for the Netherlands: in Canada for example the median delay in access across all specialties was also 28 days.[5] It can be easy to consider logistics as unimportant in a context where health, uncertainty, trust, empathy, and caring for other's wellbeing are central. But this might very well be the most important context there is to ensure the best possible logistics. Delays in access to care should not needlessly add burden to already sick, discomforted and anxious people.

Waiting is a common phenomenon in health care. On a daily basis approximately 50.000 people make an appointment at an outpatient speciality clinic in one of the 88 Dutch hospitals for which they will often need to wait 28 days or more.[6] This implies that on any given day more than 1.5 million people of the population of 16 million in the Netherlands are waiting for an appointment. In a year the population waits more than 500 million days for their appointments in specialist outpatient clinics. When in health care systems queues became too long, policy makers usually tried to solve this problem by allocating additional resources, discouraging demand or forcing health providers to treat more patients with the same resources.[7] Changes in organization and logistics potentially provide alternative solutions, but until recently they did not play a major role in discussions. Waiting was commonly considered as an unavoidable phenomenon in health care due to the constraints on capacity.[8]

In the beginning of this century the awareness grew that change in organisation and logistics might provide new answers to waiting lists and delays. An editor of the British Medical Journal for example noted in 2001 that to improve access 'Health care leaders are now recognising, as car manufacturers did before, that the health care system needs radically redesigning'.[9] This thesis investigates whether and how delays in access to specialist outpatient clinics can be optimised with changes in organisation and logistics.

Reducing delays in industry

To apply logistical concepts, health care might benefit from industry because industry has a long tradition of improving planning and logistics. In order to optimise delays while enabling efficient production, the concept of 'pull logistics' has been extensively studied and is well recognized for its effectiveness.[10,11] Strategically, the ultimate aim of a pull system is that demand determines what is produced (customer demand 'pulls' production from the supplier) versus a push system where production is scheduled by other criteria then actual demand, e.g. forecasts of demand, and is pushed onto the market.[12] When applied in production, however, pull methods not only aim to let the customer pull, but also need to optimise the efficiency of the production system. That is why, tactically, a pull system authorises the release of work based on the actual system status.[13] We shall formally define and explain the concept of pull in the theoretical framework in chapter 2.

Push and pull methods are researched and described in the field of Operations Management (OM).

OM is concerned with the design, operations, and improvement of production systems.[14] There is a growing interest from health care in the translation and application of methods and knowledge from OM, often with the objective in mind to deliver patient centred care by designing systems that create continuous patient flow without delays.[15]

Knowledge from research in industry applied to specialist outpatient clinics

There is an extensive body of knowledge on the general science of waiting in the field of OM that will be used in this thesis. There has only been sporadic research on applying knowledge from the field of OM in outpatient care, and almost no research specifically on specialist outpatient clinics. The majority is theoretical research on specific planning techniques to optimise waiting times in the waiting room of outpatient clinics. See the literature review in chapter 3 for a complete overview.

An exception is research on the application of the Advanced Access model. Advanced Access is developed by general practitioner Mark Murray in the late '90's and is first published in 2000. To a certain extent, Advanced Access translated the pull method into a practical application to improve the organisation of outpatient clinics for family practices and general practitioners. The aim of the Advanced Access model is to 'do today's work today'.[16] The basic premise is that this can be achieved by improving the way demand and supply are matched (or: synchronized) and by increasing the efficiency in the way supply is delivered without adding capacity.[16-18] The validity of the Advanced Access model has been established in several small and one large study for family practices and general practitioners in the US and the UK, whereby delays reduced remarkably with 50% or 60% on average.[19-21] However, the studies did not validate the sustainability of the results over the years after the application of the model, nor did they compare the model to the theory on waiting and the method of pull.

There is thus research available on the application of Advanced Access in Primary Care, but there is no research available on Advanced Access or any other approach on the application of OM techniques in specialist outpatient clinics to optimise delays, which is more complex than family practices or general practices. Furthermore, little is known on the specific problem and causes of delays for neither Primary Care nor specialist outpatient clinics. Only limited research is available on the translation and application of knowledge from the field of OM in general and pull methods in particular to health care and no research specifically on the application of pull methods in specialist outpatient clinics.

In 2002 the model of Advanced Access was adapted to specialist outpatient clinics and further developed with knowledge from OM. This was introduced in the Netherlands as 'Working without a waiting list (WWWL)' (in Dutch: 'Werken zonder wachtlijst (WZW)'). Specialist outpatient clinics applied the WWWL model by participating in Breakthrough Collaboratives organised by the Dutch Institute for Health care Improvement CBO. The aim of the Breakthrough collaboratives WWWL is to improve delays in access to Dutch specialist outpatient clinics by creating a learning and stimulating environment for specialist outpatient clinics and spread the knowledge throughout the Netherlands. Since 2001 over 200 specialist outpatient clinics have participated in the Breakthrough collaboratives in groups of 5 to 30 outpatient clinics in each collaborative.

Soon after the start of the first collaborative remarkable results in improving access was reported.[22-26] These claims have, however, not been validated with scientific studies. Did the specialist outpatient clinics substantially reduce delays in access? Were the results sustainable over a period of years? How did they achieve their results? Was it a successful application of pull methods?

2. Problem statement

With this thesis we aim to increase the body of knowledge on the science of waiting optimisation in specialist outpatient clinics specifically, and for health systems in general. The problem of delays to specialist outpatient clinics, the knowledge on waiting, the established effectiveness of applying pull methods in industry to optimise delays and the success stories surrounding Advanced Access and WWWL bring us to the following problem statement and research questions:

> *How can delays in access to specialist outpatient clinics be optimised by improving the way supply and demand are matched?*

The terms 'delay in access', supply', 'demand' and 'matched' refer to patients requesting an appointment (demand), specialist outpatient clinics providing appointments (supply) and the processes how demand and supply are matched. Exact definitions are provided the Theoretical Framework (chapter 2). 'Optimised' is chosen instead of 'minimised', because some level of delay access will be required in order to work efficiently.

3. Goals and research questions

We applied the knowledge on the science of waiting from the field of OM to the practice of specialist outpatient clinics to answer the problem statement. We needed to better understand the context of the problem, develop methods to analyse the problem and test whether methods are effective. We set two goals for our research to answer the problem statement.

1) Design a method to improve the way demand and supply are matched in specialist outpatient clinics to optimise delays

We developed a theoretical framework based on existing literature and tested it by explorative research on how specialist outpatient clinics match supply and demand. The research questions are:

1. How are demand and supply matched in specialist outpatient clinics?
2. What is required to optimise the way demand and supply are matched?

Research question 1 concerns the processes that match supply and demand, their relations, the information that is used and the decision rules that are applied. The outcome of question 1 is used to identify and understand the variables that need to be considered to optimise the way they match supply and demand, which we use to answer question 2.

2) Determine whether and why WWWL is an effective method to optimise delays in access to specialist outpatient clinics

We evaluated the short and long term effects of Working Without a Waiting List (WWWL) for 18 specialist outpatient clinics. We use these findings to determine the effectiveness of WWWL and to develop a framework to optimise delays. The research questions are:

1. What changed in the way the specialist outpatient clinics match supply and demand?
2. What are the results in reduction of delays in access?
3. Which changes are still present three years later and what new changes have taken place?
4. What is the delay in access three years later?
5. Which factors determined whether the improvements in delays in access were sustained according to the involved actors?

4. Thesis outline

This thesis consists of three parts. The theory from part I and the research of part II together enable us to design a method to optimise delays in access to specialist outpatient clinics in part III.

Part I THEORY

To design a method of improving the way demand and supply are matched, we developed a theoretical framework in chapter 2 that combines the knowledge on waiting with the general characteristics of specialist outpatient clinics. We described the science of waiting from the field of Operations Management (OM), focus on the application of pull methods and apply it to specialist outpatient clinics. In the final section we provide a general pull based Production Planning and Control framework and the variables that need to be considered to apply it.

We complemented the theoretical framework with a literature review in chapter 3 to describe what is known on the problem of delays in access to specialist outpatient clinics. Chapter 4 further complements the theoretical framework with a description of what is known about how processes are improved in health care by applying the 'lean'-philosophy. Lean is a comprehensive concept that is a source of WWWL and Advanced Access and includes the application of pull methods.

Part II RESEARCH

In chapter 5 we studied the interventions and results of the application of WWWL by 18 specialist outpatient clinics. In chapter 6 we continued this study with an explorative research on the sustainability of the results and the factors that the involved actors consider as determinants of sustainability.

In chapter 7 we explored in more detail the way specialist outpatient clinics match supply and demand. We investigated the characteristics of a specialist outpatient clinic that can match demand and supply with relatively little dependence on other departments, and a specialist outpatient clinic that has high dependency, in particular the planning of the operating theatre. We concluded with an overview of the variables that determine the mix of push or pull methods to optimise delays to specialist outpatient clinics and the conditions that need to be taken into account to optimise the way specialist outpatient clinics match supply and demand.

Part III SYNTHESIS AND DISCUSSION

The results from the theoretical framework and the research chapters were combined in chapter 8 to answer the problem statement. We not only provided an overview of the findings, but we also created synthesis by combining the theory of part I with the findings of part II to create a Production Planning and Control (PPC) framework for specialist outpatient clinics. This PPC framework is used to create a typology of the way specialist outpatient clinics match supply and demand with different combinations of push and pull methods. We used the push-pull typology to

analyse the results and sustainability of Working Without a Waiting List (WWWL) and determine the most effective typologies.

In chapter 9 we discussed the implications in general for the optimisation of delays in access to specialist outpatient clinics and to what extent we consider pull systems viable for this purpose. We finished with recommendations with implications for practice and for further research to optimise delays in health systems.

5. Dutch context of specialist outpatient clinics

Contrary to many other countries, in the Netherlands specialist outpatient clinics are part of the hospital and most are physically situated inside the hospital. The assistants are employed by the hospital. All specialists in this research are self-employed except for one academic specialist outpatient clinic. Each specialty has a contract with the hospital. A new appointment and a follow-up appointment twelve months later generate revenue for the hospital and a separate revenue for the specialists, paid by the insurance company.

Like most Dutch hospitals all involved hospitals are private not-for-profit organisations. In the past decade, the Dutch government gradually started to introduce market dynamics in the system, especially by shifting more buying power to the insurance companies while stimulating patients to choose between insurance companies and hospitals. Each Dutch person has by law a health insurance policy. The size of the specialist outpatient clinics range from three to twelve specialists in one specialist outpatient clinic, sometimes more. The size of the hospitals mostly range from approximately 130 to 1.350 beds and 30.000 to 250.000 first outpatient specialty appointments. The hospitals involved in this research range from 350 to 1.000 beds and 85.000 tot 180.000 first outpatient specialty appointments per year.

Traditionally the specialists in the Netherlands enjoy a large autonomy: medically as well as how they behave in the organisational processes. Most hospitals have a form of dual management making specialists co-responsible for the policies, management and organisation of care delivery.

References

1. Institute of Medicine: Crossing the Quality Chasm. A new health System for the 21st Century. Washington: National Academy Press, 2001.
2. Ministry of Health, Welfare and Sport. Waiting times in hospitals (in Dutch). May 2002. http://www.minvws.nl/rapporten/cz/wachttijden_in_ziekenhuizen.asp, accessed 31 Jan 2010.
3. National Institute for Public Health and the Environment (RIVM). How long is being waited on hospital care? (in Dutch) http://www.rivm.nl/vtv/object document/o7864n34418.html, accessed 31 January 2010.
4. Björnberg A, Garofé BC and Lindblad S. Euro Health Consumer Index 2009, Brussels: Health Consumer Powerhouse, 2009.
5. Ministry of Industry. Access to Health Care Services in Canada, January to December 2005. Ottawa: Statistics Canada; 2006.

6. National Institute of Public Health and the Environment. National Public Health Compass: number of first outpatient specialty visits in 2005 (in Dutch), http://www.nationaalkompas.nl/zorg/ziekenhuiszorg/gebruik-trend/#nwid_2, accessed September 5, 2013.
7. Hanning, M. Maximum waiting-time guarantee - an attempt to reduce waiting lists in Sweden. Health Policy 36 (1996), 17-35.
8. Vissers JMH, Van der Bij JDH and Husters RJ. Towards Decision Support for Waiting Lists: an Operations Management View. Health care Management Science 4; 2001, 133-142.
9. J. Smith. Redesigning health care. BMJ 2001;322;1257-8.
10. Hopp WJ, Spearman ML. Factory physics. New York: Irwin/McGraw-Hill, 2000.
11. Liker JK, The Toyota Way: 14 management principles from the world's greatest manufacturer. New York: McGraw-Hill, 2004.
12. Ohno T. Toyota Production System: beyond large scale production. Tokyo: Productivity, Inc.; 1988.
13. Hopp WJ, Spearman ML. To pull or not to pull: what is the question? Manufacturing & Service Operations Management; Spring 2004;6;2;133-48.
14. Krajewski IJ, Ritzman LP. Operations Management. 6th Edition ed. Upper Saddle River, NJ: Pretice Hall; 2002.
15. Walley P. Designing the accident and emergency system: lessons from manufacturing. Emerg Med J. 2003;20(2):126-30.
16. Murray M, Tantau C. Same-day appointments: exploding the access paradigm. Fam practice manag 2000:7:45-50.
17. Murray M, Berwick DM. Advanced Access: reducing waiting and delays in primary care. JAMA. 2003;289:1035-1040.
18. Oldham J. Advanced Access in Primary Care. Manchester: National Primary Care Development Team, 2001. http://www.acfp.ca/docs/advanced_Access.pdf (accessed 29 Jul 2008).
19. Pickin DM, O'Caithain A, Sampson FC, Dixon S. Evaluation of ' advanced access' in the National Primary Care Collaborative. Br J Gen Pract 2004;54:334-340.
20. Dixon S., Sampson FC, O'Caithain A, Pickin DM. Advanced Access: more then just GP waiting times? Familiy Practice 2006;23:233-239.
21. Murray M, Bodenheimer T, Rittenhouse D, Grumbach K. Improving timely access to primary care. Case studies of the Advanced Access model. JAMA 2003;289:1042-1046.
22. Rouppe van der Voort M, Janssen S, Minkman M, Slot S and Schellekens W. De balans opgemaakt: Doorbraak-aanpak succesvol om toegangstijden tot een week te bekorten, Medisch Contact; 2004; 14; 546-49.
23. Van Bodegom J, Rouppe van der Voort M and Van Merode F. Controle over vraag en aanbod, Medisch Contact; 2004; 12; 469-71.
24. Blaauw I, Ton E, Simons S, Raadgever B and Rouppe van der Voort M. Doel gemist, missie geslaagd, Medisch Contact; 2004; 13; 500-3.

25. Kuper I, Kruisen M, Sauerwald V and Rouppe van der Voort M. De winst van gedeelde zorg, Medisch Contact; 2004; 10; 370-72.
26. Blanken R, et al. Wachtlijst weggewerkt. Het succesverhaal van een dermatologische polikliniek. Medisch Contact; 2004; 9; 328-31.

Part I THEORY

In part I we discuss the available body of knowledge in relation to our problem statement how specialist outpatient clinics can optimise delays in access. In chapter 2 we describe the science of waiting and present a theoretical framework that enables us to investigate how delays in access to specialist outpatient clinics can be optimised. In chapter 3 we complement the theoretical framework with a literature review on what is known on the problem of delays in access to specialist outpatient clinics. Chapter 4 further complements the theoretical framework with a literature review on how processes are improved in health care by applying the 'lean'-philosophy, a comprehensive concept that has been a source of Working Without a Waiting List (WWWL) and Advanced Access, two approaches to optimise delays in access that are relevant for this research.

CHAPTER 2

Theoretical framework

1. Introduction

The problem statement of this research is how to optimise delays in access to specialist outpatient clinics. In this chapter we create a theoretical framework to study the phenomenon of delays in access. We use the body of knowledge on the science of waiting from the field of Operations Management (OM) because it has proves successful in industry to deal with these types of problems. Paragraph 2 describes the science of waiting. Paragraph 3 describes the method of pull, which has proven an effective OM method to optimise delays. Paragraph 4 discusses the application of the science of waiting and the application of pull in specialist outpatient clinics. Paragraph 5 presents a general pull based Production Planning and Control framework and discusses the applicability for specialist outpatient clinics.

2. The science of waiting

2.1 Introduction

Paragraph 2.2 explains the basics of Queuing Theory, which forms the basis of the science of waiting. Paragraph 2.3 analyses the general causes of delays from the perspective of Queuing Theory. Paragraph 2.4 discusses strategies how uncertainty can be managed. Paragraph 2.5 relates this to the complexity of queuing systems, 2.6 specifies it for shared and single queues and 2.7 elaborates on the role of information. Paragraph 2.8 concludes with optimisation questions for delays in access.

2.2 Queuing Theory

Applying the science of Queuing Theory requires that a specialist outpatient clinic is interpreted as a queuing system.[1] A system is a complex whole the functioning of which depends on its parts and the interaction between these parts.[1] Systems have the following characteristics:[2]

1. The properties or the behaviour of each entity of the collection has an effect on the properties or behaviour of the collection as a whole.
2. The properties or the behaviour of each entity, and the way they influence the whole, are at least dependant on the properties and behaviour of one other entity of the collection.
3. Each possible sub-collection of the entity is characterised by the two afore-mentioned properties. Therefore it is impossible to divide a system in independent parts.

Because of these characteristics, a system as a whole always has characteristics that none of its elements have: 'the whole is more than the sum of the parts'. Characteristics that can only be attributed to the whole and not to individual elements are called 'aggregate' properties. Delays can be an example of such an aggregate property.

A queuing system comprises of an arrival process, a service process and a queue.[3] A specialist outpatient clinic is as a system of servers where patients are seen, with varying rates of arriving patients, varying process times and varying delays. From a system perspective a delay in access is an aggregate result of the involved parts not functioning properly together to match supply and demand. Demand to a specialist outpatient clinic is the number of patients requesting an appointment (the arrival process, each time implying a certain service time), including demand that is generated by the specialist when he advises a patient to schedule a follow-up appointment. Supply is the number of appointments offered. A queue is the number of patients waiting in line for an appointment, resulting in a delay in access. The delay can be measured as the time between

requesting an appointment and the availability of an appointment. See Figure 2.1 for a graphical display. The queue is displayed as a reservoir of patients waiting for an appointment to be supplied. If demand is higher than supply the water level, or queue, will increase and vice versa.

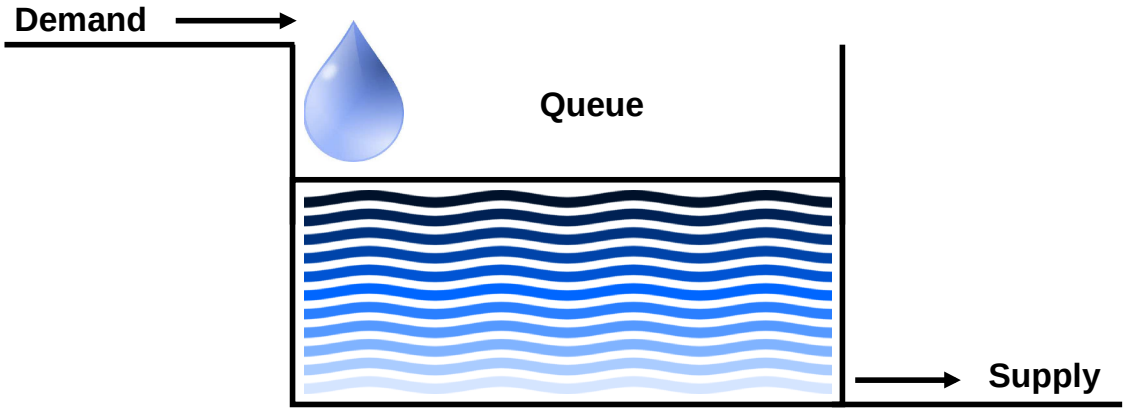


Figure 2.1 A simple queuing system

2.3 Causes of delays

Delays in a simple queuing system can have three causes:[4]

1. Demand is larger than supply
2. There is uncertainty in the arrival process
3. There is uncertainty in the service process

The 'arrival process' is the daily demand for appointments. The 'service process' is the number of appointments supplied each day and the different appointment lengths. If demand is larger than supply the queue increases continuously. For example, if every week 150 people ask for an appointment and 100 appointments are supplied, the queue will increase with 50 every week. In Figure 2.1 the water level between demand and supply will keep rising. Our research, however, is aimed at situations where demand and supply are equal on average but poorly matched. This results in a continuous delay that fluctuates within certain boundaries.[5] We are therefore concerned with the second and third causes of delays, which is essentially about managing uncertainty. Concerning delays in access to specialist outpatient clinics, the uncertainty concerns the variation in the volume of demand and supply. See for example Figure 2.2. The thin yellow line represents the number of requested appointments per week and the thin blue line represents the number of appointments supplied. The thick lines represent trend lines. The averages of the lines are equal but they are out of sync. Where the yellow lines are above the blue the patients experience a delay and the queue is increasing. Where the yellow line is below the blue line the queue is decreasing. The more supply and demand fluctuate out of sync, the longer the average delay will be.[6]

Conclusion: the first driver of delays is uncertainty about the volume of supply and demand. The second driver is the utilisation percentage of capacity. When there is uncertainty in the arrival and service time, in this case in the number of appointments requested and supplied, the delay for the service can grow in a highly non-linear fashion when the utilisation of the capacity approaches 100%, as depicted in Figure 2.3 and explained in the next example.[1]

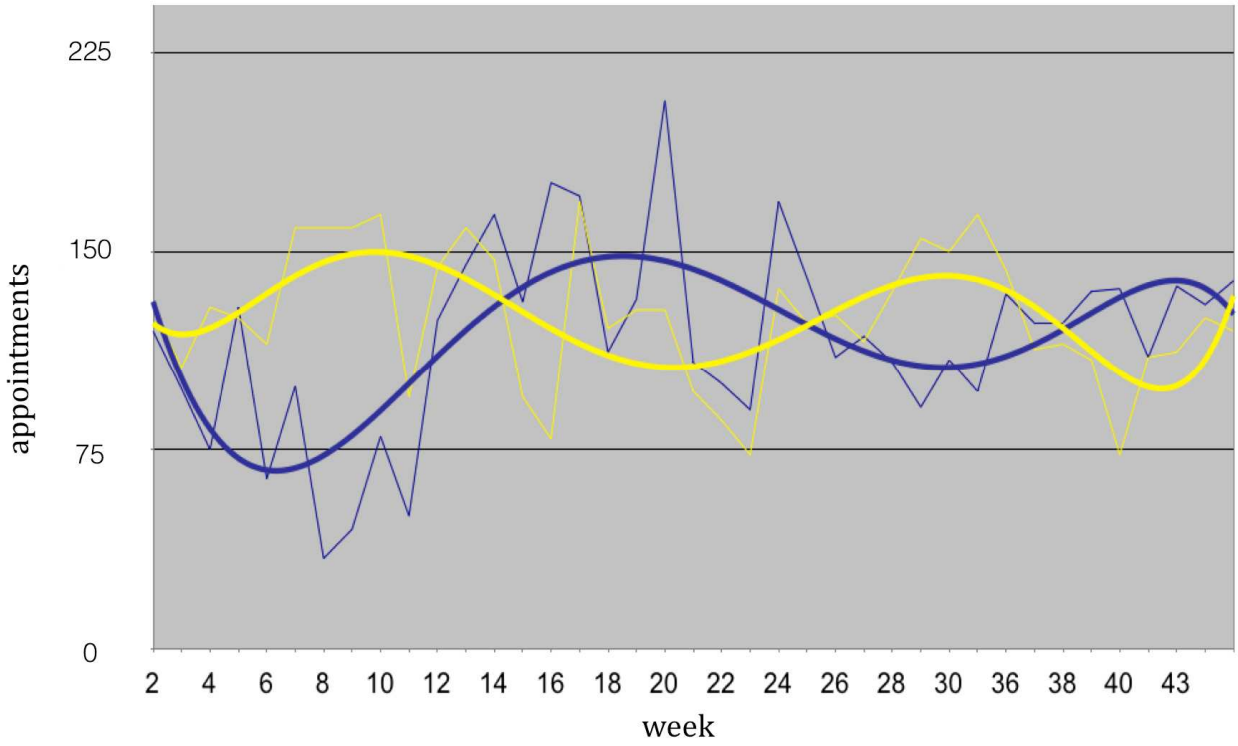


Figure 2.2 Supply and demand of appointments per week, with trend lines

If demand for example fluctuates between 50 and 150 every other week and supply is always 100, the queue will need to fluctuate from 0 to 50 every other week to work at 100% utilisation with minimal delay. Now imagine that it's random whether demand will be 50 or 150 every week. Then it could be that the next three weeks demand will be 50. Then a queue of 150 at the start is needed to work at 100% utilisation. It's also possible that the next eight weeks demand will be 50 (statistically once every five years). There needs to be a permanent queue of at least 400 to be able to deal with that situation when it arises. The same logic applies to fluctuations in supply. If supply is not 100 every week, but also fluctuates between 50 and 150, the needed buffer to work at 100% utilisation increases in the same fashion. The more variation in either demand or supply and the more uncertainty on the variation, the longer the queue needs to be to be able to always work at 100% utilisation. This relation is not linear, but exponential because the relation is inverse. If the utilisation goal is increased from 50% to 75%, at a certain level of uncertainty, the required queue doubles. From 75% to 87,5% it doubles again, etcetera. Strictly speaking, in order to be certain that 100% utilisation is possible in every statistical situation imaginable, the queue needs to be infinite.

Compare it to driving on the highway at 120 kilometres per hour. When the traffic gets busier, the cars still doing 120, get closer to each other. When it gets very busy however the traffic becomes nervous and suddenly you need to slow down and come to a stand still. This sudden change from 120 to standing still can be caused by a small percentage of extra cars on the road, when approaching 100% utilisation. The uncertainty is the difference in speed between the cars, individual cars slowing down or accelerating and the variation of cars joining and leaving the motorway. Close to 100% utilisation there is not enough space for drivers to adapt fast enough to changes around them without causing other cars to slow down.[7]

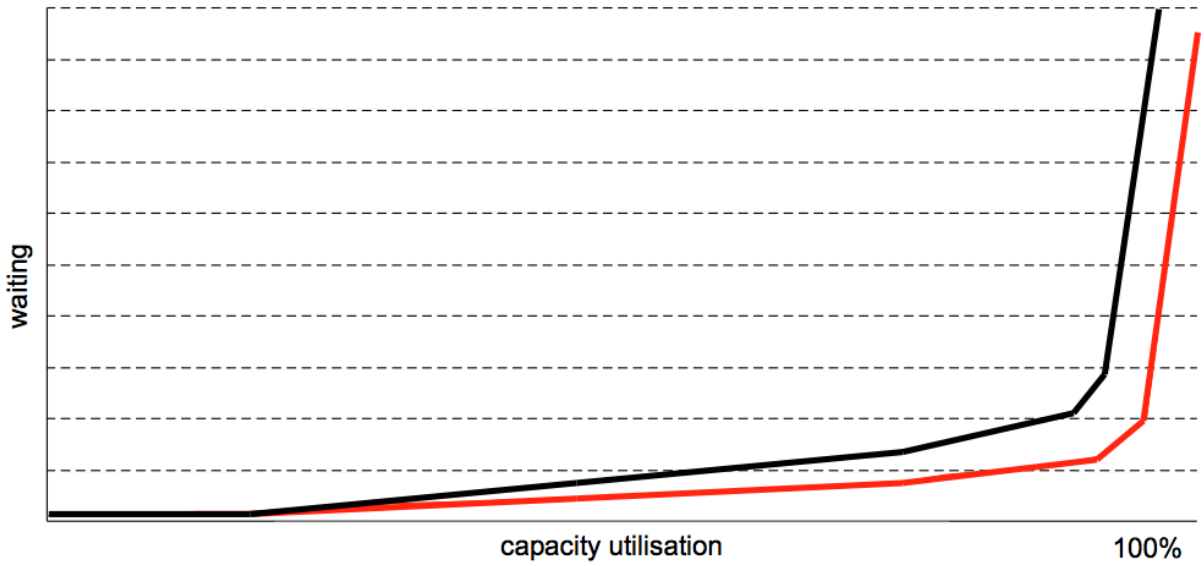


Figure 2.3 The relation between waiting and capacity utilisation

The more uncertainty there is in the system, the earlier the exponential rise of the curve starts.[1] In Figure 2.3, the upper curve depicts the relation when there is more uncertainty in the arrival and service process then the lower curve. A strategy to achieve short delays while optimising utilisation is thus to reduce variation in the system.

2.4 Managing uncertainty

As stated before, managing delays is concerned with managing uncertainty. Uncertainty is the difference between the amount of information required to perform a task and the amount of information available in the organization.[8] The greater the uncertainty about the volume of supply and demand, the greater the amount of information that must be processed among decision makers in order to achieve a given level of delays in access (more about types of information and information processing later). General organisational strategies dealing with uncertainty are to:[8]

1. Increase the ability to preplan
2. Increase the flexibility to adapt to the inability to preplan
3. Decrease the level of performance required for continued viability

Optimising delays in specialist outpatient clinics is concerned with managing uncertainty in the volume of demand and supply of appointments. To deal with the uncertainty in the distribution of demand, an outpatient specialty clinic can forecast fluctuations in demand and plan supply accordingly on an aggregate level (strategy 1). For the remaining uncertainty it can react to unpredicted fluctuations on an operational level (strategy 2).[9] To deal with uncertainty, one has to either plan late (short-term planning) or react to the unexpected situation. The latter we call reactive decision-making.

The responsiveness of a system can be defined as 'the actions or behaviour of a system using a set of capabilities to purposefully and timely address changes triggered by external stimuli, like a change in demand'.[10] In a responsive system, demand determines the supply of capacity, which needs to be flexible enough to adapt. Flexibility in supply can be defined as the extent to which a system is able to create variety in the mix and/or volume of resources in a system, so that the

system state can be changed without adding resources or fundamentally changing the resources.[11] Increasing flexibility is a strategy to increase the responsiveness of a specialist outpatient clinic.

The more uncertainty about the variation in demand, the more buffer-capacity is required to keep waiting times to a minimum (Figure 2.3). The more flexibility to utilise capacity where demand is, the less buffer capacity is needed and the higher the utilisation of capacity can be. There are several types of flexibility to match supply and demand when managing uncertainty:[12]

- Mix flexibility: ability to produce different products at the same point in time. In a specialist outpatient clinic this concerns the ability to serve different mixes of appointments in one session and different mixes of sessions in a week. Using shared queues (slots where different types of appointment can be booked) is a strategy to increase the flexibility in the mix to adapt to variation in the volume of demand per type of appointment.
- Changeover flexibility: ability to deal with changes in the product mix over time. In a specialist outpatient clinic this translates into being able to deal with changes in the mix of types of services/appointments over time, for example changes in the ratio between new and follow-up appointments. Again, sharing queues reduces this problem; reversely: the more queues, the more frequent changes in the product mix need to be dealt with.
- Volume flexibility: ability to easily make changes in the aggregate production amount. How much and how fast can a clinic increase or decrease the number of appointments offered on a single day or week? How soon is a new specialist hired, supporting staff hired and if necessary, more space created to perform sessions?

The variability in demand volume can be exogenous variation (out of control of the clinic) and endogenous variation (potentially within the control of the clinic).[13] For example, the difference in the number of patients that request a new appointment in two different weeks is natural variation. If an outpatient clinic cancels sessions and as a consequence there are less requests for follow-up appointments several months later, then that is artificial variation that can be controlled.

2.5 Managing complexity

The level of uncertainty and associated information that needs to be processed can go beyond the processing capacity of a specialist outpatient clinic as a consequence of increasing complexity. We define complexity as the number of linkages and information flows between entities that affect the way supply and demand are matched. This is relevant for matching supply and demand, because the complexity determines the required coordination of the dependencies between the involved elements. For a system to be stable, the number of states of its control system must be greater or equal to the number of states in the system being controlled ('Ashby's law').[14] If for example in a specialist outpatient clinic the scheduling of a group of specialists of one discipline also needs to be coordinated with the scheduling of a group of specialists of another discipline, the complexity increases and the number of states in the system increases. As a result, either the coordination efforts must increase, or the performance will decrease (delays increase or utilisation decreases). One strategy to manage this problem is by reducing complexity, in other words to reduce the number of linkages for example by reducing the dependency of a clinic on other entities in order to supply appointments.[8]

2.6 Single and shared queues

An important factor that determines the complexity of queuing systems and determines the level of uncertainty that needs to be managed is whether 'single queues' or 'shared queues' are used.[1]

> *Grocery stores typically have individual queues for checkout lanes, while banks often have a single queue for all tellers. The reason banks do this is to reduce congestion by pooling variability in the process times. If one teller gets bogged down serving a person who insists that an account is not overdrawn, the queue keeps moving to the other tellers. In contrast, if a cashier is held up waiting for a price check, everyone in that line is stuck (or starts lane hopping, which makes the system behave more like the combined-queue case, but with less efficiency and equity of waiting time.*

Specialists generally perform personal sessions, meaning that each session in the specialist outpatient clinic is planned per specific specialist and the patients in each session have an appointment with that specific specialist. Personal sessions are 'single queues', meaning that patients 'line up' in a queue for each specific specialist, similar to grocery stores. In contrast, in an emergency department for example the first specialist that is available will see the next patient. This is a system with a 'shared queue'; similar to what happens in banks.

When booking an appointment in a specialist outpatient clinic however, it becomes a bit more complicated. If a patient expresses no preference for a specialist, the assistant books a new appointment with the first specialist that has an appointment available. For these 'appointments without preference', the system performs as a shared queue like the banks. When a patient requests an appointment with a specific specialist, the patient lines up in a single queue. This happens when patients are referred to a specific specialist, or because the patient chooses a specific specialist by reputation, or when a patient books a follow-up appointment with his or her 'own' specialist. Sometimes specialist outpatient clinics also have sessions that are not specialist-specific, for example two sessions a week for urgent appointments and the specialists rotate in performing those sessions. University and teaching hospitals can also plan sessions that are performed by any of a group of residents.

Within sessions, a similar hybrid nature is present because within sessions several types of queues for the same specialist can be present. Sessions can have flexible slots for appointments that can be used for different types of appointments (e.g. a new or follow-up appointment). These are shared queues like the banks. Sessions can also have fixed slots that can only be used for a specific type of appointment ('fixed appointments', for example at 9:00am on Monday a slot can only be booked for a new appointment). Each fixed type of appointment is a single queue like the grocery store.

From a queuing theory perspective, the more uncertainty has to be dealt with, the more shared queues perform better than multiple single queues with respect to utilisation and delays. Shared queues dampen the effect of variability by pooling resources to distribute the workload more efficiently. In Figure 2.3, besides reducing variation as the first strategy, shared queues form a second strategy to make the exponential curve rise closer to 100% utilisation with the same delay. For specialist outpatient clinics, reserving capacity in the schedules exclusively for specific types of appointments ('fixed appointments') will make the system more vulnerable to variation and as a consequence either the delays will increase or the utilisation will reduce.

If sharing queues increases the linkages and information flows it increases the required capacity to process information to coordinate the activities. This can reduce performance. On the other hand, if queues exist within a resource, as is the case when the capacity of a specialist is divided in several types of sessions and reserved appointments within the sessions, sharing the queues by reducing the number of reserved appointments reduces the amount of information needed because there are fewer flows that need to be managed. If for example the delay for one queue increases and another queue stays short, there will be a point where a decision needs to be made whether an exception can be made to schedule a patient in another queue. Shared queues can reduce this form of coordination, which is another reason why shared queues can improve performance.

2.7 The role of information

This paragraph discusses the required information to perform a task, in our case to match supply and demand in a timely manner while optimising utilisation. What information is used to make the decisions concerning the release of capacity? How are the effects of decisions known? What feedback or feedforward information is required to decide corrective actions to optimise the match between supply and demand? This paragraph explores the nature of feedback and feedforward and the relevance for creating pull.

Feedback information

We explain what feedback is and why it is important for a management system with the following example:

> *In the winter of 1763–1764, Watt was asked to repair a model of Newcomen's engine that was used for demonstration lectures at the university. In acquainting himself with the model, Watt was impressed by how much steam was required to run the engine. He undertook a series of experiments on the behaviour of steam and found that a major problem was the temperature of the cylinder walls. Newcomen's engine wasted most of its heat in warming the walls of its cylinder, since the walls were cooled on each cycle as cold water was injected to condense the steam, forcing the piston back under air pressure. Early in 1765, Watt remedied this wasteful defect by devising a modified type of steam engine. In retrospect, it sounds like a simple idea. After pushing the piston up, the steam was admitted to a separate container, called the condenser, where the steam condensed at a low temperature. With this system, the cylinder containing the piston could be kept hot all the time, and the condenser could be kept cool all the time. The use of the separate condenser allowed huge fuel savings. Watt's engine could do twice as much work as Newcomen's with the same amount of fuel. Watt also added many other refinements, such as automatically controlled valves that were opened and closed by the reciprocating action of the piston itself, as well as a governor that controlled the amount of steam reaching the engine, to maintain a constant speed for the engine. The latter idea of using part of the output of the process to regulate the process itself, is called feedback. It is an essential part of the design of many modern mechanical and electronic systems.* [15]

Relevant for optimising the way supply and demand are matched is the way capacity is released and becomes supply to meet demand. The trigger to release capacity is feedback or feedforward information. Feedback on the status of the process in the outpatient clinic itself can concern the process of matching supply and demand or it can concern the deliverance of appointments itself. Feedback on the process of matching supply and demand can be the current waiting time, the

percentage of urgent appointments or the average number of overbookings (added to a session even though there is no more space in the agenda). Feedback on the deliverance process can for example be the percentage of no shows or how much longer the sessions took compared to the planned time.

Feedback has many interesting properties that can be used when designing systems. For example, feedback can be used to make a system resilient toward external influences and it can create linear behaviour out of non-linear components.[16] The principle of feedback is simple: base corrective actions on the difference between desired and actual performance. One of the key uses of feedback is to provide robustness to uncertainty.[16] Imagine for example a specialist that cancels outpatient sessions. This creates variation in the supply of appointments that can lead to an exponential increase of waiting (Figure 2.3). If for example the desire is to offer a steady supply of new appointments, the system could automatically adjust the ratio of offered new to follow-up appointments to adjust for cancellations. If for example half of the sessions in a week are cancelled, the percentage of new appointments would be doubled. In this way, feedback can be used to create linear behaviour, in this case the levelling of supply of new appointments.

Feedback can also have disadvantages. It can create instability, causing oscillations (e.g. amplification of a microphone) or even runaway behaviour. An example is when a specialist becomes demotivated when he receives feedback on a rising level of no shows and he looses discipline and starts cancelling sessions on a short notice, which leads to a rising number of rescheduling of patients, which leads to less commitment from patients, leading to a further rising of the percentage of no shows. Another potential drawback is that the combination of different forms feedback can introduce unwanted noise into a system, requiring careful filtering of signals that are caused by the measuring.[16]

Feedforward information

Feedback information is reactive: there must be an event before corrective actions are taken. In some circumstances it is possible to measure a disturbance before it enters the system and this information can then be used to take corrective action before the disturbance has influenced the system: feedforward information.[16] Feedforward information provides information on the current status of a process, which predicts the future status of a process. An example is the percentage of scheduled appointments versus open appointments for the next six weeks. If this percentage rises above a certain level, it can predict that in the near future a delay in access will occur. Another example is registering how many patients agreed to a follow-up appointment in a future month that are not yet booked. This gives an indication of that part of follow-up demand that can be expected in the future.

Both feedback and feedforward information can be used for managing patient flow. Feedforward can offer the possibility to react faster to changes because it makes a development earlier visible. Experience indicates that it is often advantageous to combine feedback and feedforward, and the correct balance requires insight and understanding of their respective properties.[16]

2.8 Optimisation questions

Strategies to improve the performance of a queuing system come at a cost and it depends on the specific circumstances how effective each strategy is. The consequence is that optimising delays in

access requires the optimisation of several choices. In this chapter four optimisation questions have been discussed:

- Reduce waiting vs capacity utilisation
- Invest in information processing vs reducing the need for coordination
- Invest in planning vs reacting
- Invest in reducing variation vs increasing flexibility to deal with variation

3. Pull

3.1 Introduction

In industry Just-In-Time (JIT) manufacturing is an important concept to optimise delays. JIT is considered the basis of the Japanese economic success in the second half of the 20th century.[1] JIT caused a revolution in the science of Operations Management (OM), spread around the world and has become an important chapter in in the history of manufacturing management.[1] JIT is a collection of techniques to manage flow by increasing the responsiveness of a system. Pull methods are an alternative to push methods, which were the common methods until JIT was invented. The ultimate application of JIT would mean producing with zero inventory. In the case of a specialist outpatient clinic, the ultimate application of JIT would be direct access for all patients, regardless of the urgency. A central principle of JIT is to switch from 'push' to 'pull' as an operating system.[1] The concept of 'pull' has proven to be successful in managing flow and increasing the responsiveness of the system to demand. It has been extensively studied in factories and is well recognized for its effectiveness.[1] Paragraph 3.2 describes what preceded pull and paragraph 3.3 formally defines pull. Paragraph 3.4 discusses the distinction between the strategic and tactical level of pull, 3.5 provides an example of a pull system and 3.6 discusses the potential benefits of pull. Paragraph 3.7 discusses related concepts to pull and 3.8 elaborates on the concept of cycle mix planning. Paragraph 3.9 discusses the combination of push and pull methods and 3.10 concludes with a discussion of first and second order change in relation to the application of pull methods.

3.2 What preceded pull: MRP, MRP II and ERP

In the 1960's the upcoming use of computers led to the development of Material Requirements Planning (MRP) for scheduling and inventory control. In 1972 this development received a huge boost when the American Production and Inventory Control Society (APICS) launched its 'MRP crusade'. Before MRP most production control systems produced an item if the inventory of that item fell below a specified level.[1] MRP, however differentiates between the external demand of final products and internal demand for components. Essential is the notion that external demand is uncertain, but internal demand for components is known. The relationship between final products and components is described by the Bill Of Materials (BOM). MRP uses the BOM to compute schedules of what should be started into production based on external demand, which makes it a push system (see further). Despite widespread enthusiasm MRP and its improvement for controlling the purchasing of components, quickly needed to deal with many production problems like long lead times and system nervousness.[1] This was mainly caused by the MRP assumption that lead-time is constant, which implies the assumption of infinite capacity.

To deal with the shortcomings, MRP II was developed. MRP II included demand management, forecasting, capacity planning, master production scheduling and more. MRP II further integrated

business functions into a common framework. In the 1990's Enterprise Resource Planning (ERP) became its successor. ERP integrated the framework into a management tool that can consolidate and track enormous quantities of data. SAP and other suppliers created a multi-billion dollar ERP market. ERP, however, like MRP still has fundamental problems because of the assumption that lead times are fixed and capacity is infinite, still causing problems in long lead times and system nervousness.[1] Furthermore, the disaggregation of aggregate estimates to individual products and capacities is only possible for deterministic processes.[9] Deterministic processes are processes for which all parameters describing the processes are known and constant.

Planning and control in health care is different from manufacturing. In manufacturing industries logistics is about managing inventory, in service industries like health care, logistics is about managing capacity.[10] Manufacturing industries can produce parts or end products before demand is expressed. An inventory enables a manufacturer to respond fast to demand and/or to prevent idle time for the next resource to produce due to upstream variation in supply or downstream variation in demand. In service industries like health care, nothing can be produced without the patient being present, making the use of inventory as a buffer before demand is expressed impossible. Inventory is only possible in the form of waiting lists for patients, for external demand, but not for internal demand. To respond fast to external demand in service industries like health care (short waiting lists or short delays in access), the buffer can only be found in capacity. The trade-off is idle time: accept more idle time to respond faster or vice versa. But there are strategies to minimise the buffer capacity in order to respond fast. Just-In-Time (JIT) production is of interest for any system attempting to reduce delays, because it addresses the problems of the MRP and ERP systems and aims to minimise delays.

There is another consideration to take into account when applying industrial concepts to health care. In contrary to the manufacturing environment where MRP was invented, in many hospital processes it is uncertain beforehand which capacities will be needed for individual patients. For example, in the specialist outpatient clinic it can often only be decided during the appointment which diagnostic capacities are required. The MRP and ERP systems can only be applied to the deterministic processes of hospitals and not for the non-deterministic processes. To optimise delays in specialist outpatient clinics a method is needed that can deal with uncertainty in external as well as internal demand and that takes into account that there is no possibility to use inventory as buffers.

3.3 General definition and application of pull

The inventor of JIT, Taiichi Ohno from Toyota, explained the concept of pull in 1973 as:[17]

> *Manufacturers and workplaces can no longer base production on desktop planning alone and then distribute, or push, them onto the market. It has become a matter of course for customers, or users, each with a different value system, to stand in frontline of the marketplace and, so to speak, pull the goods they need, in the amount they need and the time they need them.*

More specifically, 'pull' refers to the mechanism that triggers the movement of work in the system. Fundamentally, in a pull system the trigger for work comes from inside the system. For factories pull can formally be defined as:

*A **push** system* schedules *the release of work based on demand, while a **pull** system* authorizes *the release of work based on system status.* [1]

The difference between push and pull systems is depicted schematically in Figure 2.4.

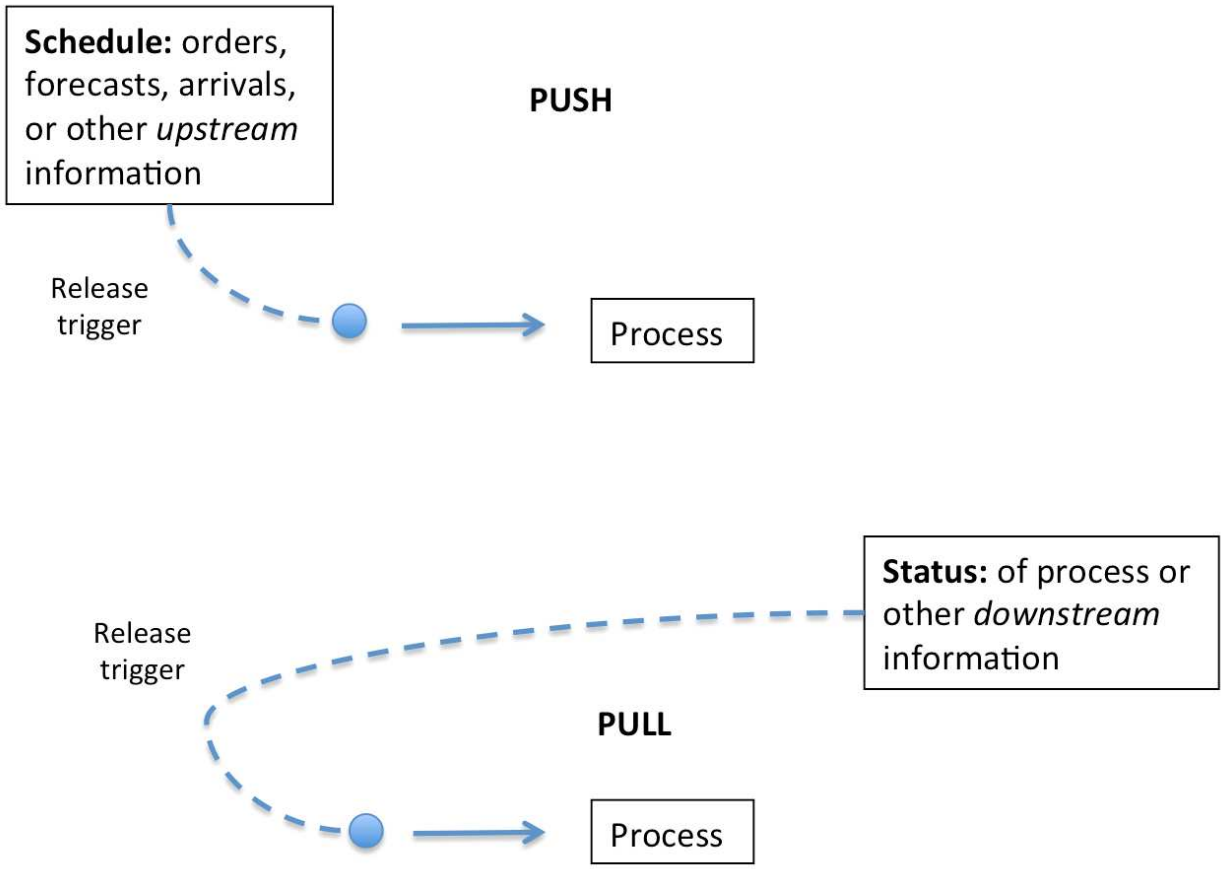


Figure 2.4 Release triggers in push and pull production systems[1]

In a push system, the trigger to *schedule* production is based on *upstream* information. Upstream information can for example be a forecast that in August demand is expected to decrease. In a pull system the trigger to *release* work comes from the status of the process itself or another *downstream* station. The status of the process can for example be movement of a product to the next station, triggering a signal upstream that they can move work. Downstream information can concern for example be the length of the queue for the next activity. Note that a pull system acts real-time: actions are based on the actual conditions of that moment. If the situation changes, a pull system will translate that change directly into a change for the trigger; in a push system decisions on 'when to produce what' are planned for the future, without knowledge of the conditions when the production will take place. What is considered inside or outside the system depends on the scope of your manufacturing system. For example, car manufacturers generally consider the dealers that sell their cars as outside the manufacturing system. Toyota in contrast lets dealers install part of the accessories and considers dealers as inside the manufacturing system.

3.4 Strategic and tactical level of pull

Given the strategic intent to let the customer (demand) pull the goods they need, it may be confusing that push is based on demand and pull is based on the system status. This apparent

contradiction can be explained by making a distinction between the strategic meaning of pull and the tactical application of pull. The explanation of Ohno is on a strategic level: the intent to produce exactly what and when customers want it.[17] This implies that there is minimal inventory in the system, because inventory indicates that either you have produced something the customer has not yet asked for (push), or it takes long between the expression of demand and responding to demand, which contradicts the intent to deliver when the customer needs it. A pull system will therefore enable an organisation to respond better to (changes in) demand with minimal inventory as a tactical goal.[18]

The best-known tactical application of pull is the 'kanban' system that Toyota developed.[1] The kanban system authorizes production as inventory is consumed. The goal is to minimize 'overproduction': producing nothing that is not needed at that time, avoiding inventory. In a push system, there is no control of the amount of inventory in the system. This often results in long throughput times and high levels of 'Work In Process (WIP)', because inventory piles up between steps as a consequence of variation in the processes. So, even though push systems are based on demand, they are often either not responsive to actual demand or they base their schedules on *forecasted* demand leading to large inventories of finished goods that are pushed onto the market.[1]

Contrary to this problem of non-responsiveness and high costs of inventory a characteristic of pull systems in factories is a limit on the maximum amount of inventory. The limit means that you can't perform your activity if the person after you ('downstream') has not signalled that he needs your activity. Defined more formally:

> *A pull production system is one that explicitly limits the amount of work in process that can be in the system. By default, this implies that a push production system is one that has no explicit limit on the amount of work in process that can be in the system.* [18]

We agree that a pull system has an explicit limit, but we do not agree that the definition of push is 'by default' every production system that has no explicit limit. It's imaginable that a push system can have an explicit limit on the amount of work in process.

3.5 Example of a pull system

Imagine a small sports model car factory with the strategic intent to let customer demand pull production with short waiting times and the tactical choice to minimise inventory by using a kanban system. There is only one model, but with three different colours. The factory assembles large parts by adding three big parts to the chassis in three sequential steps and in the fourth step the colour is painted. Figure 2.5 visualises the production process; the circled steps 1 to 5 follow the pull method and are explained in the text below. WIP is 'Work in process' which are unfinished cars. Inventory represents parts that are use to assemble the car. The production system can look like this:

Inventory and Work In Process (WIP):

- In the parking lot behind the factory three finished cars are parked, one of each colour.
- In between the four production steps (three assembly and one painting) is space for three chassis as inventory and all three spaces are filled.

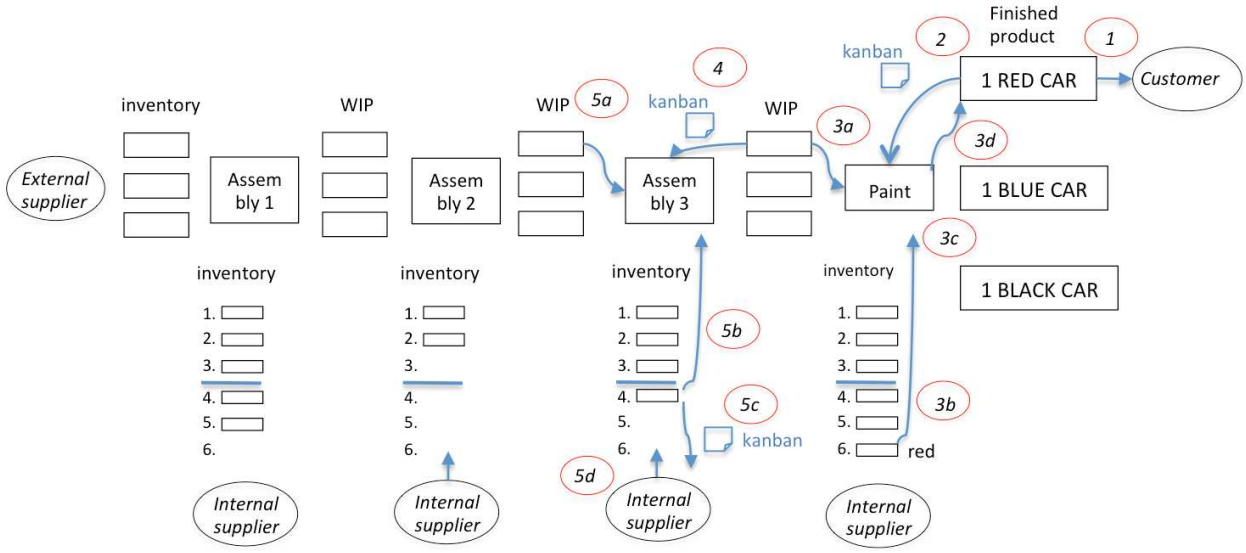


Figure 2.5 A pull-based production process

- When a step has three parts of inventory left to assemble, a (kanban) card goes to the supplier and immediately three more parts are delivered. This is indicated with the line between '3.' and '4.'

The production process:

- Each time a customer buys a car, the car with the chosen colour is driven from the parking lot to the customer (the circled step 1).
- When a car is driven away, a (kanban) card with the specified colour is handed to the person who does the painting (step 2). The painter moves a car from the queue before his step (step 3a), takes a pack of the requested colour red from inventory slot 6 (step 3b), paints the car in the specified colour (step 3c) and delivers it to the 'RED CAR' slot (step 3d).
- When the painter moves a car from the queue, a (pull) signal is given upstream to the step before the painter, where the doors are assembled (step 4). The person handling the doors pulls a chassis with body from his queue of three chassis in WIP (step 5a), takes a door from inventory slot 4 and adds the doors. Because there are now only three parts of inventory left the painter takes the kanban card and places it in a box (step 5c), signalling the internal supplier to deliver three doors (step 5d).
- Etcetera.

Observation:

- The WIP is maximum 16 chassis (4 chassis that are used during the 3 assembly and 1 paint steps and 4 times 3 chassis as WIP before each step).
- The inventory of parts to assemble is maximum 24 (4 times 6 parts in inventory; the bucket of paint for one car is considered a part).
- Everything that is produced is exactly what the customer demanded, even though the production of that moment does not go to that customer (the customer already received his requested car from the inventory of finished cars).
- If there is no customer demand, nothing is produced.

- As long as the rate of demand does not exceed the rate of production there is no waiting time for the customer. If the rate of demand is higher, a continuously increasing waiting time (demand queue) develops. If the rate is lower, for example half the rate of production, the assembly line will be standing still half of the time. There is a minimal inventory of finished cars, but the utilisation of this system is vulnerable to variation in the rate of customer demand.
- If a step is delayed (for example because of a machine breakdown), the steps before that step only assemble one more part on a chassis and then stop (because they do not receive a signal anymore to continue). The steps are completely dependent on each other.
- The steps after that delay continue to produce a maximum of three more, because that's the maximum WIP between each step. After that, they will wait until the delayed step starts producing again. Again, the steps are completely dependent on each other.
- If am internal or external supplier is not able to deliver parts when a signal is given, only three more cars can be produced (because there are only three parts of inventory at that moment). This makes the system vulnerable to supplier reliability.

3.6 Benefits of pull

A pull system like the one described in the example has minimal inventory, short throughput times and maximal flexibility in the mix of what to produce based on the actual demand (versus forecasted demand). The potential benefits of pull systems are:[18]

1. *Reduced WIP and cycle time*. Because of the maximum inventory between each step there is minimal WIP and cycle times.
2. *Smoother production flow*. The limit on WIP reduces variation in the process, leading to a more predictable output stream because no step can produce more than the other steps.
3. *Improved quality*. A pull system makes the parts of the process depend more on each other because there is almost no inventory that can serve as a buffer. This forces each step to create a stable outcome with no defects, or the next step will shut down. Problems are directly visible and require immediate corrective action. This creates pressure to address root causes of problems to ensure consistent quality.
4. *Reduced cost*. Inventory is expensive; pull methods reduces the cost of inventory because no step can produce more than is required by the next step. Perhaps more importantly, the increased dependence forces the system to make efficiency improvements because there are almost no buffers between steps to dampen the effects of unexpected problems.

A pull system achieves these benefits because 1) there is less congestion in the system because it is a closed system; 2) it is easier to control because the minimal inventory can be observed directly; and 3) WIP is bounded.[18]

In most factories there is a large variety of a product. For example: the colour, engine specifications and accessories of a type of car can lead to a large variety of different versions of that car. Traditionally, factories would produce large batches with the same specifications in order to lose as little time as possible with setup times between runs with different specifications. This again leads to long waiting times for customers to have their specific product, or it leads to large inventories. Pull systems prevent both problems.

3.7 Related concepts

To achieve the benefits of pull other concepts usually need to be implemented as well. In Toyota for example the concept of 'levelling' the production (a constant flow with minimal variation in the volume of demand per workstation per minute or hour) is fundamental to make their pull system work because without a levelled production and without inventories the capacity utilisation would be to low. A pillar of levelling is the use of 'Takt time': the pace of assembly that is the same for every step in the process, and that is based on the pace of demand. To achieve Takt time throughout a factory requires a high level of standardisation in order to reduce variation that will cause steps to wait for each other. Pull also requires a very flexible production system because different types of models are produced on the same day to meet actual customer demand. To enable this, reducing 'setup times' is even more important than in a push system, because in a pull system setup changes occur more often (setup times = the time it takes to make the required changes to produce a different model, in the example in 3.5 it could be changing the spray mouth to paint the next car in a different colour). Toyota developed all these and more principles and methods as part of the Toyota Production System, with aim that the elements integrate into one system. It is the attention for the interaction between the elements that makes the system work.[17] In the next paragraph we explain in more detail 'cycle mix planning', one other related concept that can be important to make pull work.

3.8 Cycle mix planning

In real life, there are many aspects that need to be planned in advance, whether working with a push or a pull system. For example the number of people working in the specialist outpatient clinic and their contracts are to a certain extent fixed and to a certain extent flexible. A pull system will try to use the available flexibility to release appointments as much as possible based on the system status. This is an optimisation question. A pull system can make use of extensive planning; the criterion for push or pull is how capacity is finally released. The challenge of planning in a pull system is to constantly optimise the planning system to minimise deviations from the plan. We use descriptions of the book 'Toyota Supply Chain Management' to provide an example how this can be achieved with cycle-mix planning.[19]

> *Cycle-mix planning at Toyota*
>
> *The sales division of Toyota provides a rolling three-month plan with the first month categorized as a firm order and the next two months as a forecast. The aggregate production plan is agreed upon between the sales division and the manufacturing division. The sales division wants to remain flexible and respond quickly to market changes and to limit the use of incentives to sell vehicles. The manufacturing division wants to operate at full capacity and produce high-profit vehicles. The agreed month implies that the sales division commits to buying these units and the manufacturing division has agreed to produce them. The total volume of cars for the coming month cannot change anymore; the specifications of the cars (engine, colour, options, etc.) can change up to a week in advance. Manufacturing then assigns a production date and sequence to each vehicle for the operational planning. The objective is to create a levelled schedule that avoids congestion, workload imbalance and inventory. Also, about 5% of the time is planned as a buffer for continuous improvement activities during production. The horizon is based on the predicted mix of demand that can be planned in a levelled production schedule and is then repeated cyclical with small margins to change the*

mix based on actual demand. If the mix of demand changes over time (e.g. more sunroofs are desired), the cycle mix will be reconsidered.

Manufacturing schedules staff with standard work hours and preschedules usually a limited amount of overtime, for example one hour per shift. Day-to-day adjustments to the overtime are made just prior to the start of each shift based on operational conditions of that day. Suppliers receive detailed schedules that specify exactly how much of what needs to be delivered at an exact time. Under normal circumstances there is a maximum of +5 and -5% variation of the volumes to be delivered. Dealers can specify changes until a week before production. Toyota ensures that this deadline is longer than the lead-time of at least 80% of the suppliers. Of the remaining 20% a small inventory is accepted.

As described Toyota is able to manage its supply chain externally and internally in an extremely balanced and smooth manner because of a vary carefully decided product mix planning. The objective of product mix planning is to strongly reduce the number of variants of vehicles. Product mix planning is initially performed annually before a new model launch and can be adjusted monthly to reflect changes in demand and/or seasonal trends. To decide the right product mix, a balance is sought between: 1) serving as much as possible the range of different variants that customers desire and 2) choosing a mix of variants that can be manufactured efficiently and in a smooth flow with as little batching as possible. On an aggregate level this implies for example that the variant that requires the most work (like a sun roof) will be a maximum of for example fifty produced in a week. On an operational level it can imply that this variant will never be scheduled twice directly after another (since for each variant the same average time is available, this could either cause congestion or overburden for the workers).

Fluctuations in demand per type of vehicle can be absorbed by Manufacturing by changing the product mix that is produced in the month plan, by extra overtime or extra temporary workers on a daily basis and by changing the 'Takt time', the rhythm how many cars are produced per hour. The latter can be changed approximately once every two months. Sales can influence demand with marketing campaigns or sales incentives.

3.9 Combination of push and pull

In reality factories almost always have hybrid systems with push and pull elements combined. It is for example possible to implement a kanban based pull system in a factory that gets a production schedule based on forecasted demand (= push). It depends on the context which mixture provides the best results. If demand fluctuates heavily and unpredictably, it's probably wiser to produce at least partly on forecasted demand or by buffering time that customers need to wait, to ensure efficient production. The choice where to apply push and where pull as a tactical system is an optimisation challenge, depending on factors such as variation in rate of demand, impact of setup times on total throughput time, ability to work with levelling and Takt time, but also the maximum lead time that is strategically chosen to deliver fast to customers. Depending on the context and strategy of the organisation, many other considerations can be relevant. The potential benefits of pull are less when mixed with push methods, which is also an optimisation question.

3.10 First and second order change

To make a pull system work may require first and second order change. When a thermostat measures that the temperature in the house is below a specified degree and it turns the heating on, than this is an example of a change. When you change the specified temperature that turns the heating on, that is an example of first order change: a problem is solved (the temperature setting does not meet the requirements) by applying existing knowledge. Learning is restricted to evaluating an action within the existing parameters.[20] When you add a sensor that measures the outside temperature and program the thermostat to analyse the relation between changes in the outside and inside temperature and let it experiment by turning the heat on at a change in the outside temperature before it affects the inside temperature and then learn whether it was too early or too late (there is a delay in turning the heating on and it having effect on the temperature), than you have created second order change. The parameters for action and for first order learning themselves are changed. Second order change requires learning in the way learning itself takes place. Introducing a pull system changes the existing parameters how demand and supply are matched and will require new forms of feedback and feedforward information and second order change to learn the right parameters for action.

4. The science of waiting and pull applied to specialist outpatient clinics

4.1 Introduction

This part applies the science of waiting and pull methods to specialist outpatient clinics. Paragraph 4.2 describes the general characteristics of how specialist outpatient clinics match supply and demand. Paragraph 4.3 applies the science of waiting by discussing buffering options of the uncertainty and variability of supply and demand for specialist outpatient clinics. Paragraph 4.4 defines pull for specialist outpatient clinics and 4.5 discusses more specifically cycle mix planning. Paragraph 4.6 discusses the potential benefits of pull for specialist outpatient clinics. Paragraph 4.7 places the matching of supply and demand of specialist outpatient clinics in the larger system of the hospital.

4.2 Characteristics of specialist outpatient clinics

A common approach for this purpose is a Production Control Framework. A Production Control Framework deals with the balance between service (delay in access) and efficiency (capacity utilisation), at all levels of planning and control.[21] Planning concerns the decision process that determines what should be done. Control concerns the process that assures the planned results are obtained.[9] Different frameworks are available. Vissers et al describe an elaborate Production Control Framework for health care with five levels from patient scheduling up to strategic market choices.[21] The framework assumes a homogeneous product range and a primary process geared to this product range. In hospitals this can be found in 'focused factory' types of hospitals, for example for oncology radiotherapy. Our research, however, concerns general specialist outpatient clinics in hospitals that service a heterogeneous product range.

The number and distribution of the type of appointments (the 'appointment mix') and the number of urgent or non-urgent appointments is uncertain. To deal with this uncertainty, specialist outpatient clinics need demand forecasting and deal with the remaining uncertainty by operational

planning, almost real time.[9] The operational planning of a specialist outpatient clinic should include short-term planning and it should react to the situation as it occurs, which we call reactive decision-making.[9] The framework of Vissers et al does not support this requirement. We use the framework of Bertrand to investigate the characteristics of specialist outpatient clinics as a preparation to develop a Production planning and control (PPC) framework for specialist outpatient clinics (chapter 8).[22]

The framework of Bertrand makes a distinction between:

1. Aggregate Production Planning: the medium/long term matching of available resource capacity with the required capacity. In specialist outpatient clinics this concerns mainly the scheduling of sessions several weeks or months ahead.

2. Operational Production Planning: the coordination and matching of available resource with demand at the detail level. This involves the timing of work orders or, in other words, the periodic assignment of available resources capacity to individual products. In specialist outpatient clinics this concerns the scheduling of appointments with individual patients.

Both aggregate and operational planning need to deal with three typical characteristics of specialist outpatient clinics. On the aggregate planning level, opening or cancelling sessions determines the availability of the specialist's time. Specialists are in a hospital a 'shared resource'. A shared resource usually refers to a resource that serves two different product families.[1] Specialists can work as a shared resource in that sense, for example working in the specialist outpatient clinic as well as the emergency department. Furthermore, they can also work in different departments for the same patient (or more general, the same 'product family'), for example in the specialist outpatient clinic, in the operating theatre and in the clinical ward. This means that there can be conflicting forces from different departments that try to have the scarce resource, time from the specialist, more available for their own process.

Secondly, a large part of the appointments need to be booked with a specific specialist. This concerns follow-up appointments and specific referrals. Since most patients have two or more appointments the specialist specific appointments usually account for more than half of the appointments, making the majority 'single queues' (paragraph 2.6).

Thirdly, specialists are a 'resource' that has a mind of its own to determine when to take a holiday for example or to work extra sessions. Considering the second characteristic, this has direct impact on the availability of capacity, especially for the follow-up appointments and specific referrals, which are single queues for a resource that might be limited available in a specific period. Both types of single queue and shared queue appointments are mixed in the same sessions, creating a complex mix of scheduling processes to manage.

These characteristics of specialists as a 'resource' complicate the realisation of sufficient supply of appointments, because there are multiple and often conflicting interests that need to be weighed when 'releasing capacity based on a system status'. See Figure 2.6 for a basic diagram of the scheduling process of sessions in specialist outpatient clinics and Figure 2.7 for a more detailed process diagram. Capacity is released on an aggregate level in Figure 2.6 at 'Schedule sessions' and

on an operational level at 'Schedule patients'. The complication to realise sufficient supply is relevant at 'Doctor communicates absence' and 'Cancel sessions' in Figure 2.7.

In Figure 2.7 several boxes are inputs for both scheduling activities:

- 'Decision rules'. This refers to the formal or informal rules that are applied when scheduling sessions or patients. A decision rule can be for example to open an extra session if demand rises above a certain level. Another rule can be that only General practicioners are able to schedule priority appointments.
  'Blueprint': this is the basic agenda for each session and for the number and type of sessions in a week. The basic agenda is commonly used and states for example that a particular session always starts with a new appointment, followed by three follow-up appointments etc. There is usually also a blueprint for the number of general sessions with mixed pathology and a number of specific sessions for particular pathology.
- 'Information': any information that is used to decide how to schedule either sessions or appointments. For example information on the current delay to the clinic.

On the operational planning level it is the private life and preference of the patient that determines when the appointment is scheduled and thus when the appointment is released. For example, at a certain access time a decision rule could be to book appointments further in the future to free up time for the specialist to perform more surgeries in the operating theatre.

A third subject of interest to study how a specialist outpatient clinic matches supply and demand is the way a specialist outpatient clinic designs the blueprint for the sessions. There the decisions are made that determine how capacity will be either scheduled or released. There are two types of blueprints:

1. The schema for the outpatient sessions for each week. Which specialist does how many of what type of sessions: general sessions, or for specific pathology etc.
2. The schema that determines at what time what type of appointment can be booked. For example first a new patient, the three follow-up appointment and the blocking of appointments for urgent appointments until 24 hours in advance etc.

What considerations are applied when these schemas are designed? Are the design principles for managing flow applied, for example the relation between capacity usage and exponential waiting times? Are the logistical considerations for single or shared queues applied? This determines the number of different appointment types that are fixed in the session schemas and the sessions with specific types of appointments. Together they determine how vulnerable the system is for fluctuations in the volume of supply and demand.

4.3 Uncertainty, variability and buffering

Paragraph 2.4 discussed the relevance of managing uncertainty in the volume of supply and demand. Regardless of its source, all variability in a production system will be buffered. A fundamental principle of factory physics is that there are three types of variability buffers:[18]

1. Inventory
2. Capacity
3. Time

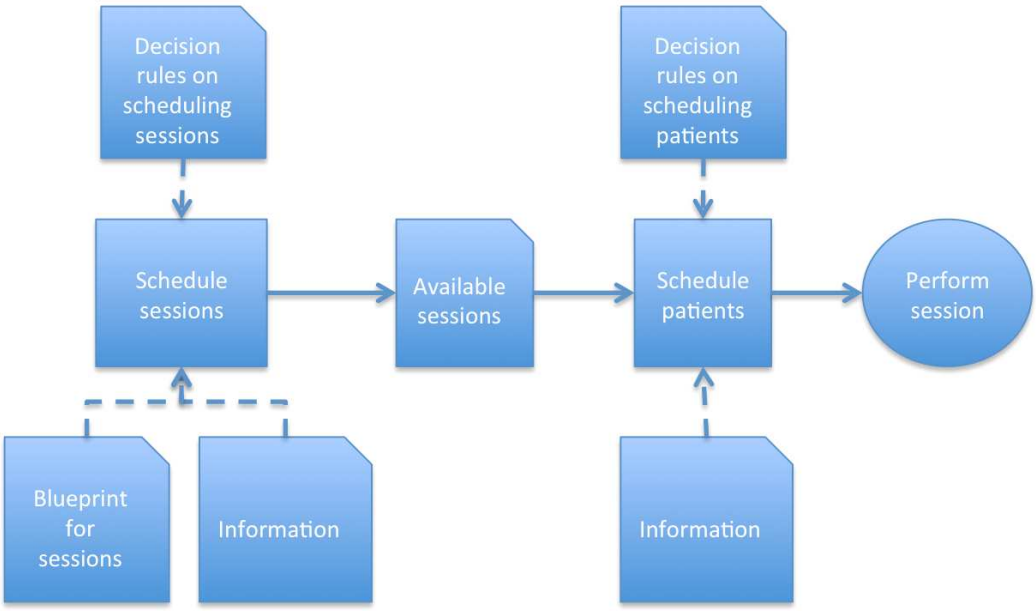


Figure 2.6 Basic scheduling process of sessions and patients

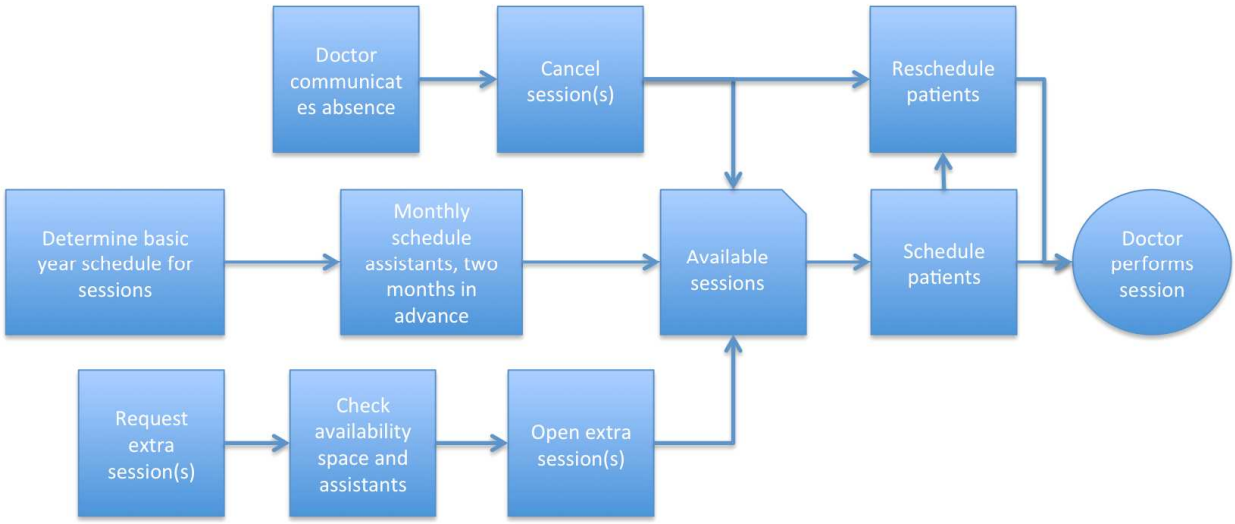


Figure 2.7 Basic scheduling process for sessions and patients in more detail

In a specialist outpatient clinic access time is obviously 'time' as a buffer but it also acts as inventory (WIP), since demand is immediately planned in the 'production schedule'. There is no inventory of half or finished care, since it cannot be 'produced' without demand. If a specialist outpatient clinic offers extra supply and accepts that not all appointment slots in the sessions are always used, that would be using capacity as a buffer instead of time. The challenge when reducing delays in access is to reach an optimal balance between delay in access and utilisation of capacity in order to require as little buffer capacity as possible.

The more demand can be shaped, the more variation can be anticipated, the more unnatural variation in supply and demand can be avoided and the more flexible the clinic is in the way it supplies appointments, the higher the capacity usage can be with minimal delays in access.

4.4 Definition of pull in specialist outpatient clinics

In this and the next paragraphs the pull method from industry as described in paragraph 3 is translated to specialist outpatient clinics. In this paragraph pull is formally defined. On a strategic level for specialist outpatient clinics, the intent of pull is to deliver exactly the care that a patients needs, when he needs it. The production model is output oriented, while a service environment focuses on the interaction with a customer or patient.[23] This is why we stated in the introduction of this chapter that Just-In-Time (JIT) in factories translates ultimately to zero inventory. In specialist outpatient clinics, however, JIT ultimately translates into direct access for patients. In a production environment like a factory, managing inventory and transportation is key to optimise logistics; in a service environment, managing capacity is key.[23] A push model in a specialist outpatient clinic attempts to protect current capacity by pushing work into the future, whereas a pull model will protect the future by pulling all current work into the present. For a General practitioner the appropriate definition of the present may be today, for an Emergency Department it may be this hour, for a specialist outpatient clinic, it may be this week.[5]

The formal definition of pull for a specialist outpatient clinic is different than for a factory, because of several distinguishing characteristics. On a strategic level pull prevents production that is not demanded by a customer. A clinic however cannot 'produce' an appointment ahead of patient demand. The strategic intention of pull to deliver a product or service when a customer needs it, however, is equally, if not more relevant for outpatient specialist clinics. On a tactical level, pull prevents production that is not *at that moment* needed downstream in the production process. Similar to a factory, a clinic can also cause downstream congestion by referring more patients to the next steps than the available downstream capacity can handle (for example Radiology or the Operating Room).

The distinguishing dynamic of pull manifests itself in a factory at the moment the trigger for movement of work is given: a step cannot produce until the downstream step demands it. This is possible because it is known in advance for each product what the sequence of production steps is. In a hospital, the purpose of a visit to a specialist outpatient clinic is often to determine the next step(s). This inherent uncertainty makes it impossible to create a pull system for a specialist outpatient clinic in relation to the other departments in the hospital in the same fashion as described above in the factory example (paragraph 3.5).

The consequence is that, contrary to a factory, the tactical point of interest for releasing the work is not the day of 'production', but when the capacity is made available. If for example the capacity of the operating rooms in a hospital is divided among specialties before actual demand for each specialty is known a push method is applied. If capacity is released based on actual demand (and perhaps other considerations from the system status) a pull method is applied.

We consider the time between making an appointment and having the appointment in a specialist outpatient clinic as 'Work In Process' (WIP), contrary to factories where it is considered 'lead time' and the WIP only starts when production starts. We choose this because the patient is part of the 'production process' and therefore the process starts the moment the demand is expressed. We therefore apply the general definition of a pull system to specialist outpatient clinics in relation to

the triggers that determine the availability of appointments. Our formal definition of push and pull in a specialist outpatient clinic is:

> *A push system* schedules *the availability of appointments in specialist outpatient clinics based on (predicted) demand, while a pull system* authorizes *the release of appointments based on system status.*

Predicted demand refers to any projection of expected volumes of demand per day, week or month of the year. While a pull system might use predictions of demand for long term planning, for example to decide how many people to schedule or give permission to take a holiday, it will not use it to trigger the release of sessions and of appointments.

In conclusion, when the definition of pull is combined with the general characteristics how supply and demand are matched in a specialist outpatient clinic, two levels become relevant:

1. The decision rules and information triggers that determine the scheduling of sessions on an aggregate level.
2. The decision rules and information triggers that determine the scheduling of appointments with individual patients in sessions on an operational level.

4.5 Cycle-mix planning in specialist outpatient clinics

The cycle-mix planning (paragraph 3.8) for specialist outpatient clinics is the combination of the blueprint appointment mix per session and the session mix per week or any other repeating period. The question is how this mix is determined for a specialist outpatient clinic, how well demand can be forecasted and how flexible the outpatient clinic is to react to fluctuations in the actual demand volume and mix and whether demand for either new or follow-up appointments can be shaped to level it out. The more shared queues (paragraph 2.6) an outpatient speciality clinic has, the more potential flexibility there is to absorb fluctuations in the mix of demand. If the supply of an outpatient clinic is dependant on other departments, for example the surgeons that divide their time in the specialist outpatient clinic and the operating theatre, the cycle mix needs to include both the sessions of the operating theatre and of the specialist outpatient clinic.

4.6 Potential benefits of pull

The potential benefits of pull in a specialist outpatient clinic are:[18]

1. *Reduced WIP and cycle time*. In a specialist outpatient clinic this is reduced delay in access.
2. *Smoother production flow*. This depends on he extent to which demand fluctuates compare to supply before a pull system is implemented.
3. *Improved quality*. Better access is likely to lead to less discomfort and may prevent worsening of the condition of patients while waiting.
4. *Reduced cost*. Inventory in the sense of waiting patients bears no direct costs on specialist outpatient clinics. There are extra forms of processing waste however caused by longer delays and push dynamics, for example rescheduling patients when the agenda of the patient or the specialist changes and in the case of no shows because the patient went to another hospital, doesn't cancel after his condition improved, etc.

The potential benefits of pull indicate that pull can be useful to optimise delays in access to specialist outpatient clinics.

4.7 The larger hospital system

The specialist outpatient clinics are part of larger hospital systems and influence the matching of demand and supply of other departments and are influenced by them. Figure 2.8 depicts the most involved other departments for patients that enter the hospital at a specialist outpatient clinic and the arrows show the most common flows. 'GP' stands for 'General practicioner', 'POS' stands for 'Pre-Operative Screening', OT stands for 'Operating Theater', IC stands for 'Intensive Care'.

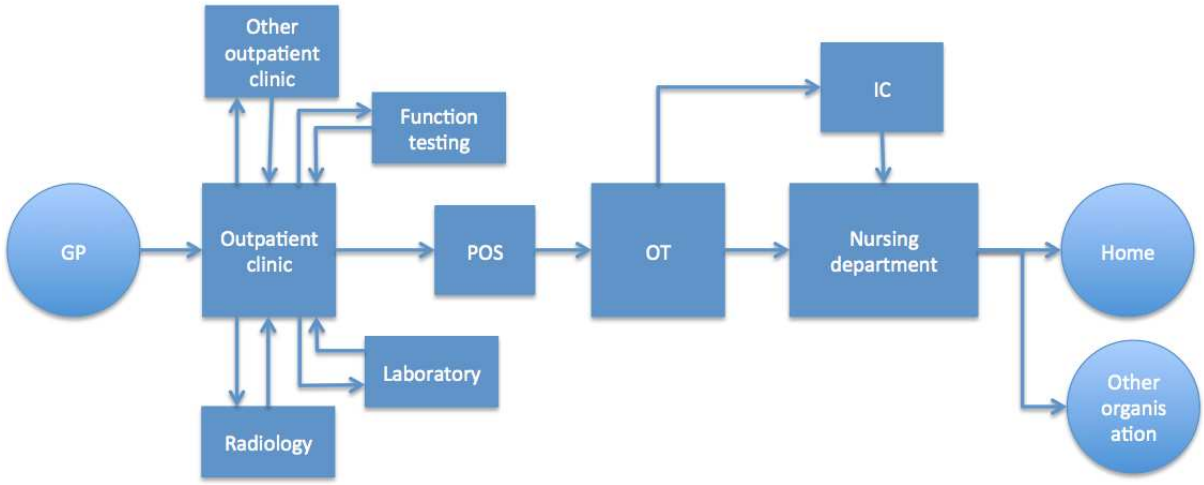


Figure 2.8 Specialist outpatient clinics as part of the larger hospital system

As mentioned before, buffering is a way to cope with variability and is the result of a lack of flexibility. Buffering may however also be the result of a lack of integration: the available capacity is not sufficiently compared to the capacity required for other phases of the diagnoses and/or treatment process.[9]

A complicating factor to manage flow in the system is that for a large part of the appointments, the next step is decided during the appointment. There is a high level of uncertainty for predicting the next step for individual appointments. This increases the need for flexibility to react fast to variation in demand.

In most specialist outpatient clinics specialists are a shared resource with other departments. In all surgical specialties the specialist is a shared resource with the operating theatre, for example General surgery, Orthopaedics, Plastic surgery, Ear Nose Throat (ENT), Urology and Gynaecology. In some specialties the specialists work almost exclusively in the specialist outpatient clinic, e.g. Dermatology or Rheumatology. Some specialists can have variable activities in diagnostic or treatment departments, for example neurologists can also work in the neurophysiology diagnostic department and internists can work in the scope department.

In the larger system, a pull system relates the release of capacity in departments to the status of the total system. If for example in January more patients request an appointment and less operations are needed, a pull system would enable a specialist to perform more sessions and less operations in January and perhaps the other way around in February when those extra patients need to be operated. It requires cross-departmental feedback, flexibility and synchronisation to make pull work in the larger hospital system. It also requires that the medical specialists can trust that the system will make the required capacity become available as soon as it is needed for the patients.

5 Production Planning and Control framework

5.1 Introduction

Paragraph 3.2 explained why the push based Planning and Production Control frameworks MRP, MRP II and ERP do not meet our requirements. Hopp and Spearmann developed a Production Planning and Control framework (PPC) for pull production systems.[1] Paragraph 5.2 discusses this PPC framework. Paragraph 5.3 presents an overview of the variables of specialist outpatient clinics that are relevant to discuss in paragraph 5.4 the potential application of the PPC framework for specialist outpatient clinics.

5.2 A pull based Production Planning and Control (PPC) framework

Figure 2.9 shows the pull based PPC framework.[1] The three levels 'Strategy', 'Tactics' and 'Control' represent the long to short term planning issues of a plant. The rectangular boxes are decisions and the rounded boxes are outcomes. The oval boxes represent inputs that are generated outside this planning hierarchy. Finally, the arrows represent the interdependence of the modules.

On the strategic level, forecasts are used to predict demand and plan capacity and personnel. This results in an 'aggregate plan' with rough predictions about future production mix and volume. On the tactical level, the aggregate plan is transformed to a general plan of action that will help prepare for upcoming production. Here, actual customer demand is used, together with a quota on Work In Process (WIP) to transform the aggregate plan into a 'master production schedule'. Total volume is determined, but the day-to-day schedule with the exact mix and volume per shift is determined later on the control level. On the control level the actual production progress is monitored against the schedule. This provides feedback as input for the capacity plan on the strategic level and input for the WIP position to determine the sequencing and scheduling for the work schedule on the tactical level. Finally, on the control level 'real-time simulations' use 'what if' scenario's to determine the consequences if for example certain customer demands are given priority.

On the aggregate level the availability of the most important resources and the production quantities is determined. This is not the sum of production per type of product, service or activity, but 'aggregated' per type of product, for example x tons of steel. The determination of the production per type of product is determined on the tactical and operational level. In order to aggregate, one ore more common denominators must be chosen, for example money or volume.

The choice of the common denominator(s) is important in order to make reliable aggregate choices. For example, the volume of different types of products often can't be added up to a total volume in products, because each product requires different amounts of production capacity.[24]

5.3 Variables for specialist outpatient clinics

Based on the description in paragraph 4.2 of the general characteristics how specialist outpatient clinics match supply and demand, we provide an overview of questions that need to be answered to optimise delays in access to specialist outpatient clinics. The purpose is to identify the variables and the relations between the variables that need to be considered for the application of the PPC framework, which is discussed in paragraph 5.4.

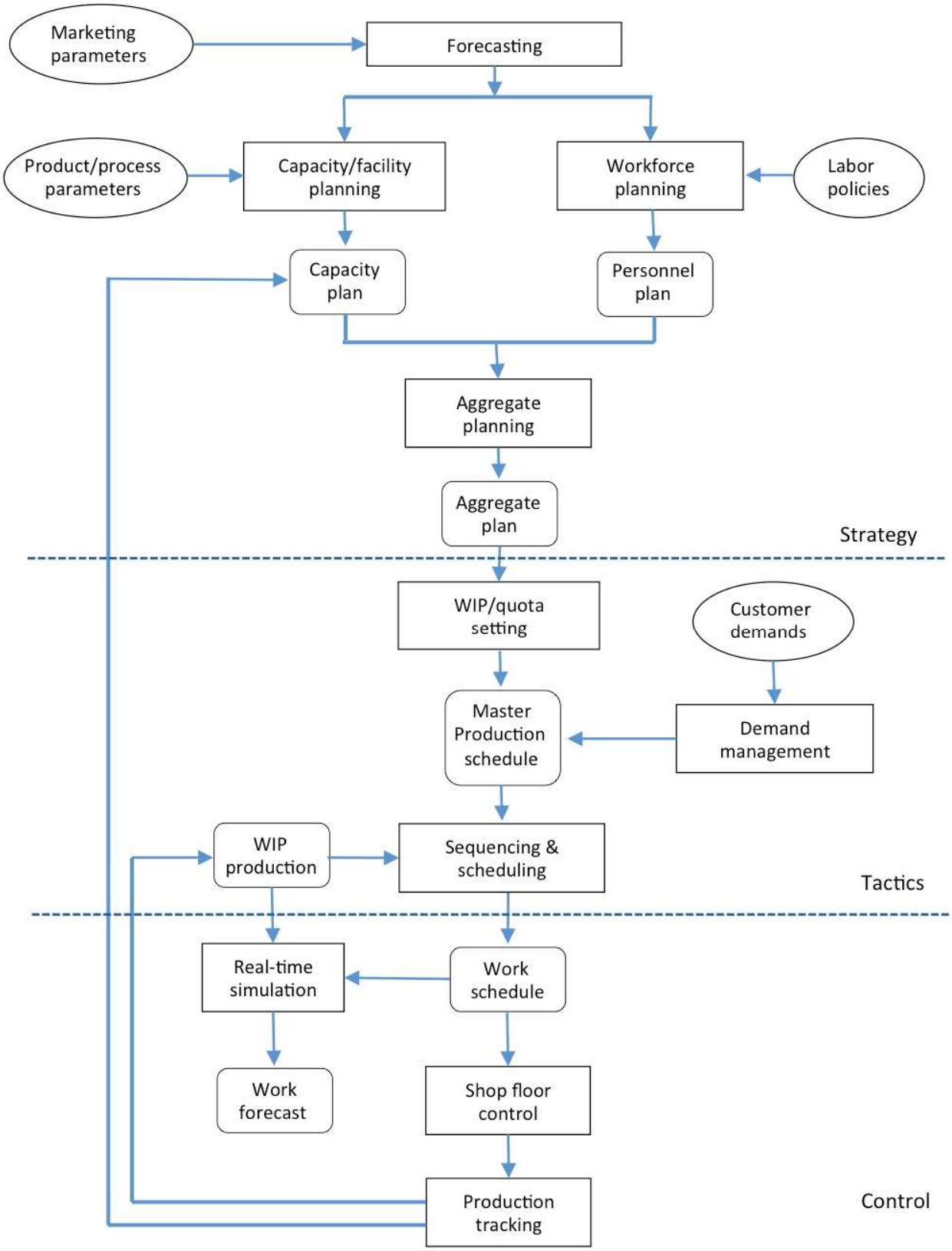


Figure 2.9 Production Planning and Control framework

- What are the norms for delays in access and utilisation levels? How are they monitored?
- What are the methods to predict demand?
- How is the amount of supply determined?
- What are the methods to minimise artificial, unwanted variation in supply?
    - Single / shared queues
    - Mix of sessions
    - Mix of appointments
    - Mix planning cycle
    - Policies to cancel sessions
- What are the methods to shape demand?
    - For new appointments
    - For follow-up appointments
- What methods are used to decide to release capacity of the specialist in the specialist outpatient clinic or the operating theatre? What information is used for this decision?
    - Information on the actual level of demand and supply? How frequent?
    - Other information?
- What are the methods to react to unpredicted variation in supply and demand?
    - What are the conditions that trigger the use of different planning rules?
    - How fast can extra supply be made available?
    - How much extra supply can be made available?
    - Can the mix of appointments be changed?
- How is capacity released? What feedback or feedforward information is used with which decision rules?
    - On an aggregate level: scheduling sessions
    - On an operational level: scheduling appointments
- How are the planning horizons of dependent resources synchronised with the planning of the sessions? (specialists, rooms, assistants, operating theatre etc.)
- What are the methods to evaluate the effectiveness of the way supply and demand are matched?
    - What are the considerations when the way supply and demand are matched have changed in the past?
- What are the methods to improve the effectiveness of the way supply and demand are matched?

5.4 A Production Planning and Control framework for specialist outpatient clinics

In a factory, personnel operate machines and other resources to produce. In outpatient speciality care the producing 'resource' is the medical specialist. When the specialist is not available, other resources because of the specialist-patient relation usually cannot produce the follow-up appointments of his 'production'. This combination of resource and personnel gives the capacity of the specialists specific characteristics which make it a third type of 'resource' that requires planning on the strategic level, in addition to the planning of rooms (capacity and facility planning) and the planning of assistants (personnel planning).

In specialist outpatient clinics demand of patients is usually directly scheduled as an appointment. This has two consequences for the PPC framework. To be able to immediately schedule patients, sessions have to be available beforehand. At least part of the outpatient sessions therefore has to be scheduled before actual demand is known. The 'master production schedule' in the form of sessions to be performed per day of the week has to be determined without the input of actual demand. It can, however, be determined with feedback from 'production tracking'. This would be a 'push' method since capacity in the form of sessions is not released on the actual system status.

A consequence is that 'demand management' is not a step between 'customer demands' and the 'master production schedule'. The 'sequencing and scheduling' step is performed directly with the input of every individual patient demand. 'Demand management' is performed on an aggregate level in the form of decision rules that determine how to schedule patients and definitions of system states that determine which rules to apply. A feedback loop from the control level is required for the personnel to determine how to schedule appointments. Demand management might be required if personnel cannot apply the existing decision rules and a new decision is required. It depends on how the actual system status is used to determine the decision rules to schedule patients and if necessary release extra supply of appointments whether it can be characterised as a push or pull method.

The most important logical aggregate denominator for resource planning on the strategic level is available time for appointments. A risk is that not all changes on the operational level are translated in the aggregate calculations. For example if the production agreements that 2% more patients can be treated are translated in 2% more capacity, this might be inadequate if the appointment mix changes over time. For example the return rate might increase (more follow-up appointments) due to new forms of treatment that require more frequent follow-up. The total increase of capacity might then for example be 5%. It might also be that the variation in demand increases. If the strategic choice is to have the same norms for access and nothing else changes, more buffer capacity is needed to provide the same access.

A question is how to use WIP quota. In a factory WIP quota can be fixed to a strict maximum: for example maximum three parts of inventory between two production units. In hospitals the relation between the steps are often not known beforehand. It is for example uncertain what percentage of patients of a specific session will require surgery. WIP quota can probably not be efficiently used to strictly maximize the number of patients between the outpatient session and surgery, but WIP quota can be used to trigger responses. For example: if the waiting list for surgery reaches 80 patients, an extra operating session is planned and one outpatient session less is planned.

In conclusion, the elements of the PPC framework are present and relevant for specialist outpatient clinics, but it requires adaptions for the specific characteristics.

References

1. Hopp WJ, Spearman ML. Factory physics. 3rd ed. New York, NY: McGraw-Hill/Irwin/Irwin; 2008.
2. Jackson M. Systems thinking: creative holism for managers. Chichester: John Wiley & Sons ltd.; 2003.

3. Kramer NJTA, Smit Jd. Systeemdenken : inleiding tot de begrippen en concepten. 2e herz. dr. ed. Leiden: Stenfert Kroese; 1979.
4. Van Merode GG. Planning en reactie in zorglogistiek: Inaugural lecture; April 19, 2002.
5. Murray M, Berwick DM. Advanced Access: reducing waiting and delays in primary care. Jama. 2003 Feb 26;289(8):1035-40.
6. Walley P, Silvester K, Steyn R. Managing variation in demand: lessons from the UK National Health Service. J Healthc Manag. 2006 Sep-Oct;51(5):309-20; discussion 20-2.
7. Sugiyama, Y., et al., Traffic jams without bottlenecks - experimental evidence for the physical mechanism of the formation of a jam. New Journal of Physics, 2008.
8. Galbraith JR. Organization design. Reading, Mass.: Addison-Wesley Pub. Co.; 1977.
9. Van Merode GG, Groothuis S, Hasman A. Enterprise resource planning for hospitals. International Journal of Medical Informatics. 2004 Jun 30;73(6):493-501.
10. David FJ, Manrodt K. Teaching Service Response Logistics. Journal of Business Logistics. 1992;13(No 2).
11. Bernardes BE, Hanna MD, A theoretical review of flexibility, agility and responsiveness in the operations management literature: Toward a conceptual definition of customer responsiveness, International Journal of Operations & Production Management, 2009;29(1):30 – 53.
12. Molema JJW. Hospital System Design: creating supply flexibility to match demand variability. Datawyse/Universitaire Pers Maastricht. 2009; (Thesis) 11.
13. Litvak E, Long MC. Cost and quality under managed care: irreconcilable differences? Am J Manag Care. 2000 Mar;6(3):305-12.
14. Ashby WR. An introduction to cybernetics. New York,: J. Wiley; 1956.
15. Cassidy D, Holton G, Rutherford J. Understanding Physics. New York: Springer-Verlag; 2002.
16. Åström KJ, Murray RM. Feedback systems : an introduction for scientists and engineers. Princeton: Princeton University Press; 2008.
17. Ohno T. Toyota Production System: beyond large scale production. Tokyo: Productivity, Inc.; 1988.
18. Hopp WJ, Spearman ML. To pull or not to pull: what is the question? Manufacturing & Service Operations Management; Spring 2004;6;2;133-48.
19. Iyer AV, Seshadri S and Vasher R. Toyota Supply Chain Management. A strategic approach to the principles of Toyota's renowned system. New York: McGraw-Hill, 2009.
20. Argyris C, Single-Loop and Double Loop Models in Research on Decision Making. Administrative Science Quarterly sep 1976;21;3.
21. Vissers JMH, Bertrand JWM and De Vries G. A framework for production control in health care organizations, Production Planning & Control 2001;12;6;591-604.
22. Bertrand, JWM and Muntslag DRM. Production Control in engineering-to-order firms. Int J Production Economics 1993;3-22.
23. Davis FW and Manrodt KB. Teaching service response logistics , Journal of business l ogistics 1992(13);2;199-230.

24. Bertrand, JWM, Wortmann J, Wijngaard J. Productiebeheersing en material management. Noordhoff Uitgevers;Groningen; 1998.

CHAPTER 3

Literature review: optimising delays in access to specialist outpatient clinics

1. Introduction

This literature review presents the available knowledge on delays in access to specialist outpatient clinics, the causes of these delays and what is known about the effectiveness of interventions to optimise delays in access to specialist outpatient clinics. Because 'specialist outpatient clinics' is a relatively narrow field we performed the search for specialist outpatient clinics as well as for outpatient clinics in general for the purpose of identifying possible causes and interventions.

2. Method

Our search was performed in 'All databases' of the Web of Knowledge in February 2012. The subject of our review has no unequivocal term. Especially 'delays in access to the specialist outpatient clinic' and 'waiting in the office for the appointment to start' are two distinct problems, sometimes with similar vocabulary. Our search therefore started broad with combined search terms that we identified in several relevant articles that were already known to us. Table 3.1 shows the search strategy. From all approximately 4.000 articles in the second or third column the titles and - if necessary - the summaries were read to determine whether they addressed delays to outpatient clinics. In total 45 articles concerned delays in access and were completely read. From the references of these 45 articles another 85 relevant articles were identified. The combined 130 articles form the content of the results in the next paragraph.

3. Results

3.1 Introduction

Of the 4.000 articles that were identified in the first round, the majority addressed "waiting in the office for an appointment to start". Of the 130 identified articles addressing delays in access to outpatient clinics, the majority concerned the practices of General practicioners and the majority concerned outpatient clinics in the United States and United Kingdom. A handful of relatively small studies concerned delays in access to specialist outpatient clinics and only one concerned specialist outpatient clinics in the Netherlands.[1] More theoretical studies on reducing waiting were published in the 1980's and 1990's and go as far back as the 1950's. Most studies on practical effects on changes were published between 2001 and 2011. We present the results in the scientific logic of understanding he extent and the causes of the problem of delays in access (paragraph 3.2 and 3.3), analysing the problem (paragraph 3.4), possible interventions (paragraph 3.5) and the (long) term effects of interventions to optimise delays in access (paragraph 3.6).

3.2 What is known about the extent of the problem of delays?

This paragraph presents what is known about how long delays in access are, what the perspective of patients is on delays in access, what the relation between delays in access and clinical outcomes is and the relation with efficiency and other relations.

How long are delays in access?

There are no scientific publications analysing the extent of delays to (specialist) outpatient clinics. Three articles refer to (non-scientific) reports. A 2004 United States report investigated access to Cardiology, Dermatology, Orthopaedics and Obstetrics-Gynaecology. The investigators called 12 to 20 physician offices per specialty in 15 metropolitan areas pretending to be a patient. 75% was not

Table 3.1 Search strategy

		1		2		3
	Search term:	Articles	In combination with:	Articles	In combination with:	Articles
1	access times	84000	outpatient clinic	441	-	-
2	waiting times	30000	outpatient clinic	635	access	115
3	delays	668000	outpatient clinic	758	access	68
4	waiting	64000	outpatient clinic	1007	access	0
5	access	386000	outpatient clinic	1258	-	-
6	accessibility	88000	outpatient clinic	663	-	-
7	Advanced Access	16000	outpatient clinic	63	-	
8	Capacity management	-	outpatient clinic	186		
	Search term:	Articles	In combination with:	Articles	In combination with:	Articles
1	access times	84000	outpatient specialty care	66	-	-
2	waiting times	30000	outpatient specialty care	43	access	20
3	delays	668000	outpatient specialty care	0	access	0
4	waiting	64000	outpatient specialty care	0	access	0
5	Advanced Access	16000	outpatient specialty care	0		
6	access	386000	outpatient specialty care	203	-	-
7	accessibility	88000	outpatient specialty care	0	-	-

able to offer an appointment within 2 weeks and 42% was not able to provide an appointment within 3 weeks.[2] The study was repeated in 2009. 57% was not able to offer an appointment within 2 weeks and 31% was not able to provide an appointment within 3 weeks.[3] A Dutch 2002 report states that 25% of the seven largest specialisms in Dutch specialist outpatient clinics could not offer an appointment within 4 weeks.[4] Two Canadian government reports from 2003 and 2005 reported that across all specialties in Canada the median access is also 4 weeks.[5,6]

Some studies that review the results of attempts to improve access provide figures on delays in access before the attempts. These are for obvious reasons not representative, but they are indicative of the extent of delays that might be encountered. In the studies of general practices delays vary from an average of 3.6 days for 462 UK practices to an average of 43 days or 6 weeks for 157 practices in Veterans Primary Care in the United States.[7,8] For specialist outpatient clinics in the United States, United Kingdom and the Netherlands reported delays from approximately 50 specialist outpatient clinics vary from 2 weeks to 8 weeks, with some as high as 19 weeks.[1,9-12] These are similar to the aforementioned delays in access in the Canadian and Dutch reports with some exceptions that have a longer delay in access.

The perspective of patients on delays in access

Eleven studies provide insight in the perspective of patients on delays in access to specialist outpatient clinics.[13-23] A 2006 US study of two internal medicine outpatient clinics with 260 patients and 46 physicians concluded that patients consider access considerably more important, that they want faster access than the physicians think the patients want. Furthermore patients consider their appointment to be more 'urgent' than the physicians.[13] A 2000 UK study of 188 patients of nine surgical specialist outpatient clinics discovered that the actual delay in access had no relation to the perception of the delay.[14] Patients that felt they had waited too long had the same access (median and distribution) as those who found it acceptable. The median access for these 188 patients was 4.8 weeks. Conclusion: the perspective of patients on delays in access is subjective and varies

widely per patient and per situation. In general, however, patients consider fast access not only important, but also more important than providers think.

The relation between delays in access and clinical outcomes

Eight studies analysed the relation between delays in access to specialist outpatient clinics and clinical outcome.[24-31] This did not concern urgent access, but routine access. All studies involved relatively small populations, most often of one hospital. Five publications report a positive relation (mortality, geriatrics, diabetes, haemoglobin A1c level, breast cancer well-being).[24-28] Two studies report no relation (depression, diabetes).[29,30] One study found no relation between clinical outcomes for diabetes and delays in access, but found a relation between continuity of care and clinical outcomes.[31] There are not enough studies to draw general conclusions, but there are indications that there can be some relation, but not a strong relation between delays in access for routine appointments and clinical outcomes.

The relation between delays in access and efficiency

The only relation between access and efficiency has been published on the problem of no shows. Two studies show strong correlations between access to primary care and no shows.[32,33] A US study of 5.901 patients that experienced a median access of 4.3 days had no shows varying from 12 to 42% depending on the delay.[32] A 2004 US study interviewed 34 patients and revealed several causes of no shows: negative emotions about going to see the doctor and perceived disrespect were important issues.[33] The issues became stronger if the delay was longer and if the patient did not understand the scheduling system. Patients were not aware of the logistical consequences and some even thought that a no-show would be a positive event for the clinician and staff because it creates time in their busy schedule. The patients did not consider logistical consequences as a key issue to influence their attendance. One study found no correlation with access, but did find a correlation with provider-continuity, waiting in the office and other factors.[34] It seems that reducing delays in access can decrease no shows, but there is not enough evidence to draw conclusions.

Other relations

We found no scientific publications investigating the perspectives of the society, insurance companies, government or employees.

3.3 What is known about how outpatient clinics match supply and demand?

We identified no publications investigating the way supply and demand are matched by (specialist) outpatient clinics from an Operations Management (OM) or similar perspective. Two publications investigate how assistants book appointments and manage urgent requests in general practices. A conclusion from 1985 is: the larger general practices become, the more complex and more rigid rules are developed to schedule appointments.[35] This leads to more hostile behaviour towards patients and tensions between patients and assistants. A reflection in a 2001 study shows that scheduling appointments in outpatient clinics is a complex social phenomenon.[36]

3.4 What is known about the causes of delays in access?

We identified one publication investigating the causes of delays to (specialist) outpatient clinics. A 1995 UK study aims to address the underlying causes of waiting times for an orthopaedic outpatient service by examining both demand and supply factors.[37] The research revealed

considerable variation in clinic capacity utilisation between consultants and across appointment categories, as well as considerable variation in the capacity templates of the individual outpatient clinics. They observe 'a lack of standardisation in capacity setting, with the size of the clinical team reflecting the individual clinical preferences of the consultants rather than the needs of the orthopaedic service. With such widely different staffing patterns, there are also significant differences in the available skill mix, which further hampers effective capacity utilisation'. The researchers reflect that two core characteristics of such services as specialist outpatient clinics have particular impact:

1. *Perishability*: services cannot be produced in advance to anticipate peaks in demand.
2. *Inseparability*: services typically involve simultaneous production and consumption, introducing an element of "uncontrollability" in the delivery process, in turn impacting the ability to achieve effective capacity utilisation.

They further note that 'Ultimately both clinicians and managers need to attempt to evolve their traditional functional responsibilities and work together to integrate clinicians in the management process'.

A considerable number of 47 publications investigate the effects of improving the way supply and demand are matched by applying the concept of Advanced Access (see paragraph 3.5 for more information on Advanced Access).[1,7-12,21,22,26,27,29-31,34,38-69] None of these studies investigate the causes of the delays in access. One publication on Advanced Access results did however include interviews with General practicioners that provide some clues into the causes of delays as they explain their reluctance to apply the model to improve delays in access.[38] Some physicians fear a reduction of autonomy if demand determines when they need to work and if they need to standardize their practices. Furthermore, some physicians fear a loss of quality when continuity of care is compromised to improving delays in access or when patients can't choose the physician they want.

Noticeably, a US publication on waiting times in the office from as far back as 1952 discusses the 'disproportionally long time which patients are obliged to wait' and observes 'that the requirement that the consultant be kept fully occupied is usually regarded as an over-riding consideration'.[70] The observation can also be posed as a question on the root cause of delays to specialist outpatient clinics. Another perspective is offered in a 1990 US publication, which seemingly proudly concludes that the residents run the specialist outpatient clinic and that they don't mind waiting and delays much because they consider quality and attention more important.[71]

3.5 What is known about possible interventions to optimise delays in access?

We identified approximately 40 publications investigating possibilities to improve the way demand and supply are matched for outpatient specialist clinics, other than with Advanced Access (see further). Most are concerned with reducing the waiting time in the office or reducing no shows. A 2010 UK study investigates the relation between capacity and access to surgical capacity.[72] They note that 'it is rare to find hospitals analysing their results very quantitatively (…) making sparse any real data describing how capacity is in practice tailored to demand'. They state that it is important to know the variation in demand to determine the required capacity and the level of acceptable waste of capacity as a trade-off for fast access.

The 1995 UK study mentioned in the former paragraph that investigated the underlying causes of waiting times for an orthopaedic outpatient service recommends 'if outpatient service is to be managed as a flexible system operating at increased capacity levels, the need for detailed, accurate and timely information becomes paramount'.[37] Considering the 2010 quote on the 'lack of any deal data',[72] this advice has not been picked up yet, at least not in the NHS.

Out of the 130 articles that we selected for this literature review, we identified 22 publications investigating scheduling techniques to optimise the matching of supply and demand.[73-92] As mentioned before, most are aimed at reducing waiting in the office or reducing no-shows. Related to optimising delays in access, one study proposed dynamic scheduling techniques for outpatient clinics with a high service load, where a dynamic percentage of slots is kept open, taking a chance of no-shows in consideration.[89] Another study notices that two reasons complicate managing supply and demand in specialist outpatient clinics:[87]

1. The need to reserve capacity for urgent appointment requests that must be treated soon after they occur.
2. The need to realise high utilisation of more-expensive specialists' time.

They further note several factors that affect the ability of appointment schedulers to utilise available providers' time efficiently and effectively: inter-arrival and service time variability, cancelations and no-shows, patient preferences for a particular day of the week, time of day and for certain physicians, degree of flexibility by the physicians in the use of their time, constraints posed by provider preferences for the sequence and mix of cases each day, appropriate level of information technology and a smooth running call center. They observe that capacity is soft: service providers can vary available capacity to a certain degree by working faster, double booking and working extra hours. They also observe that most studies optimise delays in access to outpatient clinics or waiting times in the office, but rarely both simultaneously, even though they are connected. To improve access and waiting they propose to define access rules determining how much capacity is needed for each type of appointment and for future callers with more urgent needs. For scheduling, they make a distinction between three delivery environments:

1. The single batch process: appointment scheduling decisions are not made until after observing all demand for a session, with the result that inter-arrival times are irrelevant. For example scheduling appointments every time 50 requests are accumulated.
2. The unit process: booking requests occur one at a time and at random time epochs. This is commonly assumed for primary and specialty care.
3. The periodic process: appointment requests are accumulated over discrete time periods. For example scheduling appointments once a day.

They recommend the pooling of supply in specialty care and surgery environments. They further recommend daily scheduling to take care of deviations from planned clinic time and booked appointments, for example by specialists serving as stand-by consultants and providers of emergency care.

Several studies investigate specific scheduling strategies to reduce waiting in the office and to reduce the level of wasted capacity of no-shows, e.g. by overbooking a certain percentage.[91,92] Only one study investigates planning strategies for specialist outpatient clinics: a 2007 study uses a

computer simulation to calculate the required capacity for a Dutch outpatient neurology and a gynaecology clinic.[83] The researchers express their surprise that they did not find any articles describing the use of simulation models to analyse access time. Their model takes variation of demand and supply per weekday into account. The model considers the specialist outpatient clinic as a single queue, which is a strong limitation considering the multiple queues for individual physicians who run personal sessions. The results were implemented by increasing capacity for one and decreasing capacity for the other specialist outpatient clinic. Despite the mentioned limitation the study reports that the goal of proving two weeks access for 95% of the appointments indeed was achieved and that one department created 14% extra capacity.

Several studies from the field of OM make observations that it is considered as disappointing that the substantial gathered scheduling and planning knowledge in their field for over 30 years is almost not applied in health care practice at all.[80,87] One study was not able to test the majority of the developed ideas because the doctors refused to cooperate.[80] They concluded that incremental approaches are required that test small changes for improvement. They state that 'despite many published theoretical work, the impact on outpatient clinics has been very limited'.[80]

Advanced Access

Since 2001 the number of publications in relation to delays in access to outpatient clinics suddenly strongly increases. They almost all concern the application of Advanced Access as a strategy to improve access to outpatient clinics.[56,93,94] First the publications are only aimed at primary care and later also in specialty care. Noticeably different from most publications before 2000, these publications all come from care providers themselves, often physicians. The articles have a practical nature, with references to queuing theory, industrial engineering, lean thinking, Theory of Constraints and the Toyota Production System, but with little further theoretical substantiation. General practitioner Mark Murray created the model of Advanced Access. The aim of Advanced Access is to eliminate all delays in access and to 'do todays work today'.[93] The basic premise is that this can be achieved by improving the way demand and supply are matched and by increasing the efficiency how supply is delivered.[94] Six elements are described to enable this:

1. Balance supply and demand

   Data is used to calculate the size of the patient population, level of patient demand for visits and number of appointment slots, available. Supply and demand must be (brought into) balance for Advanced Access to work. Demand is measured prospectively: what appointments patients actually ask for (external demand), and what follow-up appointments clinicians actually request (internal demand). Supply is measured as the available clinician time. Balancing supply and demand requires mainly reallocating return appointments in a way that smoothes out the overall flow of demand and physicians will have to adjust their schedules to have the most time available when demand is highest.
2. Reduce the backlog

   One-time extra sessions must be performed to eliminate the backlog.

3. Reduce the variety of appointment types

   The number of appointment types should be reduced to a minimum, for example only two types of appointments with two appointments lengths that can be booked at any slot(s) available. This way the schedule is kept simple and flexible.
4. Develop contingency plans for unusual circumstances

   The contingency plans need to deal with inevitable variations in demand and supply. Strategies can restrict prescheduled appointments or increase capacity. Examples are telephone prescriptions, dividing the work of absent physicians and expanding the use of non-physician clinicians to close the temporary gaps.
5. Adjust demand profiles

   Strategies are aimed at maximising the effectiveness of each visit by e.g. covering multiple issues at one sitting, using telephone or e-mail and group medical visits. The strategies lead to less (individual) appointments.
6. Increase the availability of bottleneck resources

   The bottleneck resource is almost always the time of the clinician. By transferring physicians tasks to others efficiency is maximised.

A 2006 publication discusses performance metrics for Advanced Access from an OM perspective.[46] It is intended to provide performance measures that can help primary care clinic directors monitor and evaluate their Advanced Access implementation. Their approach requires more sophisticated data analysis to capture physician-level variation in appointment demand and the impact how physicians cope with demand variability on the clinic's performance. Their method includes each physician's average daily demand and its variability. Their method is designed for general practices with panel sizes (maximum number of patients that can book appointments with this physician) and is therefore not of our interest for specialist outpatient clinics who do not have fixed panel sizes. Their approach is generalizable however and interesting. Their method tracks panel size and daily demand per physician, empty slots at the start of the day and how much slots are used for appointments of another physician. This enables them to calculate whether each physician has enough capacity on average and on a daily basis and the same for the group of physicians together. Their method also involves the option for each physician to alter his or her blueprint schedule how appointments are to be booked on a regular basis. They also track the percentage of patients that see their own physician and they track the backlog per physician. If their method is applied on a weekly level and uses instead of the panel size the number of patients returning for visits next year, the method could likely be transferred to specialist outpatient clinics. The article concludes that to sustain Advanced Access (1) physicians must be encouraged to flex capacity and (2) excess demand from physicians with full practices can be used to support clinic growth. They also note the importance of recruiting a new physician soon enough to prevent a large backlog to grow and they offer a combination of two indicators that can show trends early (but which are specific for general practices). They also stress the importance of continually updating the capacity flexibility strategy to accommodate changing requirements.

3.6 What is known about the effects of interventions to optimise delays in access?

Publications on the effects of interventions are dominated by research related to Advanced Access. We identified 45 publications on the application of Advanced Access in primary care and a handful on the application in specialist outpatient clinics (sometimes combined) and no publications on

other attempts to optimise delays in access. Most studies involve one outpatient clinic or a small group. In primary care a few larger studies involve in total approximately 1.000 outpatient clinics.[7,8,21,44,51,52] In specialty care one larger study involves 318 specialist outpatient clinics and one involves 18.[1,52] Almost all studies show positive results, sometimes with a critical note. In 2011 in a letter a MD reacted to a critical review on the lack of rigorous studies by in turn criticising the prevailing bias that only rigorous controlled trials have value. The letter states that 'given the difficulty of randomising such organisation practice change, there is a need for more willingness to propose and fund rigorous observational studies'.[96]

In the remainder of this paragraph we present the results of Advanced Access in primary care and in specialty care, criticism of the concept and the identified factors that determine the implementation success and we finish with the findings of a systematic review.

Results of Advanced Access in primary care

A 2004 UK study on the application of Advanced Access at 462 general practices that participated in a large-scale collaboration shows a reduction in mean access from 3.6 to 1.9 days.[7] Two thirds of the practices showed significant improvement. All practices learned and applied Advanced Access themselves. Some physicians noted a trade-off between fast access and continuity of care when patients don't see their own physician. This is surprising, because continuity of care is part of the strategy of Advanced Access to ensure quality, but also to minimise return visits. Apparently, the application of Advanced Access can lead to a pressure that compromises continuity. Constraints cited by the General practicioners to apply Advanced Access include a lack of resources, lack of time to instigate the changes and resistance to change within the practice.

Four studies from 2004 to 2009 study the system wide application of Advanced Access in the 1.826 outpatient clinics of the Veterans Health Administration.[8,11,51,52] The first 2004 study involves four case studies within the 1.826 Veterans Health Administration with 'dramatic improvements' and conclude that the concept of Advanced Access is robust across settings and types of clinics.[11] The 2005 study shows that access improved system wide for 157 general practices from 43 days in 2001 to 16 days in 2005 while the resources in the system did not significantly increase.[8] The Veterans Health System achieved this by training teams from all the practices since 1999. The teams themselves analysed their practices and initiated all the changes.

Results of Advanced Access in specialty care

A 2008 publication is one of the two in depth studies in specialty care, in this case 4 US paediatric outpatient clinics: endocrinology, headache, gastroenterology and ophthalmology.[9] They reported little success and sought to describe the reasons for the failure. They hypothesise that the imbalance between supply and demand was too large. They conclude that their attempts to use access-enhancing strategies were not successful, mainly because the teams were allowed to choose their interventions and they did not give enough priority to substantially increase physician's supply. We note that in this study, contrary to the other studies, the teams were not trained and supported to analyse their own practice, but the change strategy was deployed from the larger organisation. We hypothesise that this might have a substantial influence on the reported failure. A 2004 publication is the second in depth study, in this case a US Rheumatology outpatient clinic.[10] This study reports an improvement of access from 60 to less than 2 days. Patient satisfaction

measure and financial performance improved significantly. They conclude that "using a team approach, we are now able to give the patient the rheumatologic care they want and at a time they want and need it".

Criticism on Advanced Access

Criticism of the concept was expressed in a letter in 2002 in reaction to an article.[95] The author states four critical points against some of the proponents of Advanced Access:

1. 'He states that the proponents of Advanced Access promote the preference of patients to have rapid access, but that they leave out the consequence that this comes at the cost of (a) a lack of choice of a doctor (b) being triaged away from the doctor (c) telephone advise instead of face-to-face consultation and (d) lack of follow-up care and continuity.'
2. 'He does not support the claim that once patients realise they can be seen at any time, they stop booking just-in-case appointments'.
3. 'Demand does not fall.'
4. 'Advanced Access proposes to work smarter, not harder, but it actually is harder work.'

The letter ends with the note that Advanced Access can help some practices, but it 'does not justify the evangelical zeal of some of its proponents, for whom it almost seems to have religious status'.

A 2008 publication on 6 US primary care practices is one of the few that reports little success and cautions for high expectations.[54] 5 of the 6 practices were able to implement Advanced Access and access did improve from 21 to 8 days for 15 minute appointments and from 39 to 14 days for 30 minute appointments. None of the 5 achieved same-day access and delays to access increased in the two years after the implementation from 8 to 11 days for 15 minute appointments and from 14 to 29 days for 30 minute appointments. In 2 practices access became worse than before Advanced Access. They noted that the level of acceptance by physicians was sometimes low and follow-through varied. Perhaps related to this, they observed fluctuations in appointment supply, which stemmed from provider leave of absence due to illness, extended absence, leave of the practice and maternity leave. They also saw an increase of demand due to a substantial regional shortage of physicians leading to 4 of the 5 practices stopping the acceptance of new patients. None of the 5 practices reported substantial improvements in patient- or staff satisfaction and no-show rates.

In a letter Mark Murray, the developer of Advanced Access, reacts to aforementioned critical 2008 publication.[97] Murray notes that the essential measurements of daily demand, supply, activity and panel size are missing and that the backlog did not seem to be completely addressed. In conclusion he considers the study inadequate to understand the dynamic of the outpatient clinics to draw any conclusion on the potential of Advanced Access to address the problems faced by these practices.

Implementation factors of Advanced Access

A 2008 study investigates extensively the implementation factors that determine success.[51] It involved 78 primary care practices and 318 specialist outpatient clinics of the Veterans Health Administration and defined the extent of implementation as the degree to which the innovation is actually being used in daily operations and is practiced as intended. The researchers conclude that the implementation of Advanced Access is influenced by intentional spread activities as well as organisational factors including local management support, staff capabilities and facility context. Two factors showed a strong correlation for both primary and specialty practices: (1) management

support for Advanced Access and (2) clinic team knowledge and skills. Four factors showed a strong correlation for specialty practices only: (1) months working on Advanced Access, (2) availability of Advanced Access resource materials, (3) clinic staff review of Advanced Access performance data and (4) use of consulting physicians. Two factors showed a strong correlation for primary practices only: (1) examination rooms per clinician and (2) patients on wait list.

More in detail they conclude that leadership involves not so much personal advocacy, but, rather, the establishment of supportive structures and processes. This includes the elevation of the visibility of Advanced Access by appointing an oversight body, incorporating Advanced Access into facility priorities, holding managers accountable for improvement-related performance, explicitly designating champions for each clinic area, reporting on Advanced Access progress and targeting resources to remove obstacles to Advanced Access implementation that are beyond the reach of local departments. The team's knowledge and skills concerned mainly the ability to work together to make successful changes. This includes seeking information and effectively using that information to design, test and track process improvements, regularly assess progress and learn from efforts of others. In specialty practices the correlation with the months of applying the concept is interpreted as the change taking time due to the more complicated nature of the intervention in specialty practices. The researchers conclude that it is important to look at both implementation and routinisation for the diffusion of Advanced Access.

These findings were confirmed in a 2009 study from the same researchers in 78 departments of the Veterans Affairs medical centres.[52] In this study they also conclude that the implementation depends on both individual staff and a more complex dynamic of individuals operating within work units in a larger organisation. They again find that successful implementation in both primary and specialty practices is correlated to skills in obtaining and using information to do their work well, test changes and learn from the results. They, however, discovered that success is only correlated in primary practices to teams having authority to do its work, having team members participate in team decisions and listening to all team members. They hypothesise that this is only relevant in primary practices due to a function of differences in team size. In the Veterans Health Administration, primary care practices are much larger than specialty care practices. Surprisingly, problem recognition showed no correlation to success in both primary and specialty practices. They hypothesise that the lack of a relationship with the extent of implementation may signal a threshold effect: when the problem is widely recognised, it is not predictive. They also did not find a strong correlation between implementation and organisational culture. They suggest that this may imply that team effectiveness is more important than the culture of the organisation as a whole.

Several publications provide qualitative studies on the impact of Advanced Access.[38,62,66] In general management and supporting staff favour Advanced Access, but the physicians are more ambivalent. A fear of loss of autonomy is noted, especially because Advanced Access determines when to provide more supply.[38] The doctors consider their time to be limited in flexibility due to other work and private related responsibilities. This confirms the observation from the aforementioned 1995 publication that the clinician preferences determining the level of supply is a major cause of delays.[37] Also the fear of loss of continuity is a major concern for General practicioners. Another concern is the perception of an increase in workload, related to fewer no-shows and an earlier

presentation of minor illnesses. A 2008 publication on 8 case studies of UK general practices concludes that the practices did not apply the model as it was intended.[62] Especially the perception that patients can only receive same-day appointments (which is not advocated by the Advanced Access model) created antagonism to Advanced Access. They further observe that policies and targets provided further incentives to diverge from the model and these factors were compounded by informal organisational behaviours, notably the exercise of discretion, which led to adaptation and dilution of the model.

Systematic review of Advanced Access

A 2011 publication discusses a systematic review of Advanced Access.[64] They included 28 articles describing 24 studies of which 8 studies include the evaluation of the change in access. All the latter 8 studies show improvements in access. In 6 studies where statistical analysis was performed, 5 showed statistical significant improvements in access. Almost no studies showed an access of one day, which is the goal for primary care practices. About 25% reached an access of 48 hours. 11 studies reported the impact on no-shows ranging from 0% to 24% reduction. 5 studies showed a significant reduction and the researchers observed that the baseline rates were as high as 16% - 45% for these studies. 7 studies reported neutral to positive outcomes on visit volume, physician compensation and productivity outcomes. 1 study out of 4 shows a statistically significant improvement of overall patient satisfaction. The researchers note that the qualitative studies have found that the real-world implementations of Advanced Access often focus on same-day access to the exclusion of other core principles that can compromise the patients' experience, for example by decreasing the continuity of care or chronic patients being lost for follow-up appointments for primary care as well as specialty care (because they are not scheduled far in advance anymore). They summarise that proponents of Advanced Access suggest that it reduces waiting times, improves continuity of care and reduced no-show rates. Sceptics point out that Advanced Access is difficult to implement, may instead reduce continuity of care and may leave patients with chronic conditions lost to follow-up.

4. Conclusions

There is little known about the extent of the problem of delays in access to (specialist) outpatient clinics, but there are many indications that a large number of primary and specialty practices deal with longer delays in access than generally desired from the point of view of the patient, at least in the US, the UK and the Netherlands. There is little scientific insight in the way (specialist) outpatient clinics match supply and demand and there is only some scientific insight in the possible causes of delays. The most important seems variability in both demand and supply and the inability to deal with this. The most important cause for that inability may be the common practice that clinical preferences rather than the needs of patients determine the availability of supply on a daily basis.

Knowledge on possible interventions to optimise delays in access seems divided in two separate worlds. On the one hand there is a large body of knowledge available from the discipline of Operations Management (OM) to improve queuing systems like outpatient clinics. Unfortunately for our research the application of this knowledge has mostly been directed at reducing waiting in the office and reducing no-shows, not at reducing delays in access. Also unfortunate is that there is

almost no scientific knowledge on the application of the body of knowledge in real life. It seems that the OM knowledge has too little relation with the practical problems that outpatient clinics face.
On the other hand there is a large body of knowledge on applying the concept of Advanced Access that is aimed at 'redesigning the clinical office', mostly in general practices in the US and the UK, but also some US specialty practices. The evidence shows substantial improvements, but also shows considerable issues. The biggest concerns relate to a fear of loss of autonomy under physicians and a loss of quality of care due to less continuity in the care between the patient and the physician. This is remarkable, since the concept of Advanced Access recommends maintaining continuity. There is some, but scarce evidence of sustainability of the results, mostly one or two years after implementation and there is also some, but equally scarce evidence of decay.

Interestingly, the two worlds of OM and real life practice of outpatient clinics are almost completely unrelated. Even though the developer of the Advanced Access model refers to queuing theory and industrial engineering as a source of his knowledge, this relation is nowhere substantiated. Only one article from OM investigates the theoretical application of Advanced Access scheduling techniques. So there is the theoretical body of knowledge of OM on the one hand and the practical knowledge from Advanced Access on the other with a large gap in between.

In a 2011 publication two researchers from the field of OM published a thought-provoking article in a leading journal of their field.[98] They reflect on the past decades of advancements in the field of OM and conclude that 'academic researchers had only marginal roles in innovations in organisations' and they state that the common approach to deal with well-defined problems and with one or two objectives that 'the findings of such research may have little internal or external validity for facilitating management decisions'. They state that for research in the field of OM to be of more value it must include multiple perspectives by investigating different aspects of the system, by employing different research paradigms with different methods, by using different sources of data, or by using different subsets of the same data. They call this 'triangulation'. They further state that the dominance of the mathematical-modeling paradigm constrained development of the discipline and that the biggest innovation in the field of OM came from the factories of Toyota, not academic research. They finish with four propositions that the use of triangulation can obtain new insights if the domain is wider, the complexity greater, the rate of endogenous or exogenous change is higher or the economic presence and impact across multiple organisations of the system or phenomenon.

The findings of our literature review confirm their analysis of the gap on knowledge in the field of OM and the actual practices to optimise the way demand and supply are matched in specialist outpatient clinics and we recognise their propositions to be of value for our research. In our research we attempt to bridge this gap by relating the outcomes of our field research (chapter 5, 6 and 7) to the body of knowledge of OM (chapter 2) and create a synthesis between practice and theory (chapter 8).

References

1. Van der Voort MR, van Merode F, Berden B. Making sense of delays in outpatient specialty care: A system perspective. Health Policy. 2010 Sep;97(1):44-52.
2. 2004 survey of physician appointment wait times. Merrit, Hawkins & Associates, (accessed February 2012).

3. 2009 survey of physician appointment wait times. Merrit, Hawkins & Associates, http://www.merritthawkins.com/pdf/mha2009waittimesurvey.pdf (accessed February 2012)
4. Department of health (VWS). Waiting times in hospitals (in Dutch). May 2002. http://www.minvws.nl/rapporten/cz/wachttijden_in_ziekenhuizen.asp (accessed 31 Jan 2010)
5. Canada S: Access to health care services in Canada, January to December 2003. Ottawa: Statistics Canada; 2004
6. Canada S: Access to health care services in Canada, January to December 2005. Ottawa: Statistics Canada; 2006
7. Pickin M, O'Cathain A, Sampson FC, Dixon S. Evaluation of advanced access in the national primary care collaborative. Br J Gen Pract.. 2004 May;54(502):334-40.
8. Armstrong B, Levesque O, Perlin JB, Rick C, Schectman G. Reinventing Veterans Health Administration: focus on primary care. J Healthc Manag. 2005 Nov-Dec;50(6):399-408; discussion 9.
9. Britto MT, Byczkowski TL, Anneken AM, Hausfeld J, Schoettker PJ, Farrell MK, et al. Improving access to pediatric subspecialty care: initial failures and lessons learned. Quality management in health care. 2008 Oct-Dec;17(4):320-9.
10. Newman ED, Harrington TM, Olenginski TP, Perruquet JL, McKinley K. "The rheumatologist can see you now": Successful implementation of an advanced access model in a rheumatology practice. Arthritis Rheum. 2004 Apr 15;51(2):253-7.
11. Schall MW, Duffy T, Krishnamurthy A, Levesque O, Mehta P, Murray M, et al. Improving patient access to the Veterans Health Administration's primary care and specialty clinics. Jt Comm J Qual Saf. 2004 Aug;30(8):415-23.
12. Hankinson MT, Faraone D, Blumenfrucht M. Sustained improvement for specialty clinic access. Joint Commission journal on quality and patient safety / Joint Commission Resources. 2006 Mar;32(3):142-51.
13. Barry DW, Melhado TV, Chacko KM, Lee RS, Steiner JF, Kutner JS. Patient and physician perceptions of timely access to care. J Gen Intern Med. 2006 Feb;21(2):130-3.
14. Waghorn A, McKee M. Understanding patients' views of a surgical outpatient clinic. J Eval Clin Pract. 2000 Aug;6(3):273-9.
15. Bar-dayan Y, Leiba A, Weiss Y, Carroll JS, Benedek P. Waiting time is a major predictor of patient satisfaction in a primary military clinic. Mil Med. 2002 Oct;167(10):842-5.
16. Dunn E, Black C, Alonso J, Norregaard JC, Anderson GF. Patients' acceptance of waiting for cataract surgery: what makes a wait too long? Social Science & Medicine. 1997 Jun;44(11):1603-10.
17. Gerard K, Salisbury C, Street D, Pope C, Baxter H. Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences. J Health Serv Res Policy. 2008 Apr;13 Suppl 2:3-10.
18. Hart M. Incorporating outpatient perceptions into definitions of quality. J Adv Nurs. 1996 Dec;24(6):1234-40.

19. Owen SA, Maeyens E, Jr., Weary PE. Patients' opinions regarding direct access to dermatologic specialty care. J Am Acad Dermatol. 1997 Feb;36(2 Pt 1):250-6.
20. Rubin G, Bate A, George A, Shackley P, Hall N. Preferences for access to the GP: a discrete choice experiment. Br J Gen Pract. 2006 Oct;56(531):743-8.
21. Salisbury C, Goodall S, Montgomery AA, Pickin DM, Edwards S, Sampson F, et al. Does Advanced Access improve access to primary health care? Questionnaire survey of patients. Br J Gen Pract. 2007 Aug;57(541):615-21.
22. Sampson F, Pickin M, O'Cathain A, Goodall S, Salisbury C. Impact of same-day appointments on patient satisfaction with general practice appointment systems. Br J Gen Pract. 2008 Sep;58(554):641-3.
23. Thomas S, Glynne-Jones R, Chait I. Is it worth the wait? A survey of patients' satisfaction with an oncology outpatient clinic. Eur J Cancer Care (Engl). 1997 Mar;6(1):50-8.
24. Prentice JC, Pizer SD. Delayed access to health care and mortality. Health Serv Res. 2007 Apr;42(2):644-62.
25. Prentice JC, Fincke BG, Miller DR, Pizer SD. Outpatient wait time and diabetes care quality improvement. Am J Manag Care. 2011 Feb;17(2):e43-54.
26. Radel SJ, Norman AM, Notaro JC, Horrigan DR. Redesigning clinical office practices to improve performance levels in an individual practice association model HMO. J Healthc Qual. 2001 Mar-Apr;23(2):11-5; quiz 5, 52.
27. Solberg LI, Crain AL, Sperl-Hillen JM, Hroscikoski MC, Engebretson KI, O'Connor PJ. Effect of improved primary care access on quality of depression care. Ann Fam Med. 2006 Jan-Feb;4(1):69-74.
28. Henselmans I, Sanderman R, Smink A, Ranchor AV, De Vries J. Waiting in breast cancer care and emotional well-being ('Wachten in de mammazorg en emotioneel welbevinden'). Ned Tijdschr Geneesk. 2010;154:B491.
29. O'Connor ME, Matthews BS, Gao D. Effect of open access scheduling on missed appointments, immunizations, and continuity of care for infant well-child care visits. Arch Pediatr Adolesc Med. 2006 Sep;160(9):889-93.
30. Subramanian U, Ackermann RT, Brizendine EJ, Saha C, Rosenman MB, Willis DR, et al. Effect of advanced access scheduling on processes and intermediate outcomes of diabetes care and utilization. J Gen Intern Med. 2009 Mar;24(3):327-33.
31. Sperl-Hillen JM, Solberg LI, Hroscikoski MC, Crain AL, Engebretson KI, O'Connor PJ. The effect of advanced access implementation on quality of diabetes care. Prev Chronic Dis. 2008 Jan;5(1):A16.
32. Gallucci G, Swartz W, Hackerman F. Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center. Psychiatr Serv. 2005 Mar;56(3):344-6.
33. Lacy NL, Paulman A, Reuter MD, Lovejoy B. Why we don't come: patient perceptions on no-shows. Ann Fam Med. 2004 Nov-Dec;2(6):541-5.
34. Bennett KJ, Baxley EG. The effect of a carve-out advanced access scheduling system on no-show rates. Fam Med. 2009 Jan;41(1):51-6.

35. Arber S, Sawyer L. The role of the receptionist in general practice: a 'dragon behind the desk'? Social Science & Medicine. 1985;20(9):911-21.
36. Gallagher M, Pearson P, Drinkwater C, Guy J. Managing patient demand: a qualitative study of appointment making in general practice. Br J Gen Pract. 2001 Apr;51(465):280-5.
37. Laing AW, Shiroyama C. Managing capacity and demand in a resource constrained environment: lessons for the NHS? J Manag Med. 1995;9(5):51-67.
38. Ahluwalia S, Offredy M. A qualitative study of the impact of the implementation of advanced access in primary health care on the working lives of general practice staff. BMC Fam Pract. 2005 Sep 27;6:39.
39. Aiello K. Open access appointing in Army primary care clinics. Mil Med. 2005 May;170(5):370-4.
40. Belardi FG, Weir S, Craig FW. A controlled trial of an advanced access appointment system in a residency family medicine center. Fam Med. 2004 May;36(5):341-5.
41. Boushon B, Provost L, Gagnon J, Carver P. Using a virtual breakthrough series collaborative to improve access in primary care. Joint Commission journal on quality and patient safety / Joint Commission Resources. 2006 Oct;32(10):573-84.
42. Bundy DG, Randolph GD, Murray M, Anderson J, Margolis PA. Open access in primary care: results of a North Carolina pilot project. Pediatrics. 2005 Jul;116(1):82-7.
43. Cherniack EP, Sandals L, Gillespie D, Maymi E, Aguilar E. The use of open-access scheduling for the elderly. J Healthc Qual. 2007 Nov-Dec;29(6):45-8.
44. Dixon S, Sampson FC, O'Cathain A, Pickin M. Advanced Access: more than just GP waiting times? Fam Pract. 2006 Apr;23(2):233-9.
45. Forjuoh SN, Averitt WM, Cauthen DB, Couchman GR, Symm B, Mitchell M. Open-access appointment scheduling in family practice: comparison of a demand prediction grid with actual appointments. J Am Board Fam Pract. 2001 Jul-Aug;14(4):259-65.
46. Gupta D, Potthoff S, Blowers D, Corlett J. Performance metrics for advanced access. J Healthc Manag. 2006 Jul-Aug;51(4):246-58; discussion 58-9.
47. Kennedy JG, Hsu JT. Implementation of an open access scheduling system in a residency training program. Fam Med. 2003 Oct;35(9):666-70.
48. Knight AW, Padgett J, George B, Datoo MR. Reduced waiting times for the GP: two examples of "advanced access" in Australia. Med J Aust. 2005 Jul 18;183(2):101-3.
49. Kopach R, DeLaurentis PC, Lawley M, Muthuraman K, Ozsen L, Rardin R, et al. Effects of clinical characteristics on successful open access scheduling. Health Care Manag Sci. 2007 Jun;10(2):111-24.
50. Lewandowski S, O'Connor PJ, Solberg LI, Lais T, Hroscikoski M, Sperl-Hillen JM. Increasing primary care physician productivity: A case study. Am J Manag Care. 2006 Oct;12(10):573-6.
51. Lukas CV, Meterko MM, Mohr D, Seibert MN, Parlier R, Levesque O, et al. Implementation of a clinical innovation: the case of advanced clinic access in the Department of Veterans Affairs. J Ambul Care Manage. 2008 Apr-Jun;31(2):94-108.

52. Lukas CV, Mohr DC, Meterko M. Team effectiveness and organizational context in the implementation of a clinical innovation. Quality management in health care. 2009 Jan-Mar;18(1):25-39.

53. Mallard SD, Leakeas T, Duncan WJ, Fleenor ME, Sinsky RJ. Same-day scheduling in a public health clinic: a pilot study. J Public Health Manag Pract. 2004 Mar-Apr;10(2):148-55.

54. Mehrotra A, Keehl-Markowitz L, Ayanian JZ. Implementing open-access scheduling of visits in primary care practices: a cautionary tale. Ann Intern Med. 2008 Jun 17;148(12):915-22.

55. Meyers ML. Changing business practices for appointing in military outpatient medical clinics: the case for a true "open access" appointment scheme for primary care. J Healthc Manag. 2003 Mar-Apr;48(2):125-39.

56. Murray M, Bodenheimer T, Rittenhouse D, Grumbach K. Improving timely access to primary care: case studies of the advanced access model. Jama. 2003 Feb 26;289(8):1042-6.

57. O'Hare CD, Corlett J. The outcomes of open-access scheduling. Fam Pract Manag. 2004 Feb;11(2):35-8.

58. Parente DH, Pinto MB, Barber JC. A pre-post comparison of service operational efficiency and patient satisfaction under open access scheduling. Health Care Manage Rev. 2005 Jul-Sep;30(3):220-8.

59. Phan K, Brown SR. Decreased continuity in a residency clinic: a consequence of open access scheduling. Fam Med. 2009 Jan;41(1):46-50.

60. Pickin M, O'Cathain A, Sampson F, Salisbury C, Nicholl J. The impact of Advanced Access on antibiotic prescribing: a controlled before and after study. Fam Pract. 2010 Oct;27(5):554-5.

61. Pinto MB, Parente D, Barber JC. Selling open access health care delivery to patients and administrators: what's the hook? Health Mark Q. 2002;19(3):57-69.

62. Pope C, Banks J, Salisbury C, Lattimer V. Improving access to primary care: eight case studies of introducing Advanced Access in England. J Health Serv Res Policy. 2008 Jan;13(1):33-9.

63. Rohrer JE, Bernard M, Naessens J, Furst J, Kircher K, Adamson S. Impact of open-access scheduling on realized access. Health Serv Manage Res. 2007 May;20(2):134-9.

64. Rose KD, Ross JS, Horwitz LI. Advanced Access scheduling outcomes: a systematic review. Arch Intern Med. 2011 Jul 11;171(13):1150-9.

65. Salisbury C, Montgomery AA, Simons L, Sampson F, Edwards S, Baxter H, et al. Impact of Advanced Access on access, workload, and continuity: controlled before-and-after and simulated-patient study. Br J Gen Pract. Gov't]. 2007 Aug;57(541):608-14.

66. Solberg LI, Hroscikoski MC, Sperl-Hillen JM, O'Connor PJ, Crabtree BF. Key issues in transforming health care organizations for quality: the case of advanced access. Jt Comm J Qual Saf. 2004 Jan;30(1):15-24.

67. Tantau C. Accessing patient-centered care using the advanced access model. J Ambul Care Manage. 2009 Jan-Mar;32(1):32-43.

68. White B. Starting a revolution in office-based care. Fam Pract Manag. 2001 Oct;8(9):29-35.

69. Witt MJ. Advanced Access works! Improved patient satisfaction, access, and P4P scores. J Med Pract Manage. 2006 Sep-Oct;22(2):107-11.
70. Welch JD, Bailey NT. Appointment systems in hospital outpatient departments. Lancet. 1952 May 31;1(6718):1105-8.
71. Finlay W, Mutran EJ, Zeitler RR, Randall CS. Queues and care: how medical residents organize their work in a busy clinic. J Health Soc Behav. 1990 Sep;31(3):292-305.
72. Pandit JJ, Pandit M, Reynard JM. Understanding waiting lists as the matching of surgical capacity to demand: are we wasting enough surgical time? Anaesthesia. 2010 Jun;65(6):625-40.
73. Bailey NTJ. A Study of Queues and Appointment Systems in Hospital Out-Patient Departments, with Special Reference to Waiting-Times. Journal of the Royal Statistical Society Series B. 1952;14(2):185-99.
74. Rising EJ, Baron R, Averill B. A Systems Analysis of a University-Health-Service Outpatient Clinic. Oper Res. 1973 September 1, 1973;21(5):1030-47.
75. O'Keefe RM. Investigating Outpatient Departments: Implementable Policies and Qualitative Approaches. The Journal of the Operational Research Society. 1985;36(8):705-12.
76. Bean AG, Talaga J. Predicting appointment breaking. J Health Care Mark. 1995 Spring;15(1):29-34.
77. Aharonson-Daniel L, Paul RJ, Hedley AJ. Management of queues in out-patient departments: the use of computer simulation. J Manag Med. 1996;10(6):50-8, 3.
78. Benneyan JC. An introduction to using computer simulation in health care: patient wait case study. J Soc Health Syst. 1997;5(3):1-15.
79. Clague JE, Reed PG, Barlow J, Rada R, Clarke M, Edwards RH. Improving outpatient clinic efficiency using computer simulation. Int J Health Care Qual Assur Inc Leadersh Health Serv. 1997;10(4-5):197-201.
80. Cayirli T, Veral E. Outpatient Scheduling in Health Care: a Review of Literature. Production and Operations Management. 2003;12(4):519-49.
81. Robinson LW, Chen RR. Scheduling doctors' appointments: optimal and empirically-based heuristic policies. Iie Trans. 2003 Mar;35(3):295-307.
82. Lee VJ, Earnest A, Chen MI, Krishnan B. Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases. BMC Health Serv Res. 2005;5:51.
83. Elkhuizen SG, Das SF, Bakker PJ, Hontelez JA. Using computer simulation to reduce access time for outpatient departments. Qual Saf Health Care. 2007 Oct;16(5):382-6.
84. Kaandorp GC, Koole G. Optimal outpatient appointment scheduling. Health Care Manag Sci. 2007 Sep;10(3):217-29.
85. Qu X, Rardin RL, Williams JAS, Willis DR. Matching daily health care provider capacity to demand in advanced access scheduling systems. European Journal of Operational Research. 2007 Dec 1;183(2):812-26.
86. Green LV, Savin S. Reducing Delays for Medical Appointments: A Queueing Approach. Oper Res. 2008 November 1, 2008;56(6):1526-38.

87. Gupta D, Denton B. Appointment scheduling in health care: Challenges and opportunities. Iie Trans. 2008 2008/07/21;40(9):800-19.
88. Muthuraman K, Lawley M. A stochastic overbooking model for outpatient clinical scheduling with no-shows. Iie Trans. 2008 2008;40(9):820-37.
89. Liu N, Ziya S, Kulkarni VG. Dynamic Scheduling of Outpatient Appointments Under Patient No-Shows and Cancellations. Manufacturing & Service Operations Management. 2010 March 20, 2010;12(2):347-64.
90. LaGanga LR. Lean service operations: Reflections and new directions for capacity expansion in outpatient clinics. J Oper Manag. 2011 Jul;29(5):422-33.
91. Qu X, Rardin RL, Williams JAS. Single versus hybrid time horizons for open access scheduling. Computers & Industrial Engineering. 2011 Feb;60(1):56-65.
92. Wang W-Y, Gupta D. Adaptive Appointment Systems with Patient Preferences. M&Som-Manufacturing & Service Operations Management. 2011 Sum;13(3):373-89.
93. Murray M, Berwick DM. Advanced Access: reducing waiting and delays in primary care. Jama. 2003 Feb 26;289(8):1035-40.
94. Murray M, Tantau C. Same-day appointments: exploding the access paradigm. Fam Pract Manag. 2000 Sep;7(8):45-50.
95. Lamb A. Why advanced access is a retrograde step. Br J Gen Pract 2002;52(485):1035.
96. Solberg LI. Advanced Access--fad or important?: comment on "Advanced Access scheduling outcomes". Arch Intern Med. 2011 Jul 11;171(13):1159-60.
97. Murray M. Evaluating open access: problems with the program or the studies? Ann Intern Med. 2008 dec 16;149(12):909-10;author reply 911.
98. Singhal K, Singhal J. Imperatives of the science of operations and supply-chain management. J Oper Manag. 2012;30(3):237-44.

CHAPTER 4

Literature review: improving processes in hospitals with the 'lean'-philosophy

Published in Ned Tijdschr Geneeskd. 2013;157:A5541 as 'Sustainable process improvement with the 'lean'-philosophy'. Marc B.V. Rouppe van der Voort, Frits G.G. van Merode, Henricus G.N. Veraart

Foreword

For the research of this thesis, process improvement is relevant because optimising delays is a form of process improvement. The lean-philosophy is relevant because the concepts of Advanced Access and Working Without a Waiting List (WWWL) that are part of this research are partly based on the lean-philosophy and specifically the pull method that we investigate originates from the same source as the lean-philosophy.

Abstract

Process improvement occurs more and more frequently, especially with the help of the 'lean'-philosophy. This management philosophy improves quality by continuously removing 'waste'.

As a consequence of the interdependence of processes local improvements can cause negative effects elsewhere. An integral system approach is required to prevent this. Several hospitals claim that they are able to achieve this.

Research on process improvement by applying the 'lean'-philosophy reports many positive outcomes defined as increased safety, quality and efficiency. Methodological shortcomings and lack of rigorous evaluations make it impossible to ascertain the impact.

Clear is that the investigated applications are fragmented with an overly focus on the instrumental aspects of the philosophy, a lack of integration in the total system and a lack of attention for the human dimensions.

Process improvement is needed to improve both quality and efficiency of health care. It requires that hospitals develop integral systems that combine methods for process design with continual improvement of processes and associated forms of people management. Vital is that physicians take the lead to manage and improve processes integrally.

1. Introduction

Outpatient sessions, operating theatre schedules and integrated care are examples of care processes that physicians deal with on a daily basis. Health care needs to become more efficient and is changing continuously. Most physicians are involved in efforts to improve process and projects with fancy names like 'Business Process Redesign', 'One-stop-shop diagnostics', 'Working with a waiting list', 'Faster Better, 'Six Sigma', 'Theory of Constraints' and 'lean'. Why is there is so much attention for models to improve processes and what do they contribute?

We answer these questions with the help of three case-stories and a description of the essence of process improvement. We use the 'lean'-philosophy, because this approach is widely applied and still increasing in Dutch hospitals as well as in the United States and the United Kingdom. [1]

2. What is the lean-philosophy?

The 'lean'-philosophy is based on the Toyota Production System of which the effects are proven on quality, safety, flexibility and cost reduction.[2,3] A 1987 study showed that Toyota was building cars with half of the problems in quality, in half of the throughput time, half of the inventory and with half of the people then other car manufacturers. The involved researchers decided to label this 'lean' ('efficient' and 'agile' production').[3] Toyota itself speaks of a system aimed at improving quality by continuously reducing 'waste'.[4] Waste concerns every step that does not add value for the customer and in this approach waste is a root cause of quality problems. Later the lean-philosophy developed into a collection of principles, methods and instruments for the design, control and improvement of processes (Table 4.1).[5] Table 4.2 shows 3 case studies where the lean-philosophy is applied in health care.

Table 4.1. The 'lean'-philosophy in 5 steps[6]

step	action
1	Determine the value of processes from the point of view of the customer or patient: 'what is the question the patients asks?' and 'what is of importance while answering this question?'.
2	Identify the value stream (every step in the process that adds value), and eliminate waste.
3	Create 'flow': no waiting in between steps.
4	Let the customer 'pull' the process: actual demand determines the release of capacity. In contrast to e.g. an allocation of operating theatre capacity to each specialism one year ahead ('push').
5	Continuously improve further with the 'plan-do-check-act' quality improvement cycle.

Table 4.2. Examples of the application of the 'lean'-philosophy in health care

Case 1 Standardisation on the operating theatre

Orthopaedist A convinces his colleagues to standardise the hip and knee surgeries: a standard operating technique, one type of implants and suture material. They further agree to perform the hip and knee surgeries on a fixed day, four on a row. The goal is to make quality predictable, the process controllable and improvable. Furthermore, the goal is to reduce the average length of stay and to reduce the purchasing and inventory costs of materials. The reality is less successful: orthopaedist B is convinced that is technique is superior to the agreed technique and he does not change his technique. The consequence is that the operating theatre personnel is forced to continuously change operating techniques and have two different sets of materials and physical operating setups. This reduces the impact on efficiency and makes further improvement of the process more difficult.

Case 2 Increase of demand for the specialist outpatient clinic

The delay in access to the specialist outpatient clinic of a group of rheumatologists is three months. This creates a lot of hassle concerning the priority of urgent patients and semi-urgent patients. The rheumatologists hear that other specialist outpatient clinics have improved and sustained their delays in access by improving the way they match supply and demand. The rheumatologists temporarily perform extra sessions, change the planning techniques of sessions and appointments and create more flexibility. They improve the coordination of their absence and agree on measures to see more patients in busy periods. After several weeks of a reduced delay in access, it rises again. The long delay is access of the hospitals in the surrounding area lead to an increase of demand for their specialist outpatient clinic. The rheumatologists conclude that they will have to accept a long delay in access for new appointments. They are however able to offer patients fast access for follow-up appointments. This enables them to offer chronic patients control over their follow-up frequency. They plan less often a follow-up appointment and teach the patience to decide themselves when they need to come while offering them fast access. This increases the satisfaction of patients while maintaining the same medical outcomes and it decreases the pressure on the sessions because, remarkably, it reduces the number of follow-up appointments.

Case 3 To substitute or not to substitute in diagnostic pathways

A large heart centre develops 'diagnostic pathways'. Instead of a physician a nurse determines which diagnostic pathway is required for a patient. The type and sequence of diagnostic tests are standardised for each pathway, making the execution very efficient and fast. Usually the patients gets the results within several hours. A mental health care organisation also wants to introduce diagnostic pathways. The psychiatrists resist however. They are sceptical of the expected efficiency. They consider it important that the diagnostic process starts with a wide angle. Instead of substitution they want an increase of expertise in triage. This prevents corrections that are needed when during the execution of the diagnostics tests or treatment the patient turns out to be in the wrong pathway.

3. Application of the lean-philosophy

The lean philosophy distinguishes seven forms of waste.[4] Table 4.3 shows the 7 forms of waste for each case study. The case studies show that the reduction of waste can increase the value for patients. Value for patients is each action that directly contributes to improving the health

Table 4.3 The 7 forms of waste in the 'lean'-philosophy applied on the 3 case studies

Forms of waste	Case study 1: operating theatre	Case study 2: Outpatient specialty clinic	Case study 3: Diagnostic pathways
1. Overproduction*			Triage by a physician that can be performed by a nurse
2. Waiting		Delay in access for follow-up appointments	
3. Transport	Transport of material		Changing the physical layout to accommodate the diagnostic pathways can require less transport
4. Processing	Faster recovery reduces unnecessary steps	Follow-up appointments that add no value	Less planning and coordination because of standardisation
5. Inventory	Fewer different types of materials and instruments		
6. Movement			Changing the physical layout to accommodate the diagnostic pathways can require less movement
7. Defect**	Unnecessary length of stay		A patient in the wrong diagnostic pathway

* overproduction occurs when more care is delivered to a patient, including care that can be delivered by less qualified personnel

** Defects: the effects of steps that do not result in optimal care for this patient

condition of a patient. The commonality in the case studies is that the optimal care process for the patient is leading to determine how the process is designed and how the scarce time of people is facilitated to concentrate on the actions that add value to patients.

System approach

The Lean-philosophy does not work if waste is only removed locally; it requires a system approach. In reality care processes run criss-cross through many departments and multiple processes require time of the same people and resources. This implies that the improvement of one process can cause negative effects on another process. For example the reservation of time slots for urgent CT-scans in the heart diagnostic pathway can cause delays for the diagnostics steps of stroke patients. This is why not only the relations within processes, but also between processes need to be considered.

Managing the system integrally is the essence of process improvement.[8] Figure 4.1 shows the house that Toyota uses to display her production system. It represents the understanding that the parts work together to make the whole function better. If one of the elements is not functioning the whole house stands weak. The company culture is based on the core values of 'respect' and 'kaizen' ('everybody improves everywhere always'). In hospitals this integral approach implies that people from other disciplines or departments have more influence on your processes to manage and improve the total process.[9]

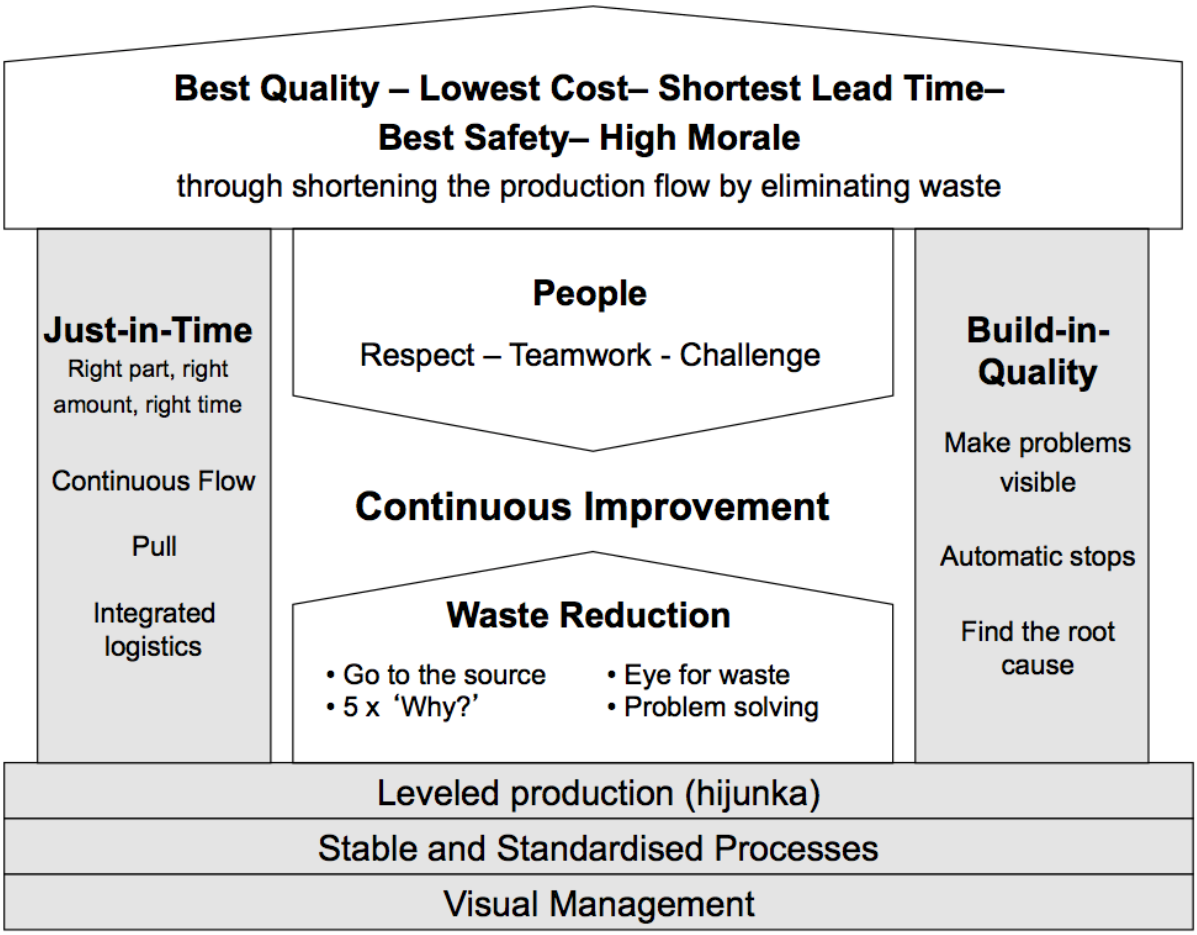


Figure 4.1 The Toyota Production System

4. Is the lean-philosophy effective?

In many sectors in industry it is demonstrated that the application of the lean-philosophy can improve results.[10] A study of 28,000 companies ascribes 23% of the difference in results to the application of the lean-philosophy.[11]

Literature

Several hundreds of studies have been published on the application of the philosophy in hospitals. Most have been performed in the last seven years, mostly in the United States and a handful in the Netherlands. Eight literature studies have been published on the application in hospitals.[5,12-18] The studies show a consistent pattern. Almost all studies report positive outcomes, defined ass safety (for example the occurrence of less infections), quality (for example less waiting) and efficiency (for example less setup times between operations). Because of methodological shortcomings, a lack of rigorous evaluations and the risk of bias it cannot yet be determined what the impact is of process improvement with the lean-philosophy.

Local improvements

Several literature studies conclude that despite the integral approach of the lean philosophy, the application merely results in local improvements within departments or for specific processes and not yet a system impact for the whole hospital.[8,19] This is ascribed to an overly focus on the instrumental application of the philosophy, al lack of integration in the total system and a lack of

attention for the human dimensions.[8] This is partly further explained by the improper application of the philosophy. In Germany for example lean is primarily applied for cost reduction, while loosing most of the original meaning of the philosophy.[20]

A few hospitals claim substantial results for their entire system. Virginia Mason Medical Center ranks in a comparison with 1160 hospitals on both quality and efficiency in the top 1% and ascribes this to the application of lean-philosophy.[21] ThedaCare, a 'health system' that includes five hospitals, claims to have achieved a productivity improvement of 12% in 3 years (in worth of $27 million) and to have reduced delays in access for all specialist outpatient clinics to 1 day.[22] So far however, both hospitals have published only studies that prove partial results. The only Dutch publication that describes the philosophy on the hospital level also claims a hospital wide impact, but has the same methodological limitations that prevent definite conclusions.[23]

5. Reflection

The philosophy will not be more than a hype if the application stays fragmented and easy elements of the deal system are used in isolation and called 'lean' as if they are generic, independently applicable ideas and methods. This will not result in a sustainable process improvement because the underlying problems and the relation between them is insufficiently studied. The random application of lean methods will provide quick results without lasting effect and will eventually result in frustration.

Integral approach

The lean-philosophy is not a toolbox to reduce costs and improve efficiency. Sustainable improvement is only possible with an integral system approach. This concerns the management of the relation between processes, the combination of methods for process design and process improvements with new methods for personnel management that include values like 'respect' and 'kaizen'.[11] Only then synergy is possible and root causes can be addressed without the risk of sub optimisation: changes that result in local improvements but reduce quality elsewhere.

The lean philosophy can be of good service to improve efficiency, but for short-term cost reductions it is not suitable. The development of an integral system approach takes time. Forcing fast results works at the expense of the required climate to integrate process improvement in the routines of the organisation.[5]

Custom-tailoring

Process improvement needs a custom approach. It is important that proven solutions that work elsewhere are not copied, but to search within the own context for possibilities to apply the solution.[24] Senior management will need to enable the creation of a integral management system over a long period of time. It must include rewards that stimulate the coherence of processes. Senior management needs to be enthusiastic but also participate in the application of the philosophy. Put in other words: senior management must manage in accordance with the lean principles.

In contrast to a car manufacturer, where the customer is not part of the process, in a hospital a patient is part of the process. Furthermore, in health care social values and norms have a large influence. The knowledge is owned by the professionals who by education, law and status enjoy a

large autonomy. The power of the lean-philosophy is to use this knowledge. By providing care professionals room to develop their own ideas, changes become better, a custom approach automatically develops and the chance increases that the required behavioural change will occur.

6. Conclusion

Physicians can frustrate process improvement (case study 1), but they can also initiate it and provide leverage to accomplish change. The fact that physicians are highly autonomous but also very dependent on other care professionals bring us to the conclusion that physicians will need to take the lead to develop an integral system. Vital for the success of the application of the lean-philosophy will be whether physicians see the potential for quality improvement by removing waste together to improve processes.

References

1. Benders J, Rouppe van der Voort M, Berden B, redactie. Lean denken en doen in de zorg : acht verhalen uit de praktijk. Den Haag: Lemma; 2010.
2. Womack JP, Jones DT, Roos D. The machine that changed the world. New York, NY: MacMillan Press; 1990.
3. Krafcik JF. A New Diet for United-States Manufacturing. Technol Rev. 1989 Jan;92(1):28.
4. Ohno T. Toyota Production System: beyond large scale production. Tokyo: Productivity, Inc.; 1988.
5. Schweikhart SA, Dembe AE. The Applicability of Lean and Six Sigma Techniques to Clinical and Translational Research. J Invest Med. 2009 Oct;57(7):748-55.
6. Womack JP, Jones DT. Lean thinking: banish waste and create wealth in your corporation. 1st Free Press ed. New York: Free Press; 2003.
7. Merode Fv, Knopper K. Innovatie door specialisatie en toewijding. Gezondheidszorg nieuwe stijl in India. Zorgvisie Magazine. 2012(4):20-2.
8. Liker JK, Morgan JM. The Toyota way in services: The case of Lean product development. Acad Manag Perspect. 2006 May;20(2):5-20.
9. Weick KE, Sutcliffe KM. Managing the unexpected: Resilient performance in an age of uncertainty (2nd ed.). San Francisco: Jossey-Bass; 2007.
10. Shah R, Ward PT. Defining and developing measures of Lean production. J Oper Manag. 2007 Jun;25(4):785-805.
11. Shah R, Ward PT. Lean manufacturing: context, practice bundles, and performance. J Oper Manag. 2003 Mar;21(2):129-49.
12. Poksinska B. The current state of Lean implementation in health care: literature review. Quality management in health care. 2010 2010;19(4):319-29.
13. Nicolay CR, Purkayastha S, Greenhalgh A, Benn J, Chaturvedi S, Phillips N, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical health care. Br J Surg. 2012 Mar;99(3):324-35.

14. DelliFraine JL, Langabeer JR, 2nd, Nembhard IM. Assessing the evidence of Six Sigma and Lean in the health care industry. Quality management in health care. 2010 2010;19(3):211-25.
15. Glasgow JM, Scott-Caziewell JR, Kaboli PJ. Guiding inpatient quality improvement: a systematic review of Lean and Six Sigma. Joint Commission journal on quality and patient safety / Joint Commission Resources. 2010 2010;36(12):533-40.
16. Mazzocato P, Savage C, Brommels M, Aronsson H, Thor J. Lean thinking in health care: a realist review of the literature. Qual Saf Health Care. 2010 Oct;19(5):376-82.
17. Vest JR, Gamm LD. A critical review of the research literature on Six Sigma, Lean and StuderGroup's Hardwiring Excellence in the United States: the need to demonstrate and communicate the effectiveness of transformation strategies in health care. Implement Sci. 2009 Jul;4.
18. Young TP, McClean SI. A critical look at Lean Thinking in health care. Qual Saf Health Care. 2008 Oct;17(5):382-6.
19. Radnor ZJ, Holweg M, Waring J. Lean in health care: The unfilled promise? Social Science & Medicine. 2012 Feb;74(3):364-71.
20. Benders J, van Bijsterveld M. Leaning on Lean: the reception of a management fashion in Germany. New Technol Work Employ. 2000 Mar;15(1):50-64.
21. Kenney C. Transforming health care : Virginia Mason Medical Center's pursuit of the perfect patient experience. Boca Raton: CRC Press; 2011.
22. Toussaint J. Writing The New Playbook For US Health Care: Lessons From Wisconsin. Health Aff. 2009 Sep-Oct;28(5):1343-50.
23. Niemeijer GC, Trip A, De Jong LJ, Wendt KW and Does RJMM. Impact of 5 years of lean six sigma in a University Medical Center. Qual. Manag Health Care. 2012;21;262-8.
24. Vos L, Chalmers SE, La Dueckers M, Groenewegen PP, Wagner C, van Merode GG. Towards an organisation-wide process-oriented organisation of care: A literature review. Implement Sci. 2011 Jan;6.

Part II RESEARCH

In part II the results of three field research studies are presented. The study in chapter 5 evaluates the effects of the application of 'Working Without a Waiting List (WWWL)' ('Werken zonder wachtlijst (WZW)') in 18 specialist outpatient clinics. Chapter 6 describes the results of an evaluation study of the sustainability of the results of the WWWL application. The research questions for the chapter 5 and 6 are:

1. *What changed in the way the specialist outpatient clinics match supply and demand?*
2. *What are the results in reduction of delays in access?*
3. *Which changes are still present three years later and what new changes have taken place?*
4. *What is the delay in access three years later?*
5. *Which factors determined whether the improvements in delays in access were sustained according to the involved actors?*

We test three hypotheses with this research:

1. *Applying the WWWL method enables specialist outpatient clinics to reduce delays in access and sustain their results.*
2. *The specialist outpatient clinics that reduced and sustained their delay in access introduced pull methods.*
3. *Other departments that influence the matching of supply and demand of the outpatient clinics are required to be included in the 'pull' system to optimise delays in access.*

Chapter 7 presents the findings of our third study: an in-depth field research that identifies the characteristics how specialist outpatient clinics match supply and demand. The research question is:

1. *How are demand and supply matched in specialist outpatient clinics?*
2. *What is required to optimise the way demand and supply are matched?*

We test two hypotheses with this research:

1. *The way supply and demand are matched is predominantly a push method.*
2. *The way specialist outpatient clinics match supply and demand causes delays.*

CHAPTER 5

Effectiveness of Working Without a Waiting List

Published in Health Policy, August 2010;97;44-52 as 'Making sense of delays in outpatient specialty care: a system perspective'. Marc (MBV) Rouppe van der Voort, Frits (GG) van Merode, Bart (HJJM) Berden.

Abstract

Objectives

To assess whether delays to specialist outpatient clinics can be solved by improving the way supply and demand are matched, without adding capacity.

Methods

A systematic review of the interventions applied by 18 clinics using the model of Advanced Access and a statistical analysis of the effects of the interventions on their delays.

Results

The specialist outpatient clinics applied different combinations of interventions aimed at improving the way they match supply and demand, improving the efficiency of the way supply is organised and at reducing unnecessary demand. Fourteen clinics show statistically significant improvements. Two probably significantly improved and two clinics did not. Their access reduced on average 55%, from 47 to 21 days.

Conclusions

It seems that delays in outpatient specialty care can be solved to a large extend by improving the way supply and demand are matched. Policy makers should analyse whether delays are caused by capacity problems or matching problems. For the latter, it appears more effective to invest in the ability to react then the ability to plan. Policy makers should create incentives for clinics to keep access short and remove incentives that stimulate delays.

1. Introduction

Accessibility is vital for providing care.[1] In access to specialist outpatient clinics however, delays are a common problem since many years.[2] Almost as common as the delays seems the assumption that they can only be solved by adding capacity. Changes in organization and logistics do not seem to play a major role in policy discussions. In industry the concept of 'pull logistics' has proven to be very successful in solving delays. Basically, in a pull system the actual day to day demand determines what is produced versus a push system where production is based on other criteria then actual demand, e.g. forecasts of demand.[3] The 'pull' concept has been extensively studied in industry and is well recognized for its effectiveness.[3,4] In health care the pull concept has been translated into the model of Advanced Access by general practitioner Mark Murray. The aim of this model is to 'do today's work today'. The basic premise is that this can be achieved by improving the way demand and supply are matched (or: synchronized) and by increasing the efficiency in the way supply is delivered, without adding capacity.[5,6,7]

The validity of the Advanced Access model has been established in several small and one large study for general practitioners in the US and the UK, whereby delays reduced remarkably with 50% or 60% on average.[8-10] In recent years the Advanced Access model has also been applied to a large number of specialist outpatient clinics in several countries. In the Netherlands it has been applied in at least 200 clinics. So far, however, no research has been performed to assess the effects. This raises the question whether delays in access to outpatient speciality care can also be solved with the model of Advanced Access? And if so, what does this imply for policy makers?

This article analyses the effects for 18 specialist outpatient clinics that tested the Advanced Access model to solve their delays without adding capacity. We attempt to make sense of delays by introducing a theoretical framework that describes the causes of delays and requirements to solve them and we place Advanced Access in this framework. We then analyse how the clinics addressed delays and what the effects are and we conclude with implications for policy makers.

2. Theoretical framework

The Advanced Access model interprets a specialist outpatient clinic as a queuing system. A system is a complex whole the functioning of which depends on its parts and the interaction between these parts.[11] A queuing system comprises of an arrival process, a service process and a queue.[3] A specialist outpatient clinic is then considered as a system of servers where patients are seen, with varying rates of arriving patients and varying delays. From a system perspective a delay in access is an aggregate result of these parts not functioning properly together to match supply and demand.

Demand to a clinic is the number of patients asking an appointment, including demand that is generated by the doctor when he tells a patient to book a follow-up appointment. Supply is the number of appointments offered. Advanced Access is based on the assumption that demand and supply are equal on average, but that they are poorly matched, resulting in a continuous delay.[5] The more supply and demand fluctuate with no relation to each other, the longer the delay will be.[12]

To deal with the uncertainty in the distribution of demand, a clinic needs to forecast fluctuations in demand and plan supply accordingly. For the remaining uncertainty it must be able to react to fluctuations.[13] Advanced Access aims to let clinics organize supply more flexibly to increase the

responsiveness ('do today's work today'). The responsiveness of a system can be defined as 'the actions or behaviour of a system using a set of capabilities to purposefully and timely address changes triggered by external stimuli, like a change in demand'.[14] Demand then determines the deployment of capacity, which needs to be flexible enough to adapt. Flexibility in supply can be defined as the extend to which a system is able to create variety in the mix and/or volume of resources in a system, so that the system state can be changed without adding resources or fundamentally changing the resources.[14] Several types of being flexible can be distinguished:[15]

- Mix: ability to produce different products at the same point in time. Advanced Access addresses this with the ability to serve different mixes of appointments in a session.
- Changeover: ability to deal with changes in the product mix over time. In a specialist outpatient clinic this translates into being able to deal with changes in the mix of types of services/appointments over time.
- Volume: ability to easily make changes in the aggregate production amount. Advanced Access addresses this with the ability to increase the number of appointments offered on a single day or week.

To respond to changes in the volume of demand, information is needed. A clinic can use two forms of information: feedback and feedforward. Advanced Access uses the third available appointment to base decisions on when to increase capacity. This is an example of feedback information: information on the present condition of the clinic.[16] Alternatively, feedforward information predicts future effects by using another condition.[16] When the percentage of available appointments for the next six weeks drops below a critical point (due to an increase in demand, decrease in supply, or both) a specialist outpatient clinic can predict that soon the delays will start to grow. Both feedback and feed forward information can be used for managing patient flow. For example the percentage of free agenda space is an example how feedforward information can be used to react faster to changes. Advanced Access advocates this for 'advanced' clinics with short delays.

3. The application of Advanced Access

In 2002 the model of Advanced Access was introduced to Dutch outpatient specialist clinics as 'Working Without a Waiting List (WWWL)' projects.[17] The projects were organised by the Dutch Institute for Health care Improvement CBO as collaboratives using the Breakthrough methodology as developed by the Institute for Health care Improvement in Boston, USA.[18] During these one-year projects multidisciplinary teams learn the model of Advanced Access and the required change methodology. In this study the specialist outpatient clinics that participated in the first three projects between 2002 and 2005 were included.

4. Materials and methods

The definition of access is the third available new and follow-up appointment for a specialist, conform the method of Advanced Access.[5] Access for a specialist outpatient clinic as a whole is measured as the mean access of all individual specialists. The *third* available appointment is measured to prevent that cancellation of appointments on a short notice give an unrealistic presentation. The measure represents how well the outpatient clinic matches total demand and supply.

The specialties of the outpatient clinics under review are: Cardiology(5), General surgery(4), Orthopaedics(3), Internal medicine(2), Rheumatology(2), Neurology(2), Dermatology(2), ENT and Pediatrics. The clinics are part of seventeen hospitals with 240 to 750 beds and 100.000 to 300.000 yearly outpatient visits. Two hospitals are academic. Each clinic participated out of its own initiative and with consent of all involved specialists.

Method

Access was measured weekly on the same day by each clinic using standard instructions from the project management and were collected monthly by the project management. Of the 22 outpatient clinics that participated, 21 provided this data. Three did not apply the interventions of Advanced Access and were excluded since this article assesses the effects of applying the concept. We thus analysed the remaining eighteen outpatient clinics.

The written project presentations and reports of each outpatient clinic were systematically reviewed to create an overview of the applied interventions per clinic and to assess whether the clinics state factors that are relevant to interpret the results. The weekly measurements of access were analysed on statistical significance of improvement.

5. Results

5.1 Applied interventions

Table 5.1 shows the applied interventions during the project for each specialist outpatient clinic. Most outpatient clinics have applied multiple interventions to address the issue that supply fluctuates with little or no relation to fluctuations in demand. Twelve outpatient clinics addressed this with the intervention to forbid cancellations of sessions within six weeks, five clinics (also) addressed it by coordinating the absences of doctors and two clinics started to anticipate fluctuations in supply and demand.

Flexibility in the mix of supply is increased by thirteen outpatient clinics by reducing the number of queues (fixed slots for specific types of appointments, e.g. new, follow-up or urgent). Six outpatient clinics increased flexibility in the volume of supply with procedures to create extra sessions in periods when delays increase. They use measurements of access as feedback information for this. Two clinics use percentage of free agenda space for the next six weeks as feedforward information to react faster to mismatches in supply and demand. No clinic addressed the changeover flexibility of supply.

All outpatient clinics also applied interventions to reduce unnecessary demand or improve efficiency. The clinics reported on average seven interventions (varying from three to thirteen) with a direct impact on matching supply and demand, reducing unnecessary demand or increasing supply, which they applied during their project. Furthermore, they reported on average two other interventions, e.g. improving the flow of sending letters to General practitioners.

Seven outpatient clinics introduced 'patient initiated care'.[19] This creates a demand driven approach to follow-up appointments, which is an application of the 'pull' principle to follow-up appointments.

Most outpatient clinics reported problems they still needed to address, for example creating flexibility to increase supply on a short notice or anticipate seasonal fluctuations, indicating further room for improvement.

5.2 Effects on access

For the eighteen specialist outpatient clinics the average access at the start of the project was 47 days, varying from 15 to 130 days (Table 5.1). At the end of the projects the average access had reduced to 21 days, varying from 1 to 79 days. On average access reduced 55%, varying from 28% to 93%. The average access for new appointments reduced from 45 to 19 days (58%) and for follow-up appointments it reduced from 59 to 29 days (48%). The access of the three outpatient clinics that were excluded because they did not apply the concept of Advanced Access, increased from 45 to 50 days during the project.

Four outpatient clinics reduced access to less then a week as an average for all specialists. Most of the other outpatient clinics did reduce access to within a week for at least one specialist, but not as an average for the whole clinic. One academic specialist outpatient clinic achieved same-day access for the sessions of their residents.

For each outpatient clinic we calculated the average access for new and for follow-up appointments per week for all specialists. The average access then was calculated for each period of four weeks. We tested whether the decrease of access was statistically significant for each outpatient clinic. We tried to estimate a linear model $y=\beta_0 + \beta_1 x$ where x is time, y is access time, β_0 is the constant and β_1 is the regression coefficient. For three outpatient clinics where data was missing between measurements we interpolated data. For two measurements that were missing at the beginning we used the first measurement to replace the missing value (one hospital missed two values, another four). One hospital missed values at the end of the measurement (five values). For this hospital we knew from their report that they kept their access below the last measured value. We used the linear fitting function of *Mathematica*© at a confidence level of 95/100. We believe that with these countermeasures the shortcomings in the data do not significantly influence the results. In Table 5.2 first the access data are given for each outpatient clinic. In the next two columns the parameters of the linear fitting function are given. The first column is the constant (the mean), the second the regression coefficient. The last column gives the p-value. The regression coefficient is significant in thirteen outpatient clinics. For the other five outpatient clinics we analysed from their reports what could explain their non-significant regression coefficient.

The access for outpatient clinic 1 did not change for nine months until they worked down their backlog substantially. They kept access short since then and consider their project a success, but statistically it cannot be proven with these data. Outpatient clinic 9 applied ten interventions and their access did go down slightly and consistently after three periods, but not substantially enough to be significant. Outpatient clinic 12 did not work down their backlog. Their access was already short at the beginning with two weeks. During the project it fluctuated between one and two weeks. They did apply twelve interventions, but statistically the effects are not significant. In their report they do consider it a success. Perhaps access fluctuated above two weeks before the project, but this cannot be proven with the measurements.

In Table 5.2 first the access data are given for each outpatient clinic as an average per four weeks. In the next two columns the parameters of the linear fitting function are given. The first column is the constant (the mean), the second the regression coefficient. The last column gives the p-value. For outpatient clinic 14 the regression coefficient is not significant because of a strong increase of their access to 22 days in period 9. The reason is that during the summer holiday the two

Table 5.1 applied interventions and measurements

Interventions	Clinic 1	Clinic 2	Clinic 3	Clinic 4	Clinic 5	Clinic 6	Clinic 7	Clinic 8	Clinic 9	Clinic 10	Clinic 11	Clinic 12	Clinic 13	Clinic 14	Clinic 15	Clinic 16	Clinic 17	Clinic 18	
Improve matching supply - demand																			Σ
reduce number of queues (less designated slots)			√		√								√				√		13
Reduce number of queues to zero (no designated slots)	√			√			√	√	√		√			√		√		√	
Forbid canceling sessions within 6 weeks	√	√				√	√	√	√		√	√	√	√	√				12
No personal resident sessions and no cancelations																		√	
Pro-actively fill up empty slots by daily scanning	√			√			√		√		√	√	√	√	√		√		10
introduce patiënt-initiated-follow-up appointments	√			√	√	√	√	√						√					7
More flexibility: extra sessions in busy periods	√	√						√	√			√							**6**
More flexibility: longer sessions in busy periods																	**√**		
coordinate absences of specialists further ahead		√		√							√	√					√		5
Plan appointment length more realistic												√	√	√		√			4
Planning of supervision during sessions							√	√	√		√								4
Anticipate fluctiations in supply and demand											√	√							2
increase supply based on '% of free agenda space'		√						√											2
increase flexibility: train employees more allround											√								1
Contingency plans for sudden absence (e.g. sick)												√							1
stop overbooking (two appointment at one slot)																√			1
decrease planning horizon, e.g. 6 weeks or 6 months	√																		1
Spread sessions more evenly over the week								√											1
plan telephone appointments in the sessions																√			1
Reduce demand																			
Review follow-up policies	√	√	√	√		√	√	√			√	√				√	√		12
Intensivate supervision on follow-up policy												√						√	
combine diagnostic appointments (one-stop-shop)				√	√		√		√		√		√	√	√				8
More follow-up by telephone		√	√	√				√	√										5
allow GP's to sent patients directly to small surgery												√							1
agreements GP's: referral and follow-up care								√											1
agreements Home Care, less follow-up appointments				√															1
offer people outside region same access as their region													√						1
only 2nd opinion if all information is available							√												1
Increase efficiency in supply																			
No personal pager during sessions (assistent answers)	√	√		√		√			√	√			√	√		√			10
No emergency pager during sessions	√							√	√				√						
Delegate medical activitites to nurse, NP, PA or assist.		√	√	√	√			√	√	√		√							8
Delegate specialist administrative tasks to assistents		√		√				√		√		√							5
Phone appointments with GP (less disturbance)		√																	1
Total number of interventions (above)	**9**	**10**	**4**	**11**	**4**	**4**	**8**	**13**	**10**	**3**	**9**	**12**	**8**	**7**	**3**	**6**	**5**	**3**	
Number of other interventions with little or no direct impact on efficiency or access	4	-	2	1	2	2	-	6	1	5	2	-	3	2	2	3	1	1	
Measurements																			average
access at the first four weeks	20	30	67	27	130	49	66	43	32	14	40	15	55	42	41	44	64	15	
access at the last four weeks	11	18	46	3	79	24	23	7	23	7	17	7	31	8	29	11	21	1	
change in access	**45%**	**40%**	**31%**	**89%**	**39%**	**51%**	**65%**	**84%**	**28%**	**50%**	**58%**	**53%**	**44%**	**81%**	**29%**	**75%**	**67%**	**93%**	**57%**
no shows at the beginning (%)					4,7	9,5	7	9		10	12	-	4,5	8			10	7	
no shows at the end (%)				2,5	4,2	7,2	4	2		5	6,5	-	2,3	4			10	2	
change in no shows	**-**	**-**	**-**	**-**	**11%**	**24%**	**43%**	**78%**	**-**	**50%**	**46%**	**-**	**50%**	**50%**	**-**	**-**	**0%**	**71%**	**42%**
return rate, beginning project (follow-up:new)				1,9		1,9	1,3		1,7		3,0	0,8	1,9	0,4	1,1	1,4	5,1	2,4	
return rate, end of project (follow-up:new)				1,5		1,5	1,1		1,5		2,5	0,9	1,7	0,4	1,1	1,3	3,8	1,9	
change in return rate	**-**	**-**	**-**	**21%**	**-**	**21%**	**17%**	**-**	**12%**	**-**	**17%**	**-13%**	**11%**	**-10%**	**0%**	**7%**	**25%**	**21%**	**11%**
change in number of patients seen	**-**	**-**	**-**	**17%**	**-**	**-**	**12%**	**-**	**4%**	**-**	**-**	**9%**	**-**	**4%**	**9%**	**-**	**11%**	**12%**	**10%**

Table 5.2 Access change during the project and analysis

Clinic	1	2	3	4	5	6	7	8	9	10	11	12	Mean and regression coefficient	P-value
1	20	23	14	17	18	24	22	19	20	18	11	11	21.787878787878782 - 0.56993006993007*x	0,113563
2			30	33	33	29	25	17	0	21	15	18	35.27272727272726 - 1.58041958041958*x	0,000344
3	67	62	64	57	55	53	42	38	37	52	53	46	64.84848484848483 - 1.9510489510489497*x	0,008812
4	27		19	13	11	22	8	16	6	3	4	3	26.121212121212107 - 2.031468531468531*x	0,00024
5	130	137	136	137	115	119	121	124	121	109	92	79	145.74242424242422 - 4.216783216783216*x	0,000483
6	49	46	52	51	28	17	24	26	39	28	27	24	49.54545454545454 - 2.353146853146853*x	0,01368
7	66	74	73	68	50	45	35	32	35	28	24	23	79.42424242424238 - 5.1293706293706265*x	2,82E-06
8	43	46	36	22	16	10	8						41.575757575757564 - 3.5629370629370616*x	0,000261
9	32	28	29	21	22	18	23	25	22	25	28	23	27.121212121212118 - 0.377622377622378*x	0,274081
10					14	22	16	9	8	6	8	7	17.575757575757567 - 0.8321678321678317*x	0,023387
11	40	37	31	23	20	23	24	22	19	20	19	17	36.06060606060606 - 1.765734265734266*x	0,000367
12	15	13	7	6	5	12	12	7	11	13	8	7	11.030303030303026 - 0.20979020979020965*x	0,478985
13	55	50	46	42	42	40	34	18	24	25	22	31	54.63636363636361 - 2.9055944055944036*x	0,000158
14	42	19	10	14	9	15	8	2	22	11	7	8	24.439393939393934 - 1.6188811188811192*x	0,057498
15	41	22	25	25	22	21	25	32	26	20	17	29	29.348484848484844 - 0.6048951048951051*x	0,274004
16	44	36	43		22	21	15	13	32	19	7	11	43.530303030302996 - 2.9020979020979*x	0,000804
17	64	59	66	64	60	56	52	51	47	36	26	21	75.07575757575758 - 3.8321678321678334*x	2,53E-05
18	15	9	3	1	1	1	2	1	1	1	2	1	8.393939393939394 - 0.8041958041958043*x	0,018498
Average	47	44	40	37	30	30	28	27	28	26	22	21		

physicians involved were not available. However, they were able to reduce access directly after the summer period to 11 and than to 7. If we would correct the linear model for this summer peak by the access time which is the mean of the access times just before and after the summer we would get the following linear model: 24.8939 – 1.88112 x with a p=0.0188793. This confirms that this outpatient clinic successfully applied the interventions. Outpatient clinic 15 reported a strong increase in demand as soon as they worked down their backlog. They reported that their surrounding hospitals have a much longer access up to six months. This explains why their regression coefficient is not significant.

In conclusion, fourteen specialist outpatient clinics show statistically significant improvements. Two probably significantly improved (clinic 1 and 12), but this cannot be proven. Two clinics have not achieved significant improvement for certain (clinics 9 and 15).

5.3 Relation between interventions and effects on access

Several interventions are widely applied, but no intervention is applied by all specialist outpatient clinics, the maximum is thirteen ('Reduce the number of queues'). Also the number of interventions differs widely per clinic, from three to thirteen. Some achieved a large improvement with few interventions (clinics 14 and 18), some achieved a large improvement with many interventions (clinics 4 and 8) and some have much less improvement with many interventions (clinics 2 and 9). The clinics that did not achieve substantial effects in access applied similar interventions as those who did achieve it.

5.4 Other effects

Eight outpatient clinics reported how many patients they treated in the year of the project compared to the year before. They report on average an increase of 10%, varying from 3,5 to 17%. An increase is to be expected since they saw extra patients when working down their backlog. Four

outpatient clinics however reported a substantial increase in demand. One gave up their attempt to solve their delays as a result (clinic 15, see above). One other outpatient clinic also reported that the surrounding hospitals had much longer delays, due to a shortage of specialists. They decided to focus their efforts on solving delays for follow-up appointments. The remaining two clinics were able to deal with the increase of demand with their current capacity, perhaps because of the interventions that increased efficiency. The increase in treated patients makes clear that the reduction of access, at least for these eight outpatient clinics, is not the result from a decrease in demand.

Ten outpatient clinics reported that their 'no shows' (patient does not show up for an appointment) reduced on average from 8,2% to 4,5% (Table 5.1). Twelve outpatient clinics reported that their 'return rate' (follow-up appointments : new appointments) reduced on average 11% from 1,9 to 1,6 (Table 5.1). Several outpatient clinics reported that the more efficient methods of working as well as making the improvements together had a positive impact on morale. One outpatient clinic reported that at the beginning of the project the absence of personnel due to illness was 11% and at the end of the project it was 2% and they attribute this to the project.

4. Discussion

It seems that the concept of Advanced Access can also solve delays in specialist outpatient clinics. The reduction of access of 55% for the eighteen specialist outpatient clinics is comparable to the 50% and 60% that have been documented for General practitioners.[8,9] Unlike General practitioners however, same day access has only been achieved by one clinic for residents and generally does not seem a realistic goal for specialist outpatient clinics.

The specialist outpatient clinics under review have showed that the interventions of Advanced Access can enable them to achieve an access for new and follow-up appointments of less then a week as an average for the whole clinic over a longer period of time, but it seems difficult to achieve. Four outpatient clinics have been able to offer it for over eight weeks or longer on a row (clinics 4, 8, 12 and 18) (Table 5.1). Three other outpatient clinics have achieved one-week access, but have not been able to keep it that short for a longer period (clinics 10, 14 and 16) (Table 5.1).

The first projects started in 2002 and the last ended in 2005. A report from 2002 stated that the Dutch national average access for an appointment varied in 2002 from 21 to 84 days for the seven largest outpatient specialties, which is comparable to the access of the participating outpatient clinics.[2] In 2005 23% of the Dutch clinics could not offer an appointment within 28 days for any of their specialists, indicating that in 2005 accessibility remained a substantial problem in the Netherlands.[20] Another indication is that the three clinics that did not apply interventions did not experience a reduction in their delays, but a small increase.

The clinics reached these results without adding capacity. Two did report a substantial increase in demand due to very long delays in surrounding hospitals. It seems that this is an indicator not to start with Advanced Access or it should be combined with expanding capacity.

The results indicate that, except for the two mentioned above, improving the way supply and demand are matched can solve delays. However, the clinics have also made significant efficiency

gains by reducing demand (e.g. in follow-up appointments) and by improving the way supply is organised. This could have created the needed flexibility in their supply. It does show that it can be done with existing capacity, but perhaps not without realising efficiency gains. It could also be that by applying interventions to reduce delays to less then a week, an outpatient clinic has to make interventions that also make it more efficient. This would be consistent with lessons learned in the automobile industry, where flow with no delays in the production are advocated for the same reasons.[4]

No outpatient clinic applied interventions to shape demand (shifting demand to weeks where expected demand is low and vice-versa), nor did the outpatient clinics learn from historical data to anticipate demand. In general the outpatient clinics did not use more advanced planning methods, but simplified their planning methods and they became better at responding to demand.

Noticeably, none of the interventions are complex, radically new and neither do they require large investments. Why were the problems not resolved before? It seems that the coherent set of principles addressed not only relevant causes but also the relations between these causes on a system level. There is no single cause found for the delays. It is a mix of factors interacting with each other. To solve delays, a bundle of interventions need to be applied, which can be different for each outpatient clinic. A learning environment is needed where the outpatient clinics learn to analyse their situation from a system perspective. This appears to have been an innovative approach for the outpatient clinics.

A relevant factor is that the outpatient clinics appear highly motivated to reach the results. They chose to participate in the projects. They also received attention and support in the projects. They might not have been able to reach the results without this. That does not change the conclusions that the delays can be solved by changing the way demand and supply are matched, but could imply that the social aspect of introducing and applying the required interventions is an important factor to enable outpatient clinics to change. A qualitative study could analyse the relation between behavioural effects and the reduction of the delays.

What will happen in the future if demand increases (e.g. demographically)? Will they increase their capacity accordingly before their delays start to grow? Will the insurance companies contract more care when the delays are still short? These are questions concerning the larger system dynamics that have not been addressed during the projects, but are likely to determine the sustainability of the results. Furthermore, the contexts of the clinics continuously change. Will the clinics be able to keep applying the interventions over time or will they reintroduce the causes of the delays (e.g. create new queues)? Will they be able to translate the principles into new interventions when changing circumstances require this? Further research of their results over a longer period of time is required to ascertain the sustainability.

Perhaps the biggest challenge for the clinics was to address their basic assumptions that demand is unpredictable and insatiable and that delays are needed to work efficiently. When delays are approached as an aggregate outcome of the way demand and supply are matched, specialists and management together seem to be able to solve this problem. The lessons from the eighteen specialist outpatient clinics indicate that from a system perspective, this can be achieved by reducing uncertainty in the levels of supply and demand, reducing complexity in the way the

appointments and sessions are planned, increasing responsiveness to fluctuations in demand with feedback and feedforward information and increasing flexibility in the type and number of appointments that can be offered per week. The interventions to achieve this differ per context of each outpatient clinic and can only be discovered by analysing the specific conditions, applying a variety of interventions and learning from the effects until an effective approach emerges.

5. Conclusion

The results imply for policy makers that they should consider that delays in access might often not be a capacity problem but a system design problem resulting in a poor matching of supply and demand. To match demand and supply effectively and efficiently, policy makers should aim to completely solve delays in access to specialist outpatient clinics, to less then a week. Capacity should be made available (and not necessarily increased) before delays grow, based on increase of demand, not on the increase of delays in access.

Even though both are important, it appears more effective to invest in increasing the ability to react better then the ability to plan better. We consider the question how to increase the ability to react and on which feedback and feedforward information the most interesting topic for further research. Finally, policy makers should create incentives for specialist outpatient clinics to keep access short and remove incentives that stimulate delays.

Limitations

The clinics were not a random selection, but chose themselves to participate in the projects. Their context could differ from other clinics that did not participate. The results are studied for the duration of the projects. The clinics should be researched several years later to ascertain the sustainability of the approach of Advanced Access. The change in access for the specialist outpatient clinics under research could only be compared to limited information of the access to the other Dutch specialist outpatient clinics.

References

1. Institute of Medicine. Crossing the Quality Chasm. A new health System for the 21st Century. Washington: National Academy Press, 2001.
2. Department of health (VWS). Waiting times in hospitals (in Dutch). May 2002. http://www.minvws.nl/rapporten/cz/wachttijden_in_ziekenhuizen.asp (accessed 31 Jan 2010)
3. Hopp WJ, Spearman ML. Factory physics. New York: Irwin/McGraw-Hill, 2000.
4. Liker JK, The Toyota Way: 14 management principles from the world's greatest manufacturer. New York: McGraw-Hill, 2004.
5. Murray M, Tantau C. Same-day appointments: exploding the access paradigm. Fam practice manag. 2000:7:45-50.
6. Murray M, Berwick DM. Advanced Access: reducing waiting and delays in primary care. JAMA. 2003;289:1035-1040.

7. Oldham J. Advanced Access in Primary Care. Manchester: National Primary Care Development Team, 2001. http://www.acfp.ca/docs/advanced_Access.pdf (accessed 29 Jul 2008).
8. Pickin DM, O'Caithain A, Sampson FC, Dixon S. Evaluation of ' advanced access' in the National Primary Care Collaborative. Br J Gen Pract 2004;54:334-340.
9. Dixon S., Sampson FC, O'Caithain A, Pickin DM. Advanced Access: more then just GP waiting times? Familiy Practice 2006;23:233-239.
10. Murray M, Bodenheimer T, Rittenhouse D, Grumbach K. Improving timely access to primary care. Case studies of the Advanced Access model. JAMA 2003;289:1042-1046.
11. Jackson, MC. Systems Thinking: Creative Holism for Managers. Chichester: John Wiley & Sons ltd, 2003.
12. Walley P, Silvester K, Steyn R. Managing Variation in Demand: Lessons from the UK National Health Service. Journal of Health care Management. September 2006;51(5):309-320. Available from: Business Source Premier, Ipswich, MA. Accessed July 31, 2008.
13. Van Merode GG, Groothuis S, Hasman A. Enterprise resource planning for hospitals. International Journal of Medical Informatics 2004;73(6):493-501.
14. Santos Bernardes E, Hanna MD. A theoretical view of flexibility, agility and responsiveness in the operations management literature: toward a conceptual definition of customer responsiveness. International Journal of Operations & Production Management. 2009;29(1):30-53.
15. Molema JJW. Hospital System Design: creating supply flexibility to match demand variability. Datawyse / Universitaire Pers Maastricht. 2009;(thesis)11.
16. Astrom KJ, Murray RM. Feedback Systems: An Introduction for Scientists and Engineers. New Jersey: Princeton University Press, 2008.
17. Rouppe van der Voort MBV, Janssen SFMM, Minkman MMN, Schellekens WLMCM, Slot S. Results evaluated: Breakthrough approach can shorten wait times (in Dutch). Medisch Contact. 2004;14:546-9.
18. Rainey TG, Kabcenell A, Berwick DM, Roessner J. Breakthrough series guide: reducing costs and improving outcomes in adult intensive care. Boston: Institute for Health care Improvement, 1998.
19. Hewlett S et al, Six year RCT of patient initiated review versus regular physician-initiated follow-up in RA. Arthritis Rheum 2003:S232.
20. National Institute for Public Health and the Environment (RIVM) How long is being waited on hospital care? ('Hoe lang wordt gewacht op ziekenhuiszorg?'). http://www.rivm.nl/vtv/object_document/o7864n34418.html (accessed 31 Jan 2010).

CHAPTER 6

Sustainability of Working Without a Waiting List

Published in Journal of Health Services Research & Policy as 'Sustainability of improvements in access to outpatient specialist care in the Netherlands', 2014, Vol 19(2) 94-10. Marc (MBV) Rouppe van der Voort, Jeroen (JDH) van Wijngaarden, Stan (SFMM) Janssen, Bart (HJJM) Berden, Frits (GG) van Merode.

Abstract

Objectives

To improve access to specialist outpatient clinics without adding capacity, Dutch hospitals applied the concept of Advanced Access. Our aim was to determine whether initial improvements are sustained and to identify the factors that influence sustainability according to the involved actors.

Methods

Qualitative case studies in 14 outpatient specialist clinics. Access measurements at the start, finish and three years after the project were compared. Sustained and new interventions are analysed. Interviews with 52 practitioners analysed with the constant comparative method to identify general factors that influence sustainability.

Results

Eleven out of 14 clinics were able to sustain or further improve their reduced delays; two did not and for one it is uncertain. The outpatient clinics maintained the majority of the interventions and all introduced new interventions. Three generic factors emerged that influenced their ability to sustain the results: increased responsiveness to better match supply and demand; clinical leadership and incentives; a shared belief that they can and should control access together.

Conclusions

Reduction of delays in access can be sustained if the way of thinking and the planning system becomes demand driven and flexible, and care providers experience benefits. Unlike previous studies, senior management support and formal training were not relevant though clinical leadership and informal socialisation was. Making multidisciplinary teams responsible for improvement process appears to be vital.

1. Background

Delays in access to specialist outpatient clinics are a common problem for general hospitals.[1,2] Dutch hospitals have successfully reduced delays by applying the Advanced Access model.[3,4] But are such achievements sustainable, and which factors influence sustainability?

A common assumption is that solving delays requires additional capacity such as more doctors. In industry, logistical principles have been successful in minimising delays without adding capacity, such as the principle of 'pull logistics' in which the actual day-to-day demand determines what is produced compared with a push system where production is based on exogenous criteria.[5]

In health care the pull concept has been partly translated into Advanced Access by Mark Murray. Advanced Access aims to 'do today's work today' by improving the way demand and supply are matched and thus reduce unnecessary fluctuations in supply and increase the responsiveness to fluctuations in the volume of demand.[3] It was translated and introduced to specialist outpatient clinics in the Netherlands as 'Working Without a Waiting List (WWWL)'. Earlier research of 18 specialist outpatient clinics showed they were able to reduce delays by 55% on average, from 47 to 21 days.[4]

The limited research available on the long-term effects of redesigning care processes suggests that initiatives easily decay and loose their gains.[6-8] Organizations often are unable to maintain new work processes, as people soon fall back on old routines.[9] Organizations may also fail to adapt to changing circumstances. Sustainability is therefore an important issue to understand how changes can create long-term desired effects. Buchanan et al.[6] identified several categories of factors that affect sustainability: individual, managerial, leadership, organizational, financial, cultural, political, processual, contextual, temporal and the change substance. Within these categories, there are several key factors: a shared vision, champions, motivation, feedback, and a supporting culture.[10-13] However, according to Savaya et al.[14] 'precise assessments of program sustainability are impossible to make on the basis of existing literature'.

Fourteen specialist outpatient clinics applied Advanced Access between 2002 and 2005. They participated in Breakthrough collaboratives where multidisciplinary teams learn about Advanced Access and the required change methodology with the aim of reducing delays to less than a week.[4] Over the course of one year, they exchange experiences and receive limited guidance to analyse their processes and test interventions.

Our aim was to determine whether the outpatient clinics sustained the reduction in delays three years after the project and sustained their new work practices, and to identify the factors that influence sustainability.

2. Methods

Specialist outpatient clinics were selected from 40 that participated in the Advanced Access Breakthrough collaboratives using the following criteria (based on earlier findings that all statistically significant improvements in these projects were at least 30%[4]): A reduction in delay of at least 30% at the end of their project; standardised data on access at the start and end of their project is available; detailed information is available on the introduced work practises.

We conducted qualitative case studies in 14 specialist outpatient clinics in 12 general hospitals in the Netherlands. The clinics include eight medical specialties (Table 6.1).

Data were collected during the improvement projects and again three years later. To avoid research bias the clinics were at the time of the initial data collection not informed that there would be a follow up three years later. We chose three years as in that time it is probable that outpatient clinics had to deal with new circumstances that can threaten sustainability.

Interviews were conducted with 52 people who are involved in managing access in each outpatient clinic: 17 medical specialists, eight support staff, 17 team leaders or department heads and 10 quality improvement staff. Each informant was asked 35 closed and open questions concerning: the current application of interventions; whether they consider their results sustainable and which factors they consider most influential for sustainability. We asked in detail how measurements are used to match supply and demand, how problems are addressed, who is responsible to keep delays in access low and how new colleagues learn how to match supply and demand. In preparation, we analysed the project reports to identify which interventions were applied during the project. We asked open questions about factors influencing sustainability and also used the categories of factors identified by Buchanan et al.[6] as topics.

To determine the level of sustainability of access we used two data sources. First, each outpatient clinic measured access during three weeks before they were interviewed. Access is defined as the average of the third available appointment for new and for follow-up appointments for each specialist. Second, each respondent was asked what the access had been over the years after the project, whether the current access is representative for that period and if they could provide additional data.

The constant comparative method of qualitative data analysis was used to identify the recurrent factors that influenced the sustainability of the results.[15] Codes were developed based on the review, interpretation and comparison of verbatim quotations. Two researchers performed line-by-line review and coding of the interview transcripts. First, data of four clinics was coded independently and then discussed to negotiate consensus. If consensus could not be reached a third researcher reviewed the data. This resulted in a preliminary list of codes. Second, all transcripts were coded using the preliminary list, and also by adding new codes until no new codes emerged. This list was used to categorize the data. These categories were further analysed and discussed to identify the general factors that influence sustainability.

3. Results

3.1 Access three years after the project

The average access after three years is 18 days, slightly lower than the 20 days at the end of the projects (Table 6.1). Each clinic still has less delay in access than before the projects and for 10 clinics it is more than 30% less.

Respondents from all but two outpatient clinics stated that their current delay in access was representative of the past three years. Those two outpatient clinics stated that access had been substantially better (clinics 12 and 14). Clinic 12 supported this claim with measurements of 50 weeks that showed an average access of 16 days, representing 50% improvement compared to the

Table 6.1 Mean delays (days) in each outpatient clinic at the start and end of the improvement project and three years later

		Delay in access			
Clinic		**Start project**	**End project**	**3 years later**	**Start to 3 years change (%)**
1	Pediatrics	15	1	1	93
2	Rheumatology	43	10	5	88
3	Dermatology	35	3	7	80
4	Cardiology	56	32	13	77
5	Orthopaedics	128	81	45	65
6	Rheumatology	65	44	23	65
7	Neurology	5	1	2	60
8	Orthopaedics	31	17	13	58
9	Orthopaedics	63	20	30	52
10	Urology	21	10	12	43
11	ENT	38	20	28	26
12	Cardiology	32	13	26	19
13	Dermatology	33	19	28	15
14	Neurology	15	8	14	7
Average		**41**	**20**	**18**	**57**

start. Based on the current measurements 11 outpatient clinics (1-10 and 12) seem to have sustained their improved access or even improved it further. Two outpatient clinics (11 and 13) have not been able to sustain their results and for one outpatient clinic it is uncertain (14).

Figure 6.1 shows the specialist outpatient clinics categorised by their access at three years compared to the start: shorter (at least 30% less); similar; longer but still at least 30% better than at the start of the project; longer, similar to the start and longer than at the start of the project.

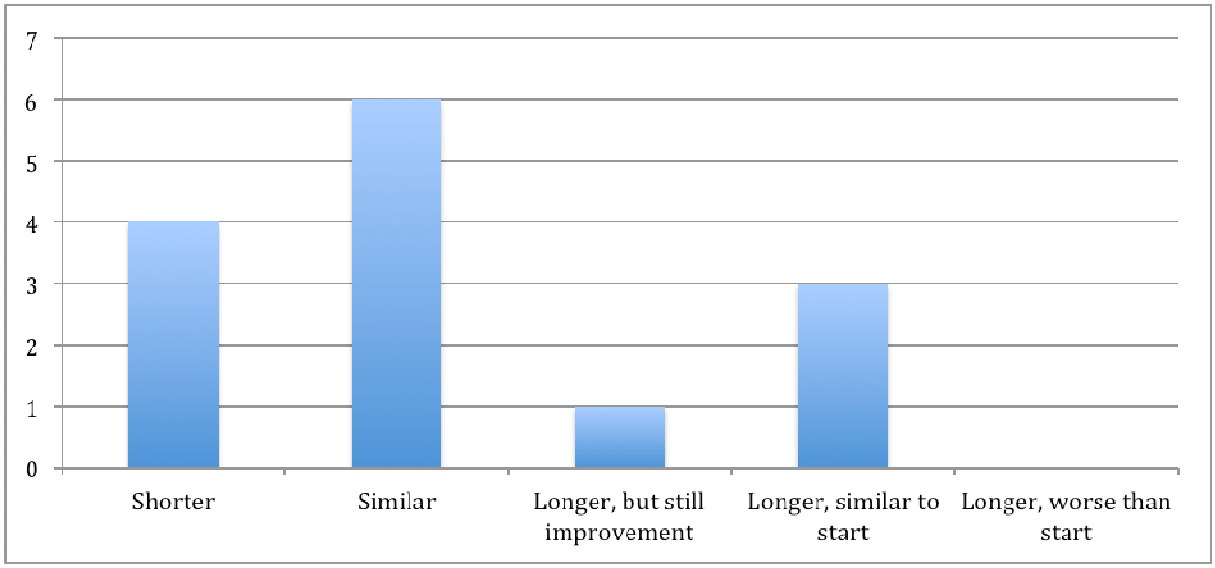


Figure 6.1 Distribution of clinics as regards access after three years compared to access at the end of the improvement project.

3.2 Sustaining work practices or continuous improvement

Table 6.2 displays the interventions introduced during the project and which are present three years later. Ten outpatient clinics were still applying all interventions while four outpatient clinics stopped several interventions, but still apply the majority. All the clinics applied new interventions after their project, on average three per outpatient clinic. Most of these were implemented in the year after the project. No new interventions have been invented. Three years after the project, the results are sustained and the outpatient clinics still apply the new work practices. The respondents state that demand did not decrease since the project and the number of specialists did not increase. So how were they able to reach this level of sustainability?

3.3 Increased responsiveness to better match supply and demand

Some outpatient clinics have sustained a short access with few interventions (e.g. clinics 4, 5 and 6), while others have sustained it with many (e.g. 2, 3) (Table 6.2). The same diversity is apparent for the outpatient clinics whose access has declined: clinic 13 still applies 10 interventions, while clinic 11 applies four. Despite this diversity, each of the successful outpatient clinics increased their responsiveness to better match supply and demand by making access visible and increasing flexibility.

Table 6.2 Interventions per clinic.
√ = intervention carried out during the project (in column 'project') and continued since the project (in column '2008'). N = New intervention after the project. X = Intervention that is stopped after the project, N/X = New intervention that started after the project and stopped again

	1		2		3		4		5		6		7		8		9		10		11		12		13		14	
Interventions	Pediatrics		Reumatology		Dermatology		Cardiology		Orthopedics		Reumatology		Neurology		Orthopedics		Orthopedics		Urology		ENT		Cardiology		Dermatology		Neurology	
	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008	PROJECT	2008
Matching supply and demand																												
1. Weekly measurement of access time	√	X	√	√	√	√	√	√	√	√	√	X	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
2. No cancelling of sessions within 6 weeks	√	√	√	√		N	√	X							√	√	√	√	√	√	√	√	√	√	√	√		N
3. Minimize number of Queues[1]	√	√	√	√	√	√	√	√	√	X	√	X		N		N	√	√	√	√			√	√	√	√		
4. Anticipate fluctuations in demand		N																N							√	X		
5. Short planning horizon				N				N								N		N								N		
6. Create possibility of extra sessions on a short notice		N	√	√		N		N		N			√	√	√	√		N	√	√			√	√		N /X		
Increase efficiency																												
7. Proactively fill empty slots				N	√	√	√	X		N		N				N	√	√	√	√				N	√	√		N
8. Delegation of non-medical tasks		N	√	√	√	√									√	√						N				N	√	√
9. Delegation of medical tasks			√	√	√	√			√	√	√	X		N	√	√								N		N	√	√
10. Appointment length based on measured times				N		N	√	√		N						N		N		N			√	√		N		
Reduce demand																												
11. Patient initiated care for follow-up appointments[2]			√	X	√	√			√	X	√	√					√	√	√	√	√	√				N		N
Number of interventions	3	5	7	9	6	9	5	5	4	5	4	2	2	4	5	9	5	9	6	7	3	4	5	7	5	9	3	6

[1] A queue is a part of capacity that is reserved for a specific use, for example time slots that can only be used to book a new appointment, or a session that is only for a certain type of pathology

[2] Patient initiated care refers to enabling patients to decide when they need a follow-up appointment [20]

Making access visible. Most specialist outpatient clinics closely monitor access by making it visible as a key indicator. Nine clinics still measure access weekly and convert these data into graphs. This not only enables a fast response to increasing delays but also helps to keep individual doctors aware of the consequences of cancelling sessions.

> *"Access needs to be well monitored. That's why I never stop measuring. If there is no monitoring or nothing is mentioned everybody starts to act according to his or her individual agenda." (medical secretary, clinic 2)*

Five outpatient clinics make the access of each specialist visible for everyone to create social pressure.

> *"We measure every week and we mail it to each doctor. He knows his own access, but also that of his colleagues. That creates a nice form of competition." (department manager, clinic 9)*

Two outpatient clinics, both with one week access, are able to respond quickly without measuring access because it is immediately obvious when access increases.

Flexibility of capacity. Many respondents identify flexibility of capacity as a crucial requirement to minimize delays. They distinguish three types of flexibilities. First, flexibility in the number of sessions per week. The clinics look for possibilities to run extra sessions when access increases, requiring extra work.

> *"What we try to do is to run sessions when they are needed… That means working on a free afternoon or day. I think that sustainability is about being conscious of your actions and about being flexible to adapt." (doctor and medical manager, clinic 3)*

In some clinics, the decision to add extra sessions has been delegated to support staff.

> *"We have agreed that an assistant can, without asking, plan an extra session. That is actually a great result in itself." (department manager, clinic 9)*

Lack of flexibility to perform extra sessions is mentioned by one outpatient clinic as a cause why improved delays are not sustained.

> *"We don't perform extra sessions anymore when access increases. I'm frustrated, because we can't offer an appointment to patients when they want it and we feel the pressure increasing." (team leader, clinic 11)*

Flexibility in the mix of appointments per session. Before the projects, all outpatient clinics had a fixed number of slots for each type of appointment. For example, a session always has six new and 12 follow-up appointments. This provided practical advantages. However, if eight patients ask for a new appointment and ten for a follow-up, the number of fixed slots does not match demand. This will either lead to unused slots or a delay in access to ensure fully booked sessions. Almost all clinics made the appointment mix flexible and still do (Table 6.2).

> *"We have changed the schedule for our sessions from fixed numbers of new and follow-up appointments to completely flexible with only five minute slots in which the supporting staff have full control to fill our sessions" (doctor, clinic 9)*

Maintaining buffer capacity. The outpatient clinics maintain buffer capacity to absorb fluctuations in demand. For example two outpatient clinics plan 10% more sessions than anticipated demand. The efficiency gains of their project made this possible.

"When you plan more time than needed on average, you have more flexibility… everything revolves around that." (doctor, clinic 2)

One of the outpatient clinics stopped planning 10% extra capacity, and as a consequence delay started rising.

"(After the project) we started planning a bit more than 100%... then someone started thinking the other way round and it goes wrong again… We don't have enough flexibility anymore." (quality manager, clinic 13)

The efficiency gains enabled the outpatient clinics to meet growing demand. Some however, think they have reached the limits.

"There is more demand this year. When you consider all changes that we have made, there is not much room for further efficiency improvement. It has become a capacity problem. I feel we are really at the edge of our possibilities now." (doctor, clinic 3)

3.4 Clinical leadership and incentives

Increased responsiveness can only be sustained if a support structure is in place, based on clinical leadership and incentives. Making access visible is used not only to monitor demand, but also as an incentive for doctors by creating social pressure and competition. Another important incentive is the direct benefit experienced as a result of a short access:

"The best motivator is the stress level when access increases. It's very important that they feel an effect on them personally (quality manager clinic 3)

When delay increases, the support staff experience the pressure first. General practitioners start calling, patients complain and it takes much more time to plan appointments. Some outpatient clinics ensure that the doctors feel the pressure by transferring calls and complaints directly to them. Another incentive is that the respondents feel proud to deliver a good service.

"It creates stability and it is a gesture towards patients and that is most important… That's what I like about it. It's patient friendly." (doctor, clinic 2)

Some also state that a short access is needed because of increased competition between hospitals.

"Our short access is the key to our success. Nothing else. Not because we operate well, not because we are friendly. It's about accessibility" (doctor, clinic 8)

In every successful outpatient clinic, at least one doctor is committed to minimize delays and invests time and energy to sustain the results. This doctor regularly checks access and takes action.

"I track our access times and free agenda space. When I think that we will run into problems I ask colleagues to run extra sessions" (doctor, clinic 2)

These committed doctors confront other doctors when they do not uphold proper work practices such as cancelling sessions at short notice. In addition, they enable support staff to uphold the required work methods, even when other doctors try to change it.

"I once deliberately let it all go wrong. Then those who are least active came to complain…I explained how it was their own fault. It all depends on how you plan your holidays and conferences… We have a mutual responsibility for keeping access short, but that also entails

an individual responsibility. I think they got the message. But once in a while someone needs to be reminded." (doctor, clinic 2)

"I believe, if the doctor would stop… well, I doubt we could continue. Someone must support it with a lot of energy and faith." (assistant clinic 2)

No respondent mentioned the contribution of senior management as a relevant factor for sustaining results. Several doctors continue despite the lack of senior management support.

"You get a pad on the back by the senior management and that's about it. It becomes a bit tiresome." (doctor, clinic 4)

During the projects senior management was usually involved as sponsors. After the projects senior management did not change incentives and did not establish supportive structures to sustain the results.

In one of the two outpatient clinics that did not sustain their results, several respondents argue that this is caused by a lack of motivation and clinical leadership among doctors:

"The doctors don't seem to consider access important anymore. Their attitude is 'that's temporarily, it will correct itself'. But we see that access has become permanently longer." (team leader, clinic 11)

3.5 A shared belief that they can and should control access together

Besides leadership and incentives, the respondents also emphasize the importance of a shared belief that they are able to control access, even that its wrong not to. This belief is shared among the doctors as well as the support staff.

The project teams had made their own analysis of demand and supply, tested interventions, involved their colleagues and measured results. This involvement created a firm believe in the work practices and their own ability to control access. As a result, delay in access is no longer seen as the result of too much demand-, but of bad planning.

"Long waiting lists used to be a status symbol, because it implied that you are a good doctor. Now it states: 'why are you not able to organize your clinic?" (doctor, clinic 10)

Many respondents demonstrate an intrinsic drive to sustain access. They continue because they believe in it. This does not imply that every colleague agrees with each new working practice. However, the shared belief overrules individual preferences and the return of old habits.

The outpatient clinics that did not sustain results, state factors outside their influence that frustrate their efforts; they feel unable to control access and less responsible.

"We have all been convinced the past years of the importance of it all and how nice it is to work with a short access… Doctors want to perform more sessions, but there are not enough nurses and supporting staff… the motivation drops. There are people who say 'if we just let the waiting lists rise again, we are likely to get more support" (doctor, clinic 13)

4. Discussion

4.1 Main findings

Eleven, maybe 12, out of 14 specialist outpatient clinics sustained their reduced delays for at least three years. The clinics also sustained the majority of the work practices. The outpatient clinics focused mainly on sustaining their results, not on further improvement. New interventions were introduced only during the aftermath of the project. No new work practices were devised.

The Advanced Access concept advocates that an outpatient clinic needs to change the system by which demand and supply are matched, to prevent the return of the backlog.[3,4] Our results confirm that the specialist outpatient clinics have remained flexible in the volume of their supply and respond quickly to fluctuations in demand with the use of feedback signals. In system terms, they apply pull methods.[5] The outpatient clinics that did not sustain their results stopped being responsive. These findings are consistent with recommendations from other research on improving flow in care processes and on managing variation in demand.[5,16] Our research shows that pull methods are not only effective to reduce delays, but also to sustain the results.

In order to be sustainable, this system change needs to be supported by incentives, clinical leadership, and a shared belief that they can and should control access. The care providers are motivated by directly experiencing benefits from reduced delays. These are both rewards for maintaining changed behaviour (patient satisfaction, beating the competition), as disadvantages when returning to old habits (pressure, stress). This is consistent with other research.[7,17] It has been suggested that measuring results is an incentive.[14,18] Our study shows that it is important to make the consequences of actions visible; measuring results is one way to achieve this. Perhaps most important is clinical leadership; a doctor who persistently supports the interventions and who confronts those who waver. This is consistent with previous studies on the importance of champions.[7,8,11] Noticeably, the champions in our study are doctors. They are better able to address their colleagues as peers to uphold agreed work methods and ask for extra efforts. Furthermore the respondents show a consistent belief that together they are able to control access, even that it is wrong not to. This shared belief legitimises confronting and overruling doctors and others who do not uphold new work practices. These findings are consistent with studies on sustainability that show the importance of cultural changes that support the relevant goals and work methods.[17]

Inconsistent with prior studies is the finding that respondents did not identify senior management support as a relevant factor.[19] Perhaps this can be explained by the high level of influence of doctors on decisions in Dutch hospitals. However, when sustainability becomes dependent on changes in the larger hospital system or if clinical leaders leave, absence of senior management support could risk sustainability. More senior management involvement in an earlier phase of these projects might have supported sustainability by providing better monitoring and control access. But if it involved setting targets and controlling results, the impact could be negative. The respondents show a real sense of responsibility; it is their project, their targets and their results. Senior management involvement might have compromised this sense of responsibility and hence sustainability.

Another difference from other studies is that formal training and education did not appear necessary to sustain the lessons learned during the project.[17,20,21] The knowledge transfer seems to

work effectively through informal socialization only. This may however be dependent on the continued presence of the initiators.

The question that remains is why these factors were present in most of these clinics. We speculate that the change strategy of the 'breakthrough collaborative' was important. Part of the strategy is to make a multidisciplinary team responsible for their own implementation process. They are taught the principles by their own peers, choose and test their own interventions, measure results and adapt these interventions. In each specialist outpatient clinic, members of the improvement team still act as ambassadors of the principles and as clinical leaders.

4.2 Limitations

The study might have benefited from a longitudinal approach in which work practices and access is tracked continuously. However, a continuous presence could introduce a problematic bias and effect sustainability. Also, we did collect data during multiple time points and reliable measurements and reports from the projects were available, plus most of our respondents worked in the specialist outpatient clinics during the entire period.

We have compensated for subjective bias and recall bias by using different sources: outcome measurements, interviews with different respondents from different sites and project reports to confirm findings (triangulation). Still, subjective bias and recall bias may have influenced the results.

4.3 Recommendations

The role of senior management deserves more attention. Longitudinal studies could be adapted to prevent researchers influencing the clinical practice.

Improvement processes, without adding capacity, can sustainably replace access delays not only in specialist outpatient clinics but throughout hospital systems in general. This requires care providers to become more demand driven and the system to become more flexible in response to fluctuations in demand. Investing resources in approaches that do not challenge conventional thinking and way a system is designed are unlikely to achieve sustainable results.

Care providers are able to control access once they understand the causes of delays and ways to tackle them. In addition, when they experience direct benefits in their work and there is a clinical leader who takes strong ownership sustainable improvement is more likely. Policy makers and senior management do not need to show leadership, but should provide a supporting structure.

References

1. Morton A, Bevan G: What's in a wait? Contrasting management science and economic perspectives on waiting for emergency care. Health Policy 2008;85:207-217.
2. Institute of Medicine: Crossing the Quality Chasm. A new health System for the 21st Century. Washington: National Academy Press; 2001.
3. Murray M, Berwick DM. Advanced Access: Reducing waiting and delays in primary care. JAMA 2003,289:1035-1040.

4. Rouppe van der Voort MBV, Berden HJJM, Van Merode, GG: Making sense of delays in outpatient specialty care. Health Policy 2010,97:44-52.
5. Hopp WJ, Spearman ML. Factory physics. New York: Irwin/McGraw-Hill; 2000.
6. Buchanan DA, Fitzgerald L, Ketley D (ed.): The sustainability and spread of organizational change. New York: Routledge; 2007.
7. Scheirer MA: Is sustainability possible? A review and commentary on empirical studies of program sustainability. American Journal of Evaluation 2005,26(3):320–347.
8. Martin GP, Currie G, Finn R, McDonald R: The medium-term sustainability of organizational innovations in the national health service. Implementation Science 2011,6;19.
9. Virani T, Lemieux-Charles L, David A, Berta W : Sustaining change: once evidence-based practices are transferred, what then? Longwoods Review 2009,6(4):89-96.
10. Kotter JP: Leading change: why transformation efforts fail. Harvard Business Review 1995,73:59-67.
11. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O: Diffusion of innovations in service organizations: systematic review and recommendations. The Milbank Quarterly 2004,82:4:581- 629.
12. Julian D.A., Kombarakaran F. 2006: Assessment of quality outcomes within a Local United Way Organization: Implications for sustaining System level Change. Am J Community Psychol 38,175-181.
13. Swain K, Whitley R, McHugo GJ, Drake RE: The sustainability of evidence-based practices in routine mental health agencies. Community Ment Health J 2010 Apr,46(2):119-29.
14. Savaya R, Elsworth G, Rogers P: Projected sustainability of innovative social programs. Eval Rev 2009,33(2):189-205.
15. Creswell JW. Research design: Qualitative, quantitative and mixed methods approaches. London, Sage Publications; 2003.
16. Van Merode GG, Groothuis S, Hasman A: Entreprise resource planning for hospitals. International Journal of Medical Informatics 2004,73(6):493-501.
17. Pluye P, Potvin L, Denis JL: Making public health programs last; conceptualizing sustainability. Evaluation and program planning 2004,27:121-33.
18. Grimshaw JM, Thomas RE, Maclennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, Wensing R, Dikstra C, Donaldson C, Hutchinson A: Effectiveness and efficiency of guideline dissemination and implementation strategies. Health technology assessment report 2004;8(6):751-76.
19. Mcgrath KM, Bennett DM, Ben-Tovim DI, Boyages SC, Lyons NJ, O'connell TJ: Implementing and sustaining transformational change in health care: lessons learnt about clinical process redesign. Med J Aust 2008,188(6):32-35.
20. Gustafson DH, Sainfort F, Eichler M, Adams M, Bisognano M, Steudel H: Developing and testing a model to predict outcomes of organizational change. Health Services Research. 2003,38(2):751-76.

21. Jacobs RL: Institutionalizing organizational change through cascade training. Journal of European Industrial Training 2002,35:299-333.

CHAPTER 7

How specialist outpatient clinics match supply and demand

1. Introduction

In this chapter we identify and explain the variables to optimise the process of matching supply and demand in specialist outpatient clinics. We have tested and further developed the theoretical framework from chapter 2 by applying it to in-depth field research on how two specialist outpatient clinics match supply and demand. In this chapter we answer the research question:

1. *How are demand and supply matched in specialist outpatient clinics?*

We studied the characteristics and planning processes of two specialist outpatient clinics: one that can match demand and supply with relatively little dependence on other departments, and one outpatient clinic with a high dependency, in particular of the planning of the operating theatre. They enabled us to identify the variables for specialist outpatient clinics and use them to test the following two hypotheses in chapter 8:

1. *The way supply and demand are matched is predominantly a push method.*
2. *The way specialist outpatient clinics match supply and demand causes delays.*

A pull method is predominant when the release of capacity is primarily based on the actual system status; a push method is predominant when the release of capacity is primarily based on an exogenous trigger. See chapter 2.3 for more information.

To analyse the process how specialist outpatient clinics match supply and demand we describe for both specialist outpatient clinics:

- Characteristics of demand.
- Characteristics of supply.
- How supply and demand are matched: the planning of sessions.
- How supply and demand are matched: the planning of individual patients.
- Reflection on the way supply and demand are matched.
- Reconstruction how supply and demand are matched evolved over time, including identification of the triggers, responses and considerations and what drives these choices.

The results are used for both specialist outpatient clinics to:

- Create a process flow diagram to understand the relation between the different actions and variables and how they relate to the delay in access.
- Understand how the specialist outpatient clinics cope with uncertainty in the volume of demand
    - Understand the enablers to increase the responsiveness to fluctuations in the volume of demand.
    - Understand the barriers to increase the responsiveness to fluctuations in the volume of demand.

Paragraph 2 describes the methods of this research. We present the results in paragraph 3, reflect on the results in paragraph 4 and identify the variables to optimise the way supply and demand are matched in paragraph 5. Paragraph 6 discusses what drives the choices how supply and demand are matched and in paragraph 7 we discuss whether the results support the hypotheses and answer the research question.

2 Methods

2.1 Research design

To investigate the characteristics of how specialist outpatient clinics match supply and demand we explored:

1. A Dermatology outpatient clinic with relatively low dependence on other departments to match supply and demand.
2. An Orthopaedic outpatient clinic with relatively high dependence on the operating theatre to match supply and demand. We observed the way both specialist outpatient clinics match supply and demand in the field and followed the processes as they unfolded. This form of naturalistic inquiry took place in a real world setting and required the researchers not to attempt to manipulate the processes.[1] Observations enabled us to get close to reality and to understand the context where actions take place.

To gain a detailed understanding different sources of information were collected and combined.[2] Because of inter-observer variability, different researchers observed the same processes. In between and after the observations the researchers interviewed actors with different roles in the processes to gain a deeper understanding of the processes and to understand the different perspectives on the events that were observed.

The aim of the research was not to generalise the findings of these two specialist outpatient clinics, but to gain a detailed understanding of the processes to learn about the dynamic of the processes and identify the variables that determine the way demand and supply are matched. This required an approach that is 'rich in information' and enabled a deepening of the understanding of the observed events.[1]

2.2 Data collection methods

We used four data sources:

- Field observations, including field-interviews
- Interviews
- Documents from the specialist outpatient clinics
- Data from the planning and other IT systems from the hospital

Field observations, including field interviews

The operational planning processes of the outpatient clinical assistants and the team leader at the desk and at the phone were observed during eleven days. Observations were logged. This included general information about the date (when), researcher (the observer), setting (what), actors (who) and context (how) and contains specific information on each patient that is planned during the eleven days including: time of event, type of appointment, appointment date, what code is used to book the appointment, the original code of the use appointment slot, which formal or informal rule was applied to plan the appointment and notes for relevant contextual information (for example if a patient did not accept the offered date and increased pressure to get an earlier appointment). During the observations, short field interviews are used to go more in depth, understand reasons for decisions, and for clarification.

Interviews
During and after the observations, semi-structured interviews were held with key informants of each specialist outpatient clinic: specialists, including the medical manager, management, secretary, assistants, current team leader and former team leader. Information derived from the interviews was used to check and get a better understanding of the operational and aggregate planning processes and the historical development. Each interview was verbatim described.

Documents
Documents were analysed to identify formal rules, such as work protocols, and also in order to check consistency of the observations and interviews. The documents provided protocols and internal memos that complemented the observations and contributed to tracing changes of the planning system. An example of a formal rule is that only a general practitioner, not a patient, can book slots for urgent appointments.

Data from the planning and other IT systems from the hospital
The basic appointment schedule is derived from the hospital planning software, as well as information on the use of the planning system, e.g. how many appointments were booked on a different code than the basic appointment schedule has planned. An attempt to retrieve fluctuations in the volume of demand from the available appointment module in the IT system failed because the type of demand cannot be recovered in a meaningful way.

2.3 Data analysis
Since we did not know beforehand the variables to be identified, our analysis strategy is an inductive analysis to describe and understand the processes to match supply and demand, and a creative synthesis to identify the variables that determine the way supply and demand are matched.[1] For the type of phenomena that we study it is valuable to pursue different perspectives to gain a better understanding.[3] To understand the system dynamics of the way supply and demand are matched we used the methods from 'Simulation based engineering' to create a system visualisation.[4] To understand the context of the way demand and supply are matched we used semi-open interviews. The system description and the contextual analysis together enabled us to identify the variables that need to be considered to optimise the way demand and supply are matched.

3 Results
3.1 Introduction
Paragraph 3.1 describes the general characteristics of the studied specialist outpatient clinics; 3.2 describes what we learnt about the characteristics of demand and 3.3 the characteristics of supply; 3.4 describes the planning of sessions and in 3.5 the planning of appointments. Paragraph 3.6 reconstructs how the way demand and supply are matched evolved over time and in 3.7 the perspectives of the involved actors on optimal access are described. Finally, 3.8 describes the way the two specialist outpatient clinics match supply and demand with a system visualisation.

3.2 General characteristics of the studied specialist outpatient clinics
Both specialist outpatient clinics are part of a large teaching hospital with 3.500 employees, 550 beds and 180 medical specialists. Dermatology does not have to deal with wards, operating

theatres, diagnostic or treatment departments, making it a relatively independent and low complex specialist outpatient clinic. Orthopaedics on the other hand is highly dependant on other departments to match supply and demand. The orthopaedic surgeons perform sessions in the outpatient clinic, but also operate in the operating theatre. Furthermore many patients require services from the Radiology department on the same day as the appointment. Orthopaedists also supervise and teach residents to become orthopaedists, which has consequences for the planning of sessions.

3.3 Characteristics of demand

Patients are usually referred by a General practitioner and sometimes by another medical specialist. Patients may prefer a specific specialist but it is common that there is no preference. Sometimes a patient needs an appointment within several days or weeks, but usually there is no medical urgency. Patients schedule the appointment with the assistant themselves.

Most patients have at least one follow-up appointment in the same year, sometimes more. Some patients need to return for a follow-up appointment on a yearly basis. The ration between first and follow-up appointments is approximately 1:3 / 1:4. There are no large predictable seasonal fluctuations in demand as far as the respondents can tell, except that there is less demand during the summer holidays and an increase for Orthopaedics just before and after the summer holidays.

3.4 Characteristics of supply

People

Dermatology works with three dermatologists, two nurse practitioners, three nurses, six outpatient clinical assistants and two secretaries. Orthopaedics works with 7 orthopaedists, 1 fellow, 5 residents, 1 nurse, 12 assistants and two secretaries. The fellow is an orthopaedist who works temporarily for the group of orthopaedists. The resident are doctors that are specialising to become an orthopaedist.

The specialists are not employed by the hospital: both groups formed a partnership and have a contract with the hospital. It can be difficult to increase the number of specialists on a short notice, but by hiring a 'chef de clinique' or a fellow temporary solutions are possible. It's relatively easy to increase the number of assistants, but it will take several weeks to understand the system and up to six months before they are able to plan patients without supervision. The specialists themselves decide to a high degree of freedom whether they are available to perform sessions on any given day. There is an agreement however that they are 42 weeks per year available and that holidays are coordinated to ensure a minimum occupancy.

Space and materials and basic process

Both specialities use a treatment room, a consultation room and an office for each specialist during sessions. While the doctor is with a patient in one room, another patient already enters the other room and is prepared by the assistant for maximum efficiency. When an orthopaedist and a patient decide on an operation, this is planned with an assistant directly in the outpatient clinic in a separate room. All three dermatologists can run sessions simultaneously with nine rooms. The orthopaedists can simultaneously run sessions with three of eight specialists who use three rooms and three residents who use two rooms. It is difficult to increase the space due to physical

restrictions of the building and the choice that each specialist outpatient clinic has its own set of rooms.

One specialist outpatient clinic uses outdated software to schedule sessions and appointments that is cumbersome to use and strongly limits planning possibilities and the retrieval of information to optimise the matching of supply and demand.

3.5 The planning of sessions

The planning of sessions consists of two types of activities:

1. Determining the weekly mix of sessions (incidental activity)
2. The planning processes of sessions and people (continuous activity)

Ad 1. Determining the weekly mix of sessions (incidental activity)
The specialist outpatient clinics work with a fixed schedule that repeats either each week or every two weeks, see appendix 1 for an example. Dermatologists perform sessions every day, each Orthopaedist two or three days a week (the other days they operate in the operating theatre). They have a two-week-schedule because they perform on average 2.5 days of operating sessions but prefer to operate full days. Each specialist performs general and specific sessions:

1. General sessions have a mixture of pathologies and a mixture of new and follow-up appointments and are the same for all specialists.
2. Specific sessions can differ per specialist, for example:
    a. Every Thursday evening the dermatologists see only new appointments with suspicious skin spots, possibly oncology. The appointments are shorter than regular new appointments and they require fast access for medical reasons.
    b. 'Rheumatology session', 'Foot session' and 'Scoliosis session': an orthopaedist and another physician see patients together in reserved slots, as part of general sessions.

Both the dermatologists and the orthopaedists consider it important when determining the session mix that the workload of revenue generating appointments is evenly distributed amongst the specialists. Personal preferences can also be decisive for individual specialists on which day which session is performed. Orthopaedics has specific other considerations for the session mix:

- An evenly spread of sessions on each day and part of the day for efficient use of the rooms and support of the assistants.
- A maximum of six sessions simultaneously because of the physical restrictions.
- A match between orthopaedists and specific residents to provide supervision.
- Availability of other disciplines for the combined sessions.
- The operating theatre schedule has priority over the outpatient session schedule.

For the whole year the fixed session mix is planned in advance. It usually only changes if a specialist or resident leaves or a new one starts.

Ad 2. The planning processes of sessions and people (continuous activity)
Sessions get fixed 8 weeks in advance. The formal rule is not to book appointments beyond 8 weeks. The relatively short planning horizon (considering that patients often need to return in 3, 6

or 12 months) is a general rule of the hospital. It is a countermeasure to reduce the number of rescheduled appointments and no shows as a consequence of cancelations of sessions.

The team leader schedules the people and cancels sessions when specialists tell him to, but also pushes back when specialists cancel too many sessions or within 8 weeks. Dermatology schedules assistants three months ahead. For example, mid-January assistants are scheduled for April. Orthopaedics schedules assistants 8 weeks ahead (because of the 8 week horizon when sessions are released). The formal rule is that specialists can only cancel sessions at least 8 weeks in advance (again because of the planning horizon of 8 weeks). The orthopaedists have the rule that at least 3 orthopaedists must be present in the hospital. Reasons for cancellations are absence of the specialist or when the specialist needs to perform operating sessions.

Changes in the schedule of the operating theatre lead to cancelations of outpatient sessions. The operating theatre has a planning horizon of 6 weeks. The consequence is that an outpatient session that became final 8 weeks in advance might be cancelled 6 weeks in advance to enable an operating session. The team leader of the outpatient orthopaedic clinic tracks the delays for outpatient appointments as well as for surgery. He pro-actively communicates with the specialists if he notices an increase in either of the access times and proposes actions to keep both delays within acceptable levels and at the same time keep the utilisation level of both capacities high. The utilisation of the operating theatre has priority.

If the specialists request extra sessions more than 8 weeks in advance they can count on support from assistants. Otherwise it can depend on the willingness of the assistants to work extra. Requests for extra sessions can be required because a specialist sometimes wants to compensate cancelations of sessions or because the business controller informs management and the specialists that production is behind schedule. Both specialities weekly measure the delay in access, but only Orthopaedics sometimes acts on it by performing extra sessions when it increases. Each orthopaedist decides individually whether he considers it necessary.

When the specialists notice that it is becoming problematic to offer patients urgent appointments fast enough this will lead to extra sessions or extra fixed slots for urgent appointments. This is not monitored, but noticed easily because patients and General practicioners immediately express their concerns.

3.6 The planning of appointments

Figure 7.1 shows the general planning process how the specialist outpatient clinics schedule appointments. The boxes represent actions and the circles the start and finish of the process. The diamond is a decision box. If an appointment ends with the decision to return for a follow-up appointment this can be scheduled immediately if it is within the planning horizon. In other cases the follow-up appointment is scheduled at some point in the future.

The specialist outpatient clinics use 7 general appointment codes:

- Two codes for new appointments:
    - N = new (first) appointment
    - NC = New appointment for a known patient with a new referral
- Two codes for follow-up appointments:

    - C = follow-up appointment
    - CT = semi-urgent follow-up appointment (within several weeks)
- Three other codes:
    - S = urgent appointment, new and follow-up (within several days)
    - AL = slot that is not predestined
    - ICC = intercollegiate consult (appointment that is requested by another specialist)

Dermatology uses a fixed appointment length for each general type of appointment, for example new appointments (N) are 15 minutes, follow-up appointments (C) are 7,5 minutes, etc., with the exception of three extra codes. For example: new appointments with possible skin cancer have a specific code, are shorter, are mostly booked in one specific session and are given priority. Orthopaedics has no extra codes, but further distinguishes 11 appointment lengths per pathology per type of code. For example a new appointment (N) for 'knee' or 'shoulder' problems are 10 minutes and 'back' problems are scheduled 15 minutes, etc.

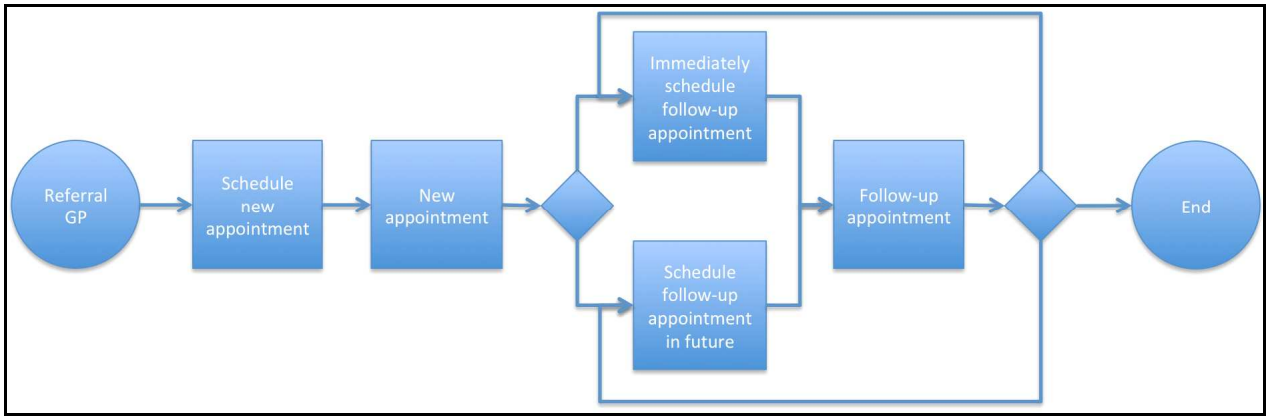


Figure 7.1 The basic planning process of appointments for one patient

When an orthopaedist performs an outpatient session and simultaneously supervises a resident that performs another session, five minutes are added to each appointment with the orthopaedist to allow time for supervision. When the orthopaedist supervises two residents, ten minutes are added to each appointment. Appointments with residents are always five minutes longer than with orthopaedists.

For Dermatology each session has a fixed layout, for example the first patient is always a new appointment, than two follow-up appointments, etcetera. The fixed mix is created to assure that enough percentage of the capacity is available for new appointments and to assure an even distribution of new appointments amongst the specialists. The order of the appointments is partly to support the logistics during the session. For example some equipment is only available in one of the two rooms. Another reason can be that for example on a fixed day a patient is given a test that needs to be checked the next day. For Orthopaedics the sessions consist almost entirely of five-minute slots that are not pre-determined (not fixed). The mix of appointments is different each time, depending on the demand for appointments. The only reserved slots are for urgent appointments and combination appointment with other physicians. When booking an appointment, one, two, three or four five-minute slots are taken and the type of appointment is selected to book it.

For both specialities, new appointments are booked with the first available specialist while considering the preferences of the patient for the date. Requests for a specific specialist are respected. Follow-up appointments are scheduled with the same specialist as the first appointment, unless the appointment is urgent. For Orthopaedics appointments might be booked with a resident that is supervised by the orthopaedist the patient saw before.

When booking an appointment for a patient it can happen that there is no slot available within an acceptable horizon. Dermatology and Orthopaedics have different ways to handle this problem. For Dermatology, if it's an urgent appointment, the assistant is almost always able to create a slot within the horizon by either booking it on another available type of appointment (a new appointment in the time slot of a follow-up appointment) or by adding an urgent extra slot that extends a session. There are formal rules that are agreed upon, for example a new appointment can be booked on a follow-up appointment, but not vice versa., urgent appointments can only be booked if the general practitioner calls, etcetera. We observed 34 formal rules and 25 informal rules concerning planning decisions. A formal rule is formally communicated. An example of an informal rule is when an assistant books two 15 min new appointments (N) on two 75 min C slots and blocks a 7.5 min C slot elsewhere in that session. Sometimes informal rules are allowed and sometimes an assistant knows that she breaks a rule, but she does not know what else to do to provide the patient the timely required appointment.

When Orthopaedics has no slot available within the requested horizon the assistants have the following options:

1. Book the appointment in the 'team' of the orthopaedist: a resident that is supervised by the required orthopaedist
2. Book the appointment on an unused five-minute appointment slot. If one or two other five-minute slots are available in the same session, they are blocked. If there is none available, it means they accept that the waiting time during that session increases.
3. Request the orthopaedist whether the appointment can be booked before the first or after the last appointment of a session. They are reluctant to do this, because the end time of each session is monitored and the norm is a maximum delay in end time of 20 minutes.
4. Request the orthopaedist whether a fixed slot for urgent appointments can be used.

If an urgent appointment can't be booked soon enough, there is always one resident 'on call' who will come to the specialist outpatient clinic.

3.7 Reconstruction how the way demand and supply are matched evolved

Which factors triggered changes in the way supply and demand are matched over the years on both an aggregate and an operational planning level? The following list shows what the respondents described as triggers from within the specialist outpatient clinic, from the hospital and from outside the hospital.

1. Within the specialist outpatient clinic

- Gradual increase in demand for new appointments each year.
    - Effect: measures to increase efficiency to see more patients per session. Most important changes are the three-room system with assistant support and reduction of the return

rate (how often patients return for follow-up visits), partly because nurses and assistants now also see patients and partly because the specialists evaluate the need for follow-up visits.
    - Effect: Increase of delays when at some point the efficiency measures do not suffice.
        - Effect: reservation of (semi-) urgent appointment slots to ensure fast access.
        - The expansion with another specialist at some point.
            - Effect: more urge to control the number of revenue generating appointments of each specialist, to ensure even contribution to the total result (the specialists evenly spread the revenues).
- The national change of the financial system from a fixed income to rewarding specialists and hospitals for actual production (specific appointments that generate income).
    - Effect: the volume of revenue generating appointments became fixed in the sessions..
    - Effect: checking financial performance and opening extra sessions if necessary.
    - Effect: efficiency measures are financially more interesting because they lead to a direct increase of income per specialist when they prevent the expansion with a new specialist (see above: effect of increase of demand).
- Further specialisation of specialists.
    - Effect: creation of more specific sessions and types of appointments and less flexibility to schedule patients with any specialist.

2. Outside the specialist outpatient clinic, inside the hospital

- Changes in the financial or administrative system of the hospital requiring certain new types of appointment codes to be added.
    - Effect: more types of appointments, which reduces flexibility when working with fixed slots.
- Hospital decision to reduce the planning horizon of all specialist outpatient clinics (to reduce rescheduling appointments of cancelled sessions and to reduce no shows).
    - Effect: reduction of the planning horizon for patients to 8 weeks and the rule not to cancel sessions within 8 weeks.

3. Outside the hospital

- National indicator for general delays in access for any appointment and specific fast access for certain types of oncology.
    - Effect: weekly measurement of access, making it more visible for the specialists and management as well as for patients and General practicioners.
    - Effect: creation of separate sessions for oncology, which reduces flexibility.
- National guidelines that include the follow-up frequency for certain pathologies and new diagnostic possibilities and new treatment possibilities.
    - Effect: increase or decrease of the return rate.

3.8 Perspectives of the involved actors on optimal access

One medical manager considers two to three weeks as the optimum delay in access. He realises that faster access would be better for the patients, but that this is acceptable for non-urgent appointments and it enables the specialist outpatient clinic to work with full sessions. The optimum is based on the wish to have just enough buffer 'inventory' of appointments to work with an as high

as possible utilisation rate.

Another medical manager and department manager considers a delay of less than a week ideal. This is based on providing the patients fast access, but also to reduce stress in the system. The disadvantage of a very short access of a few days is that more patients come in with complaints that would have passed spontaneously if the patient had waited longer. They do not consider it a problem if slots remain incidentally unused in sessions; capacity does not have to be used 100%.

One team leader considers three to five working days as the ideal access for all appointments in the specialist outpatient clinic. If this period is shorter the sessions have unused appointment slots and the practical preparations and logistics become problematic. If it is longer, waste creeps in because some of the patients require faster access.

One interviewed specialist considers access time not important (except for urgent appointments). He considers it important to help the number of patients that is agreed in the production agreements. He considers a short delay a risk because it can increase demand that should be treated by General Practitioners. Another specialist states that a longer delay is an indication of being a better doctor. The first specialist considers two weeks the ideal access, the other four weeks.

3.9 System visualisation how supply and demand are matched

In this paragraph we use the described characteristics of both specialist outpatient clinics to create a visualisation of the way demand and supply are matched. Our approach is based on Simulation-Based Engineering of Complex Systems.[4] We use the following elements:

- What are the processes on an aggregated and on an operational level? The boundary of a process for this purpose is determined by grouping the tasks that are sequential.
- What are the decisions that are taken? What information is used for these decisions? What are the rules that are used to take the decisions?
- What are the feedback and feed forward loops in the processes?

The matching of demand and supply at the specialist outpatient clinics exists of 9 (Dermatology) or 10 (Orthopaedics) processes:

Operational planning

1. Request appointment
2. Plan appointment

Aggregated planning

3. Request session cancellation
4. Cancel sessions
5. Check system status
6. Plan extra sessions
7. Request assistant holiday
8. Request assistants extra workday
9. Plan assistant holidays
10. Operating theatre planning leads to a cancellation of a session (Orthopaedics)

See appendix 2 and 3 for the description of the processes, triggers, input, rules and outputs for Dermatology and Orthopaedics respectively.

Our interest lies in using the cases to identify the relevant variables and understand the system dynamics. Figure 7.2 shows the system visualisation for Dermatology and 7.3 for Orthopaedics. In the visualisations a 'schedule' is symbolically visualised in the centre, see appendix 4 for an example of the details of the schedules for Orthopaedics. Several parameters are tracked in 'logs':

- The access in days of the third available appointment per type of appointment per specialist in 'Log access'.
- The production per specialist, measured in the volume of appointments that generate income (not the total volume of appointments) in 'Log production'.
- The number of cancellations of sessions per specialist in 'Log cancellations'.
- The number of holidays per assistant in 'Log holidays'.

In both figures the two processes stated above concern the operational planning of appointments and the other processes concern the aggregate planning of sessions. An arrow indicates that the output of one process triggers another process. The colour indicates whether the release of capacity is pull (based on the actual system status) or push (based on an exogenous trigger). A dotted arrow indicates that a process provides information for another process (but does not trigger a process). In Figure 7.3 the two processes on the left are part of the operating theatre, but directly influence the planning of the Orthopaedic specialist outpatient clinic. Again, see appendix 3 and 4 for the details of each process.

4. Reflection on how demand and supply are matched

We use the identified elements of the Theoretical Framework in chapter 2 to reflect on the way the Dermatology outpatient clinic matches demand with supply. We reflect on how they manage variation on the aggregate level of planning sessions in paragraph 4.1, on the operational level of planning appointments in 4.2, on whether push or pull methods are applied in 4.3 and on the reconstruction of how supply and demand are matched evolved over time in 4,4, and we finish with general observations in 4.5.

4.1 The aggregate level: the planning of sessions

- There is no forecasting of demand, other than negotiating a steady increase with the insurance company based on general interpretations of national developments. The return rate is also not forecasted. It is not known how much over- or under capacity is available.
- Neither the volume of demand nor supply is measured. The volume is not anticipated and there are no deliberate mechanisms to notice variation in either demand or supply, and no mechanisms to react to variations. Production is used to track financial performance, not to manage the matching of supply and demand.
- Reductions in supply have no relation to demand or any other form of system status. There are rules to prevent too much drop in supply. Access is measured to provide information to the outside world, but only part of the orthopaedists use it to react to problems in access. A lack of financial performance triggers extra sessions.

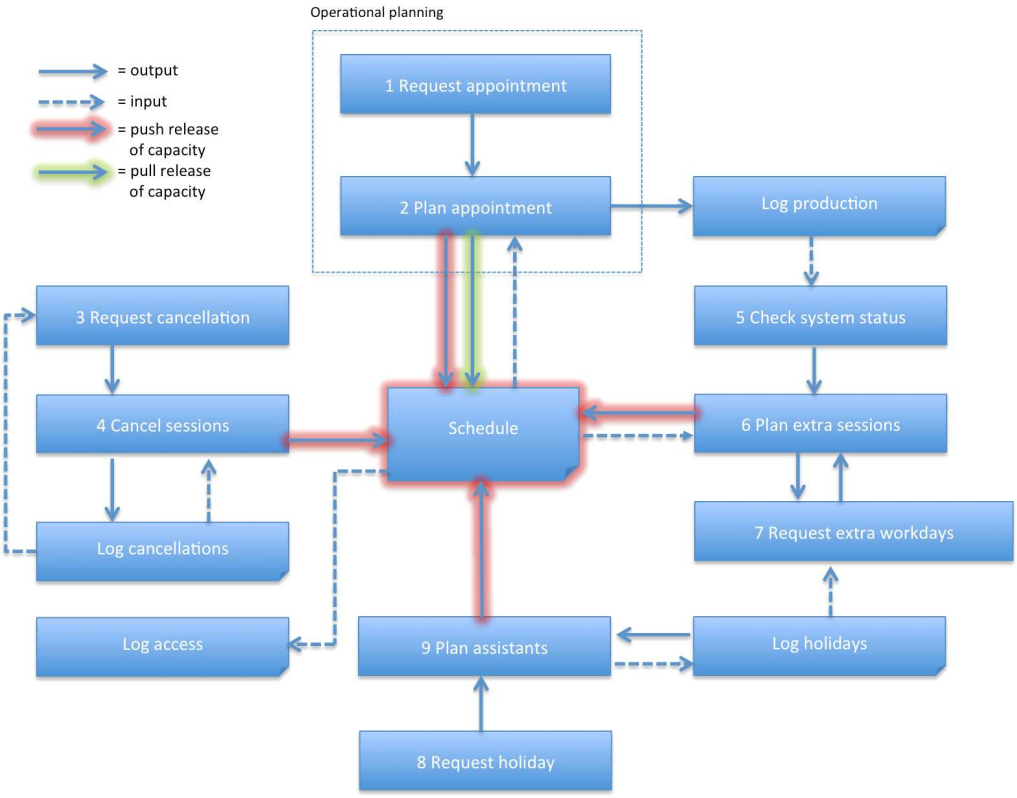


Figure 7.2 System visualisation for Dermatology

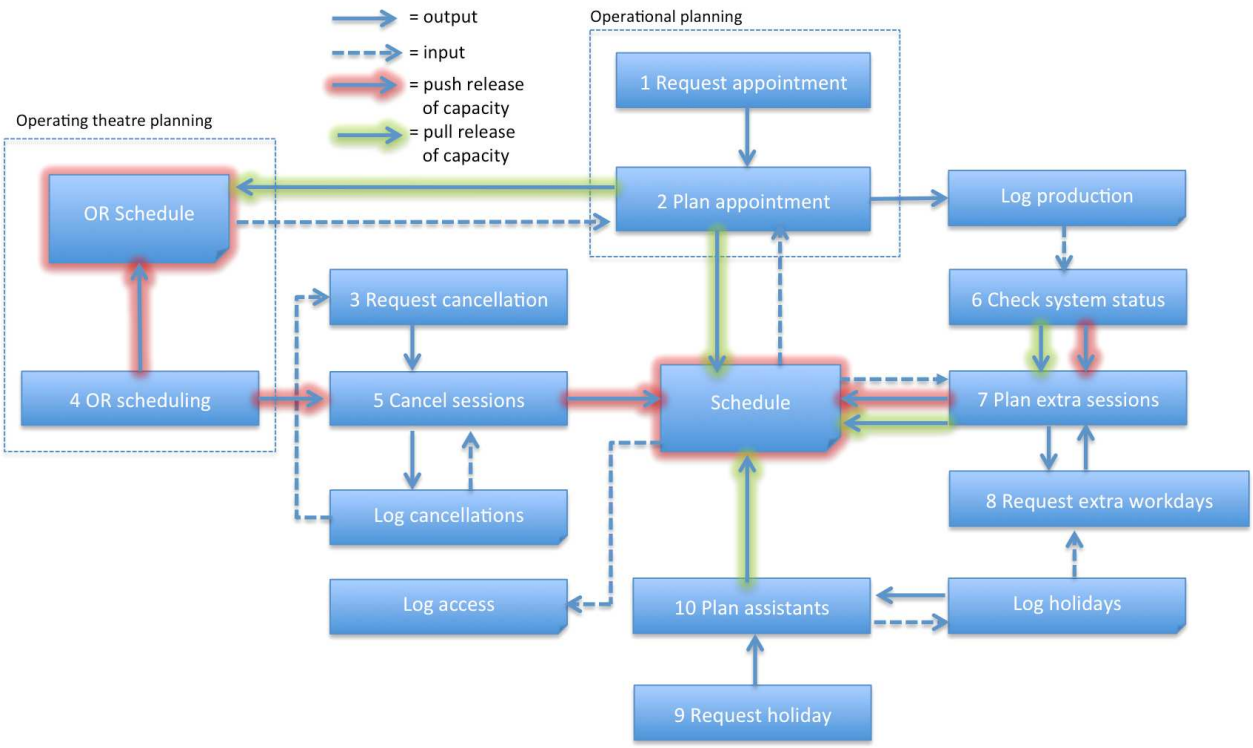


Figure 7.3 System visualisation for Orthopaedics

- Dermatology has almost no flexibility in the system to increase the volume of appointments in sessions, not in the long term and not in the short term. They also have limited flexibility to change the mix of offered appointments. However, the appointment mix on a weekly basis does change often due to the cancellations of sessions and this is not compensated in other sessions. Orthopaedics on the other hand has a high level of flexibility in the system to increase the volume of appointments in sessions, in the long term as well as the short term. The hiring of the fellow is a measure to increase the volume of appointments as a relatively fast reaction to the steady increase of demand. If necessary they open up extra session in a few months as well as in a few weeks. Furthermore, they book extra appointments, effectively extending the scheduled sessions if necessary. Also the teamwork with the residents creates extra flexibility.
- The horizons of 1 or 2 weeks for the cycle mix and 8 weeks for the planning of appointments have no relation to fluctuation in demand and are not chosen to optimise the mix of appointments in a week.
- The changeover flexibility in the mix of type of appointments per sessions and per week is potentially large. The specialist outpatient clinics can easily change the standard mix per session or sessions per week. For Dermatology this is rarely done however and only ad hoc when the circumstances change (for example a new doctor) or when a shortage of a certain type of appointment like urgent appointments has become a big problem over a long period. Orthopaedics has an almost completely flexible schedule, automatically providing maximum changeover flexibility.

4.2 The operational level: the planning of appointments

For Dermatology, the fixed appointment mix per session, per day and per week has no relation to the expected or actual mix of demand. The actual appointment mix does partially reflect actual demand, because some type of appointments is booked in the slots of other types, creating some flexibility. For Orthopaedics, there is almost maximum flexibility in the system for the mix of offered appointments. As a result, the mix of appointments per session, per day and per week is determined by actual demand. Appointment types N, NU, NO and CU have shared queues (the appointment can be booked with any specialist) and thus high flexibility, C is a single queue (needs to be planned with a specific specialist) with low flexibility.

The system to match supply and demand makes no distinction between the two types of queues. There are no compensation mechanisms or other methods to manage variation (a compensations mechanism is for example: if a specialist is several days off, the other specialists supply more N, NU, NO and CU slots). In other words: even though the specialists fix several types of appointments, there is no coordination of the volume and appointment mix neither on the aggregate nor on the operational level. The cancelation, extensions or adding of sessions by one specialist have no relation to the sessions of other specialists. As mentioned before, Orthopaedics has an almost completely flexible appointment mix.

4.3 Push and pull methods: what triggers the release of capacity?

The planning of sessions for Dermatology is a push system. There is no trigger from the actual system status that determines the volume and mix of sessions that is supplied to match with demand. The planning of appointments for Dermatology is predominantly a push system, but some level of pull is applied (by allowing actual demand for some type of appointments to overrule the

fixed slots of other types). The planning of sessions for Orthopaedics is predominantly a push method, but the status of the delay in access can trigger extra sessions, allowing some level of pull. The planning of appointments is almost completely pull based.

4.4 How the matching of demand and supply evolved over time

The reconstruction deepens the understanding of four potential causes of delays in access for specialist outpatient clinics:

1. Supply is larger than demand

When demand increases, there can be a reluctance to increase capacity with a new specialist because the current specialists can lose income if there is not enough demand for the new specialist to have a full workload (since the specialists divide income equally). Another barrier can be that it takes a long time to create extra physical space to enable the new specialist to perform sessions simultaneously.

2. Little flexibility in the appointment mix

The reconstruction shows several potential triggers to create fixed slots in the sessions for specific types of appointments. Fixing slots to ensure fast access for (semi-) urgent appointments and to ensure the sequential coupling of appointments in a week are logical countermeasures when delays increase, but should be let go when delays are reduced. Fixing slots to ensure enough space for appointments that generate revenue and to ensure an even spread of revenue among specialists is a poor countermeasure and is not necessary.

3. Little flexibility in the session mix

There are no reasons why the type of sessions is fixed on certain days of the week, the respondents simply never considered to make this flexible.

4. Not enough buffer capacity to react to fluctuations in demand

The lack of buffer capacity is the result of the desired goal to work at 100% utilisation. Efficiency improvements can not only counter the increase of demand, but also create the required buffer capacity.

4.5 General observations

- For Dermatology, the frame of reference is to provide urgent care with fast access and all other care should be acceptable from a medical point of view and patient experience. Maximum utilisation of the specialist capacity and generating exactly the income that is planned for that year has priority in the decisions with logistical consequences. Orthopaedics does take optimising the delays as a consideration in their decisions, but only has partially realised effective methods.
- Dermatology has less flexibility than Orthopaedics in the volume of sessions and in the appointment mix. There are few limitations however to create a much larger flexibility.
- A substantial waste is visible in the operational planning process for the assistants: it takes a considerable number of extra steps to find slots for patients because they cannot be offered soon enough with the available slots.
- The performance of the planning processes is not routinely evaluated.
- The planning horizons are not synchronised.

5. Variables to optimise delays in access to specialist outpatient clinics

The described results enable us to identify the relevant variables in order to optimise delays in access to specialist outpatient clinics. In chapter 2 it was explained how delays can be the result of how demand and supply are matched. If the level of supply is the same every day as demand, no delay will emerge. We group the variables in those relating to uncertainty in the volume of demand or the volume of supply. An inability to deal with these uncertainties will lead to mismatches, resulting in delays.

1. Dealing with uncertainty in the volume of demand

Outpatient specialty clinics deal with two distinct types of uncertainty in demand:

1.1 Uncertainty in weekly and seasonal fluctuation in demand for new appointments.

a. Theoretically specialist outpatient clinics can influence the distribution of demand for new appointments, for example with marketing.

b. Outpatient specialty clinics can analyse historical data and anticipate the distribution of demand for new appointments.

1.2 Uncertainty in weekly and seasonal fluctuation in demand for follow-up appointments.

a. Contrary to demand for new appointments, the distribution for follow-up appointments can be influenced relatively easily. This part of demand can be directed away from periods where demand is anticipated to be larger than supply, by shifting the period that a doctor and patient agree on when to plan the follow-up appointment. Note that if new or follow-up appointments are delayed or redirected the volume of future follow-up appointments shifts with the same delay. Seasonal fluctuations in follow-up appointments can therefore be caused by fluctuations in supply.

b. Outpatient specialty clinics can anticipate demand for follow-up appointments since most follow-up appointments have already been made at the end of the former appointment. This information can be used to calculate the required supply for follow-up appointments in future periods. Note that it is common that demand for follow-up appointments exceeds demand for new appointments and thus a substantial part of demand can be anticipated relatively easily.

2. Dealing with uncertainty in the volume of supply

The volume of supply has an aggregate (2.1) and an operational (2.2) level:

2.1 Aggregate: the number of available sessions.

a. The demand for new appointments, the return rate and the appointment length determine the number of new and follow-up appointments per session, which in turn determines the number of required sessions. The return rate and appointment length can be influenced with process improvements and choices in the level of quality. A percentage of buffer capacity is required to deal with fluctuations in demand. The more demand can be influenced and anticipated (see 1.1a,b and 1.2a,b) and the more flexible the capacity of doctors can be planned for outpatient sessions or operating sessions (see

2.1g), the less buffer capacity is required to optimise delays in balance with capacity usage. The return rate, appointment length and buffer capacity together determine the required number of sessions to meet demand.

b. The number of sessions that can be performed per week is determined by the availability of the medical specialists, rooms and assistants.
c. The cycle horizon (the number of days ahead where the mix of sessions and types of appointments is decided) can be based on the level of fluctuations in demand and supply, the level of uncertainty of these fluctuations and the desired delay in access. For practical reasons, a weekly cycle seems appropriate. The planning horizon for sessions can include several cycle horizons. For example for the next six weeks, six weekly or three two-weekly cycles can be planned. The planning horizon for sessions, specialists and assistants need to be synchronised.
d. The cycle mix (the required mix of appointments and the required mix of sessions to deliver those appointments) can be based on the predicted fluctuations of demand and corrected by the actual development of demand. Cancelling of sessions can change the required cycle mix and can lead to recalculation of the cycle mix.
e. How much supply of appointments and associated mix of sessions per week can be determined with the prediction of demand and can be adjusted based on the volume of appointments that are planned ahead in the available supply (actual demand and supply). Norms are required to continuously react to changing circumstances so that the appropriate mix of sessions can be planned. The flexibility to accommodate extra sessions is limited by the available specialist capacity, rooms and assistants above the normal availability, requiring buffer capacity or flexibility.
f. Cancelled sessions can be compensated within the same cycle horizon with extra sessions. Shared queues can be compensated by other specialists (new appointments without preference and urgent appointments); single queues only by the same specialist (new appointments with preference and follow-up appointments).
g. If the specialists also operate in the operating theatre the determination of the availability for outpatient sessions and for the operation sessions can be based on the predicted and actual demand and supply of both care units.
h. Norms determine the choices how much specialist capacity to release to either outpatient sessions or operating sessions and which type of outpatient sessions and appointments. For example a stable delay of two weeks for operating sessions can be chosen, while the delay in access to outpatient specialty appointments fluctuates with demand from one to three weeks. A distinction can be made between delays to new or follow-up appointments, of which the latter has more inherent flexibility to fluctuate.
i. The difference between the anticipated and actual demand, between anticipated and actual supply and the difference between the anticipated and actual effectiveness of the norms and decision rules can be evaluated regularly and the volumes, norms and decision rules can be adapted continuously to further optimise the system and to adapt the system to changing circumstances.

2.2 Operational: the number and mix of appointments per session

a. Specialist outpatient clinics can use either a fixed mix of appointments for each session or a partly or completely flexible mix. Fixed appointment slots can be required for the operational execution of the sessions due to dependencies on other resources of people that need to be planned in advance. The fixing of slots however reduces flexibility to match supply of each type of appointment with fluctuations in demand and is to be minimised. The required level of fixing can be determined per cycle, analogous to the planning of sessions as described above.

    i. If a period has less supply than required, this can trigger the (extra) use of a decision rule to schedule follow-up appointments that are not required within that period in a later period with a better ratio between supply and demand (see 1.2a). This will temporarily change the mix of appointments.

b. The number of appointments per session can be increased on a short notice by planning extra appointments before the start or at the end of the session, limited by the availability of the specialist and assistants.
c. Norms can trigger changes in decision rules if demand for specific appointments exceeds supply of that appointment. For example, urgent appointments can be booked on fixed slots for regular appointments when the fixed slots for urgent appointments are not available anymore.

6. What drives the decisions how supply and demand are matched?

In this paragraph we reflect on what might be drivers for the decisions that are made in specialist outpatient clinics on how they match supply and demand. We combine the results of our observations and the reconstructions (this chapter) with the results of the interviews on the factors that determine sustainability (chapter 6).

The respondents describe four drivers and assumptions that can determine the (implicit) goal they set for delays in access:

- Short delays are inefficient because they lead to low capacity utilisation.
- Urgent appointments require fast access; non-urgent appointments require an acceptable access and should be long enough to prevent patients to come for trivial problems that should be treated by the General practitioner or which disappear after a few days or weeks waiting.
- Patients and other specialists consider specialists with long delays better specialists.
- No belief that they can and should control delays in access, up to the point that if they would be able to do it, surely they would have done it already.

The respondents describe furthermore three drivers and assumptions that can determine the choices they make how demand and supply are matched.

- Desire to maximise capacity utilisation, in order to maximise revenue, but without jeopardising the quality of medical care. How patients experience the delivery of medical care is less present as a driver.
- The need for mechanisms to ensure an equal division of how much revenue each specialist generates.

- Practical problems that need to be solved, with little knowledge of system dynamics and Operations Management (OM) how the different countermeasures over time interact and can cause delays in access.

Altogether, the drivers and assumptions lead to the conclusion that the way specialists makes sense of delays in access to specialist outpatient clinics are a vital determinant how decisions are made how to match supply and demand. The interviews show that they act on facts, assumptions and interpretations (whether or not based on facts).

Sensemaking is an active reproduction of reality,[5] which we will explain further in relation to how specialists act or not act on optimising delays in access to their specialist outpatient clinic. In the research of chapter 6 several specialists stated that they used to identify long delays in access as a sign of themselves being a good or at least successful specialist. One specialist declared that before he learned how to optimise delays in access, he shouted out to the suggestion to reduce delays and offering all patients fast access: 'I'm not a gumball machine!'.

Several specialists stated that after learning how to optimise delays they came to identify maintaining short delays in access through continuous optimisation of the way supply and demand are matched as a sign of a good doctor. As one specialist stated: "Long waiting lists used to be a status symbol, because it implied that you are a good doctor. Now it says: 'why are you not able to organize your clinic?'". The perception of being a good doctor expanded from medical quality to include the way the deliverance of medical quality is organised and experienced by patients. Several specialists even expressed that they started to experience it as their social responsibility to continuously improve it; that they have to optimise delays. Some even seemed embarrassed that they did not optimise delays in access before and that for many years many patients had needlessly waited for their appointments.

The 'gumball' association reveals not only a strong negative identification with providing short delays in access. It shows an association that is in line with the results of a study where interviewed doctors who do not want to apply the methods to optimise delays in access showed a fear of loss of autonomy.[6] The 'gumball quote' reveals that long delays in access may even be considered by a specialist as a vital tool for providing outpatient specialty care. Working without delays could then lead for a specialist to a similar association like a fire fighter imagining to fight a fire without one of his vital tools in his belt. With that type of identification, changing behaviour to optimise delays will conflict with the need for maintaining a positive cognitive and affective state of the self: the need for self-enhancement.[5] This negative identification will intervene with timely decisions to perform an extra session or increase capacity with another specialist when demand increases. Only when the delay is so long that it creates large problems the tendency to act on it is triggered.

The positive association with optimising delays in access seems not only connected to a positive cognitive and affective state of the self, it even seems to have become connected with the desire to perceive oneself as competent and efficacious as a doctor.[5] The specialists did not consider it to be in their sphere of influence to optimise delays. Through learning how delays can be optimised and understanding the relation between their behaviour and choices that create the delays, the meaning of delays has changed and with it their perception of their influence on the cause of delays shifted alongside their ability to influence delays.

We assume that the presentation of the WWWL model and the approach to test the WWWL model together with several other specialist outpatient clinics created confidence that it might be possible to control delays in access. The measurements in the beginning of the projects that showed that supply and demand are in balance, that there is room for efficiency improvement and that they system is designed in contradiction with the principles to optimise delays were probably important cues for the involved actors to be convinced that they should try the WWWL model. The weekly measurement of the delay in access and the reduction of the delays after making changes probably provided the cues that confirmed the hypothesis of controlling access and stimulated the will to continue.

Several specialists made a strong connection between their view of the future way they provide patient oriented care and optimising delays in access. For example enabling patients to initiate follow-up appointments when they think they need it ('patient initiated care'), which requires fast access for follow-up appointments. From a sensemaking point of view this can be important not only for a practical need to enable short delays in access, but also from the point of identification with short delays. Connecting both innovative views (optimising delays and patient initiated care) satisfies the need for self-consistency or the desire to sense and experience coherence and continuity.[5]

It seems vital to create the opportunity for the involved specialists to shift their identification with delays in access in a way that maintains a positive cognitive and affective state, expands their perception of a competent and efficacious doctor enabling them to connect it to their view how to deliver patient care in a way that they experience coherence and continuity.

7. Conclusion

Our first research question (chapter 1) 'How are demand and supply matched in specialist outpatient clinics?' has been answered for the two specialist outpatient clinics in paragraph 3. We described their characteristics and operational and aggregate planning processes and created a system description, including push and pull descriptions. Paragraph 4 reflected on the results, paragraph 5 provided an overview of the identified variables and paragraph 6 reflected on what drives the decisions how to match supply and demand. In chapter 8 the identified variables are used to create a Production Planning and Control framework for specialist outpatient clinics and test the hypotheses on the research outcomes of the specialist outpatient clinics that applied WWWL. In this paragraph we further reflect on the results of the two specialist outpatient clinics to learn how the identified variables enable us to test the hypotheses.

Hypothesis 1: The way supply and demand are matched is predominantly a push method.

A pull method is predominant when the release of capacity is based on the actual system status (pull) and a push method when the release of capacity is based on an exogenous trigger (chapter 2.3). The system description as depicted in Figure 7.2 in paragraph 3 show that the way Dermatology matches supply and demand is predominantly a push method. The release of capacity is only partly based on the system status (pull) when assistants need to schedule a (semi-) urgent appointment, cannot find a slot fast enough and change the appointment mix or extend the length of a session to meet demand. The release of capacity for the planning of sessions, the design of the week schedules and of the sessions, and the standard methods for scheduling appointments are all

based on the release of capacity based on exogenous triggers (push). The actual system status has no influence on those decisions.

The system description for Orthopaedics shows that more pull methods are applied in the way supply and demand are matched. Figure 7.3 shows that the system status can trigger extra sessions; assistants can extend the length of a session; the planning of the assistants is based on the actual schedule for that period. Also the designs of the sessions have been altered to create a flexible appointment mix.

Most of the capacity is released with no relation to the system status, but there are effective pull methods to ensure that the system status influences the actual supply based on the system status within certain limits. The way Orthopaedics matches supply and demand for the outpatient clinic is therefore not predominantly a push method, and it is not predominantly a pull method either, but a hybrid of push and pull methods. The way Orthopaedics match supply and demand between the outpatient clinic and the operating theatre is however a push method. The operating sessions are fixed a year in advance and changes have priority over outpatient sessions, without relation to system status.

In chapter 8 the changes of specialist outpatient clinics of chapter 5 and 6 are analysed with a push – pull typology to further answer the research question.

Hypothesis 2. The way specialist outpatient clinics match supply and demand causes delays
The second hypothesis is investigated under the assumption that demand and supply are equal on a yearly basis. The way supply and demand are matched by the two specialist outpatient clinics supports the hypothesis. This concerns mainly the lack of anticipating fluctuations in demand, limited flexibility to increase volume in response to actual shortage of supply, lack of flexibility to release the cycle mix of sessions depending on actual conditions. One specialist outpatient clinic furthermore lacks flexibility in the appointment mix, but the other does work with a flexible appointment mix. Furthermore, the lack of flexibility to exchange the release of outpatient sessions and operating sessions based on the system status limits their ability to optimise delays.

The second research question of the Introduction (chapter 1) 'What is required to optimise the way demand and supply are matched in specialist outpatient clinics?' has been answered in paragraph 5 by identifying the variables to deal with uncertainty in the volume of demand and of supply. In paragraph 6 another answer to the research question is provided from the perspective of what drives decisions of specialists how to match supply and demand. It seems vital to create the opportunity for the involved specialists to shift their identification with delays in access in a way that maintains a positive cognitive and affective state, expands their perception of a competent and efficacious doctor and that enables them to connect it to their view of how to deliver patient care in a way that they experience coherence and continuity.

References

1. Patton MQ. Qualitative research and evaluation methods. 3 ed. Thousand Oaks, Calif.: Sage Publications; 2002. xxiv, 598, 65 p. p.

2. Yin RK. Case study research : design and methods. 3rd ed. Thousand Oaks, Calif.: Sage Publications; 2003. xvi, 181 p. p.
3. Singhal K, Singhal J. Imperatives of the science of operations and supply-chain management. J Oper Manag. 2012 Mar;30(3):237-44. PubMed PMID: ISI:000301902900005. English.
4. Clymer JR. Simulation-based engineering of complex systems. Placentia, Calif.: J.R. Clymer & Associates; 2007. xix, 508 p. p.
5. Weick KE. Sensemaking in organizations. Thousand Oaks: Sage Publications; 1995. xii, 231 p.
6. Ahluwalia S, Offredy M. A qualitative study of the impact of the implementation of advanced access in primary health care on the working lives of general practice staff. BMC Fam Pract. 2005 Sep 27;6:39.

Appendix 1 Actual weekly schedule per dermatologist KDK, PDM and ALT

	KDK	PDM	ALT
Monday morning			
C	14	12	5
N	4	6	3
CT	3	-	-
BEH	-	-	7
AL	-	3	-
S	1	-	1

Monday afternoon			
C	12	10	9
N	3	2	2
CT	-	3	3
BEH	-	-	-
AL	2	1	2
S	-	1	-

Tuesday morning			
C	8	4	-
N	4	4	
CT	3	3	
BEH	-	4	
AL	3	1	
S	1	1	

Tuesday afternoon			
C	2	-	2
N	3	-	3
CT	-	-	3
BEH	7	-	4
AL	-	-	-
S	-	-	1

Wednesday morning			
C	2	-	10
N	4	-	6
CT	3	-	3
BEH	6	-	-
AL	-	-	1
S	1	-	-

Wednesday afternoon			
C	11	-	8
N	3	-	2
CT	-	-	-
BEH	-	-	4
AL	-	-	-
S	1	-	1

	KDK	PDM	ALT
Thursday morning			
C	-	14	6
N	10	5	5
CT	-	-	3
BEH	-	-	-
AL	3	2	2
S	-	1	1
PP	-	-	-

Thursday afternoon			
C	6	4	6
N	-	2	2
CT	-	-	-
BEH	-	-	-
AL	10	11	10
S	-	-	-
PP	5	4	3

Friday morning			
C	16	6	16
N	2	3	4
CT	3	3	-
BEH	-	5	-
AL	1	1	1
S	1	-	1

Friday afternoon			
C	7	10	12
N	1	0	3
CT	4	3	-
BEH	-	-	-
AL	1	2	1
S	-	1	-

Total 1 week	KDK	PDM	ALT
C	78	60	74
N	34	22	30
CT	16	12	12
BEH	13	9	15
AL	20	21	17
S	5	4	5
PP	5	4	3

Explanation of the codes:

N = new appointment
C = follow-up appointment
S = urgent appointment
CT = follow-up appointment, semi-urgent
BEH = appointment with treatment
AL = flexible slot for any appointment
PP = specific treatment

Appendix 2 Processes, triggers and responses for the Dermatology system visualisation

Operational planning

1. Request appointment
    a. Trigger: random, stochastic, workdays
        i. Types of appointments: New (N), New Urgent (NU) New Oncology (NO), Follow-up (C), Follow-up Urgent (CU)
        ii. Random which appointment, stochastic, on average 96 per day in total. Average distribution: 18%N; 3%NU; 6%NO; 65%C; 8%CU.
        iii. If C choose random which specialist with 40%-30%-30% distribution between the specialist 1-2-3.
    b. Output: one requested appointment: N, NU, NO, C or CU; to 'Plan appointment'
2. Plan appointment
    a. Trigger: Output of process 'Request appointment'; N, NU, NO, C or CU
    b. Input: 'Schedule': Availability per type of appointment per day
    c. Decision rules
        i. Take first available appointment per type of appointment for N, NU and NO for any specialist.
        ii. Take first available appointment per type of appointment for C for the requested specialist.
        iii. IF first available appointment for CU for the 'own' specialist (= first N, NU or NO appointment) ≤ 3 days or less book it with the same specialist, IF it is > 3 days THEN book it with first available appointment for any specialist.
        iv. IF a N or C is sooner available than a NU or CU slot, book NU or CU on N or C
        v. IF first available N is longer than three weeks and IF there is a C available in less than three weeks THEN book the N on a C.
    d. Output:
        i. One booked appointment: N, NU, NO, C or CU in 'Schedule'
        ii. For every N, NU and NO add one to 'Log production'

Aggregated planning

3. Request session cancellation
    a. Triggers:
        i. Random, stochastic, workdays, in total on average 30 days per specialist for cancellations further than the planning horizon of eight weeks
        ii. Random, stochastic, workdays, in total on average 5 days per specialist for cancellations within the planning horizon of eight weeks, but further than four weeks
        iii. Note: three weeks cancelling are standard planned in July / August for holiday, sequentially for the three specialists
    b. Input: 'Log cancellations'
    c. Decision rules:
        iv. IF 'Log cancellations' for specialist x > 4 days for month x THEN no action
    d. Output:
        v. Request cancellation: 'Date' + 'specialist x' to process 'Cancel sessions'
4. Cancel sessions
    a. Trigger: every Monday 8:00am
    b. Input: 'Request session cancellation': 'date' + 'specialist x'
    c. Decision rules
        vi. IF two specialists request cancellation on the same date THEN cancel only one, random which one
        vii. IF a request is to cancel a day within eight weeks THEN request two extra sessions between four and eight weeks. IF both sessions are possible THEN cancel the requested date. IF it's not possible THEN no action.
        viii. IF a specialist requests a cancellation in a month which has 4 sessions cancelled for that specialist THEN no action
    d. Output:
        ix. 'cancel day x' in 'Schedule' for specialist x
        x. add day x, specialist x to 'Log cancellations'
5. Check system status
    a. Trigger: every first workday of the month, 8:05am

b. Input: ‘Production log’, volume of ‘finance generating appointments’
c. Decision rules:
    xi. IF total volume of ‘finance generating appointments’ since January 1 is > 5% less than targeted for each specialist until last month THEN request enough extra sessions to compensate for the difference for that specialist
d. Output:
    xii. ‘X requested extra sessions’ to process ‘Plan extra sessions’

6. Plan extra sessions
    a. Trigger: every first workday of the month, 8:10am
    b. Input:
        xiii. Requests extra sessions per specialist from proces ‘Check system status’
        xiv. Requests extra sessions per specialist from proces ‘Cancel sessions’
        xv. Number of sessions per specialist per week from ‘Schedule’
    c. Decision rules:
        xvi. A specialist can perform maximum eleven sessions per week
        xvii. The team can support a maximum of 32 sessions per week
        xviii. IF cancellation is requested within planning horizon THEN request extra days in process ‘request extra work days’. IF extra sessions are possible, THEN plan extra session.
    d. Output:
        xix. Plan extra sessions on date x in ‘Schedule’
        xx. Request extra workday in month x to process ‘Request assistants extra workday’
7. Request assistants extra work day
    a. Trigger:
        xxi. Requests from process ‘Plan extra sessions’ for date x
    b. Decision rules:
        xxii. IF request > planning horizon of assistants (two full months ahead) THEN action
        xxiii. IF request < planning horizon of assistants THEN 50% action
        xxiv. IF all assistants have an extra workday in that month THEN no action
    c. Output:
        xxv. Extract on day from ‘Log holidays minus extra workdays’ from random assistant that has no extra workdays in that month
        xxvi. Confirm date x extra assistant available to process ‘Plan extra sessions’
8. Assistant requests holiday
    a. Trigger
        xxvii. Random, stochastic, workdays, in total exactly 8 days per assistant for holidays further than the assistant planning horizon of two full months further
        xxviii. Note: Three weeks holiday are standar planned in July / August for two assistants, sequentially for all six assistants, synced with the holidays of the specialists
    b. Output: Request cancellation ‘Date x’ + ‘assistant x’ to process ‘Plan assistant holidays’
9. Plan assistant holidays
    a. Trigger: Requests from process ‘Assistant requests holiday’
    b. Input:
        xxix. ‘Log holidays minus extra work days’
        xxx. ‘Schedule’: number of assistants per date
    c. Decision rules:
        xxxi. IF ‘Log holidays minus extra workdays’ for assistant x =8 THEN no action
        xxxii. IF number of assistants on date x in ‘Schedule’ = 5 THEN no action
    d. Output:
        xxxiii. Add one + date to ‘Log holidays minus extra work days’
        xxxiv. Detract on assistant on date x from ‘Schedule’

Appendix 3 Processes, triggers and responses for the Orthopaedic system visualisation

Operational planning

1. Request appointment
    a. Trigger: random, stochastic, workdays
        i. Appointment types: New (N), New Urgent (NU), Follow-up (C), Follow-up Urgent (CU)
        ii. Random which appointment, stochastic, on average 96 per day in total. Average distribution: 18%N; 3%NU; 6%NO; 65%C; 8%CU.
        iii. If C choose random which doctor with equal distribution between the specialist.
    b. Output: one requested appointment: N, NU, C or CU; to 'Plan appointment'
2. Plan appointment
    a. Trigger: Output of process 'Request appointment'; N, NU, C or CU
    b. Input: 'Schedule': Availability per type of appointment per day
    c. Decision rules
        i. Take first available appointment per type of appointment for N and NU for any specialist, fellow or resident.
        ii. Take first available appointment per type of appointment for C for the requested doctor.
        iii. IF first available appointment for CU for the 'own' doctor (= first N, NU or NO appointment) ≤ 3 days or less book it with the same specialist, IF it is > 3 days THEN book it with first available appointment for any specialist.
        iv. IF a N or C is sooner available than a NU or CU slot, book NU or CU on N or C
        v. IF first available N is longer than three weeks and IF there is a C available in less than three weeks THEN book the N on a C.
    d. Output:
        i. One booked appointment: N, NU, C or CU in 'Schedule'
        ii. For every N, and NU add one to 'Log production'

Aggregated planning

3. Request session cancellation
    a. Triggers:
        xxxv. Random, stochastic, workdays, in total on average 30 days per doctor for cancellations further than the planning horizon of eight weeks
        xxxvi. Random, stochastic, workdays, in total on average 3 days per doctor for cancellations within the planning horizon of eight weeks, but further than four weeks
        xxxvii. Note: three weeks cancelling are standard planned in July / August for holiday, sequentially for the all doctors, three at the same time
    b. Input: 'Log cancellations'
    c. Decision rules:
        xxxviii. IF 'Log cancellations' for specialist x > 4 days for month x THEN no action
    d. Output:
        xxxix. Request cancellation: 'Date' + 'specialist x' to process 'Cancel sessions'
4. OT scheduling orders session cancellation
    a. Triggers:
        i. Random, stochastic, workdays, in total on average 5 days per doctor for cancellations further than the planning horizon of eight weeks
        ii. Random, stochastic, workdays, in total on average 2 days per doctor for cancellations within the planning horizon of eight weeks, but further than four weeks
    b. Output:
        i. Order cancellation: 'Date' + 'specialist x' to process 'Cancel sessions'
5. Cancel sessions
    a. Trigger: every Monday 8:00am
    b. Input: '
        xl. Request session cancellation': 'date' + 'specialist x'

xli. OT scheduling: ‘date’ + ‘specialist x’

c. Decision rules

xlii. IF two specialists request cancellation on the same date THEN cancel only one, random which one

xliii. IF a request is to cancel a day within eight weeks THEN request two extra sessions between four and eight weeks. IF both sessions are possible THEN cancel the requested date. IF it’s not possible THEN no action.

xliv. IF a specialist requests a cancellation in a month which has 4 sessions cancelled for that specialist THEN no action

xlv. IF a date is ordered by ‘OT scheduling’ THEN cancel the sessions on that day

d. Output:

xlvi. ‘cancel day x’ in ‘Schedule’ for specialist x

xlvii. add day x, specialist x to ‘Log cancellations’

6. Check system status

a. Trigger: every first workday of the month, 8:05am

b. Input:

xlviii. ‘Log production’, volume of ‘finance generating appointments’

xlix. ‘Log access’: number of days to third available appointment for each doctor

c. Decision rules:

l. IF total volume of ‘finance generating appointments’ since January 1 is > 5% less than targeted for each specialist until last month THEN request enough extra sessions to compensate for the difference for that specialist

li. IF access > 6 weeks for a doctor THEN request an extra session for that doctor

d. Output:

lii. ‘X requested extra sessions’ to process ‘Plan extra sessions’

7. Plan extra sessions

a. Trigger: every first workday of the month, 8:10am

b. Input:

liii. Requests extra sessions per specialist from process ‘Check system status’

liv. Requests extra sessions per specialist from process ‘Cancel sessions’

lv. Number of sessions per specialist per week from ‘Schedule’

c. Decision rules:

lvi. A specialist can perform maximum eleven sessions per week

lvii. The team can support a maximum of 32 sessions per week

lviii. IF a cancellation is requested within the planning horizon THEN request extra workdays in process ‘request extra work days’. IF the extra sessions are possible, THEN plan extra session.

d. Output:

lix. Plan extra sessions on date x in ‘Schedule’

lx. Request extra workday in month x to process ‘Request assistants extra workday’

8. Request assistants extra work day

- Trigger:

lxi. Requests from process ‘Plan extra sessions’ for date x

e. Decision rules:

lxii. IF request > planning horizon of assistants (two full months ahead) THEN action

lxiii. IF request < planning horizon of assistants THEN 50% action

lxiv. IF all assistants have an extra workday in that month THEN no action

f. Output:

lxv. Extract on day from ‘Log holidays minus extra workdays’ from random assistant that has no extra workdays in that month

lxvi. Confirm date x extra assistant available to process ‘Plan extra sessions’

9. Assistant requests holiday

a. Trigger

lxvii. Random, stochastic, workdays, in total exactly 8 days per assistant for holidays further than the assistant planning horizon of two full months further

lxviii. Note: Three weeks holiday are standard planned in July / August for two assistants, sequentially for all six assistants, synced with the holidays of the specialists

b. Output: Request cancellation 'Date x' + 'assistant x' to process 'Plan assistant holidays'

10. Plan assistant holidays
    a. Trigger: Requests from process 'Assistant requests holiday'
    b. Input:
        lxix. 'Log holidays minus extra work days'
        lxx. 'Schedule': number of assistants per date
    c. Decision rules:
        lxxi. IF 'Log holidays minus extra workdays' for assistant x =8 THEN no action
        lxxii. IF number of assistants on date x in 'Schedule' = 5 THEN no action
    d. Output:
        lxxiii. Add one + date to 'Log holidays minus extra work days'
        lxxiv. Detract on assistant on date x from 'Schedule'

Appendix 4 Fictive schedule for Orthopaedics

Monday	O1	O2	O3	O4	O5	O6	F	R1	R2	R3	R4	R5	∑S	∑OT
Morning	S	S	S	-	OT	-	OT	S	S	OT	-	OT	5	4
Afternoon	S	S	S	-	OT	-	OT	OT	S	OT	-	OT	4	5

Tuesday	D1	D2	D3	D4	D5	D6	F	R1	R2	R3	R4	R5	∑S	∑OT
Morning	-	-	OT	OT	S	S	S	OT	-	S	OT	S	5	4
Afternoon	-	-	-	OT	S	OT	S	S	OT	S	OT	S	5	4

Wedn.	D1	D2	D3	D4	D5	D6	F	R1	R2	R3	R4	R5	∑S	∑OT
Morning	OT	OT	-	S	OT	S	OT	OT	-	OT	S	-	3	6
Afternoon	OT	S	-	S	OT	S	-	S	S	S	OT	-	6	3

Thursday	D1	D2	D3	D4	D5	D6	F	R1	R2	R3	R4	R5	∑S	∑OT
Morning	S	OT	S	-	S	OT	OT	-	OT	OT	S	-	4	5
Afternoon	-	-	-	-	-	-	-	S	S	-	-	-	2	-

Friday	D1	D2	D3	D4	D5	D6	F	R1	R2	R3	R4	R5	∑S	∑OT
Morning	S	S	OT	OT	-	OT	S	S	-	S	OT	OT	5	5
Afternoon	S	OT	S	OT	-	OT	S	OT	-	S	S	S	6	4

Explanation of the codes:
O = orthopaedist
R = resident
S = session (outpatient clinic)
OT = operating theatre

Part III SYNTHESIS AND DISCUSSION

In part III we brought the theory from part I and the research findings from part II together. This provided not only an overview to draw conclusions, but also created a synthesis in chapter 8 to design a framework to optimise delays in access to specialist outpatient clinics. In chapter 9 we used the findings of our research to reflect on the theory of part I, we discussed the main research findings of part II, including methodological considerations and we discussed the main results of the synthesis of chapter 8 and drew final conclusions to answer the problem statement of this thesis and provide recommendations for practice and further research.

CHAPTER 8

Production Planning and Control framework and typology for specialist outpatient clinics

Parts of this chapter have been submitted as: What causes delays in access to specialist outpatient clinics? Marc (MBV) Rouppe van der Voort, Frits (GG) van Merode, Bart (HJJM) Berden.

1. Introduction

Part I presented the theory on delays in access to specialist outpatient clinics. Part II presented the results of our research on the practice of matching supply and demand in specialist outpatient clinics and the effects on delays in access. This chapter brings theory and the results of our research together, not only to provide an overview and draw conclusions, but also to design a framework and a typology of push and pull combinations to optimise delays in access to specialist outpatient clinics.

Paragraph 2 uses the research results of chapter 5, 6 and 7 to identify the causes of delays in access to specialist outpatient clinics. Paragraph 3 uses the same results to identify the methods that were applied to address the causes. In paragraph 4 the results of paragraph 2 and 3, and the identified variables of chapter 7 on how to optimise delays in access are used to create a Production Planning and Control (PPC) framework for specialist outpatient clinics. Paragraph 5 uses the identified variables and the PPC framework to create a push-pull typology for specialist outpatient clinics. The purpose of the typology is to distinguish between combinations of the most important variables of the PPC framework that determine the effectiveness of efforts to optimise delays in access. Paragraph 6 uses the push-pull typology to analyse the results and sustainability of Working Without a Waiting List (WWWL) and determine the most effective typologies. In the final paragraph 7 the following research question is answered:

What is required to optimise the way demand and supply are matched?

To answer the question we reflect on four hypotheses:

1. *The way supply and demand are matched is predominantly a push method.*
2. *The way specialist outpatient clinics match supply and demand causes delays.*
3. *The specialist outpatient clinics that reduced their delays and sustained the results introduced pull methods.*
4. *Other departments that influence the matching of supply and demand of the specialist outpatient clinics are required to be included in the 'pull' system to optimise delays.*

The hypotheses are based on the outcome of the theory on delays in access in chapter 2, 3 and 4.

2. Identified causes of delays

The results of chapters 4, 5 and 6 enable us to create an overview of the identified causes of delays in access to specialist outpatient clinics. As explained in chapter 2.2, there are three main causes for delays:

1. Demand is larger than supply.
2. There is uncertainty in the volume of demand.
3. There is uncertainty in the volume of supply.

Ad 1. Demand is larger than supply.

Supply is the number of appointments actually made available to patients. Demand is the number of appointments requested by patients. A shortage of (specialist) capacity can be the cause for insufficient supply of appointments, but it can also be the result of too much waste of capacity.

Wasted capacity can be the result of too much cancelled sessions or unused appointments within sessions. Appointments slots can remain unused because no patients are planned on a slot or because the patient does not show up for the appointment.

A2 and ad3: Fluctuations in demand and supply are not synchronised.
We distinguish three types of causes that *together* can cause fluctuations in demand and supply not the be synchronised:

a) Demand fluctuates.
b) Supply of sessions fluctuates without relation to fluctuations in actual demand.
c) The appointment mix that is supplied does not match the appointment mix of demand.

a) Demand fluctuates.
Variation in the volume of demand for a new appointment is exogenous but can have seasonal trends. Demand for follow-up appointments that are planned during the last appointment are endogenous variation and can be influenced, unplanned follow-up appointments are exogenous. Planned follow-up appointments have a correlation with seasonal trends for new appointments, and with peaks in the supply of new appointments. Demand for urgent appointments has random exogenous variation and can have seasonal variation. In conclusion, part of demand is random exogenous variation, but a substantial part of demand can be influenced.

b) The supply of sessions fluctuates without relation to fluctuations in actual demand.
Sessions can be cancelled without consideration of the required supply for that period. Sessions can also be added with no relation to the required supply of that period, for example to compensate cancelations in another period. Another cause can be that a specialist outpatient clinic is not able to respond adequately to fluctuations in demand. This can be caused by the inability to notice fluctuations (no feedback information). If fluctuations are noticed, the specialist outpatient clinic might not be able to respond to the noticed variation because of an inability to extend sessions to see more patients or an inability to perform more sessions. The latter can be caused because there is no time available from specialists to perform extra sessions, no assistants available to support the sessions or no rooms available to see the patients. The lack of availability of the specialist can be caused by inflexibility of the operating theatre sessions, the lack of buffer capacity and the lack of willingness from the specialist. Scheduling assistants further ahead than the required extra session and lack of buffer capacity for the assistants or the lack of willingness to work extra hours can cause the lack of availability of assistants. The lack of rooms can be caused by the lack of flexibility to use rooms for different purposes.

c) The appointment mix that is supplied does not match the appointment mix of demand.
The inability to match the mix of supplied appointments with the mix of demanded appointments can be caused by fixed appointment slots ('single queues', chapter 2.6) and fixed types of sessions.

In conclusion, when demand is larger than supply, the question is how much capacity is wasted and whether that waste is greater than the shortage of supply. If there is enough supply to meet demand, the next question is whether endogenous variation in supply and demand or the inability

to notice and react to fluctuations in demand are causing delays. In Figure 8.1 a summary of the identified causes of delays in access to specialist outpatient clinics is provided.

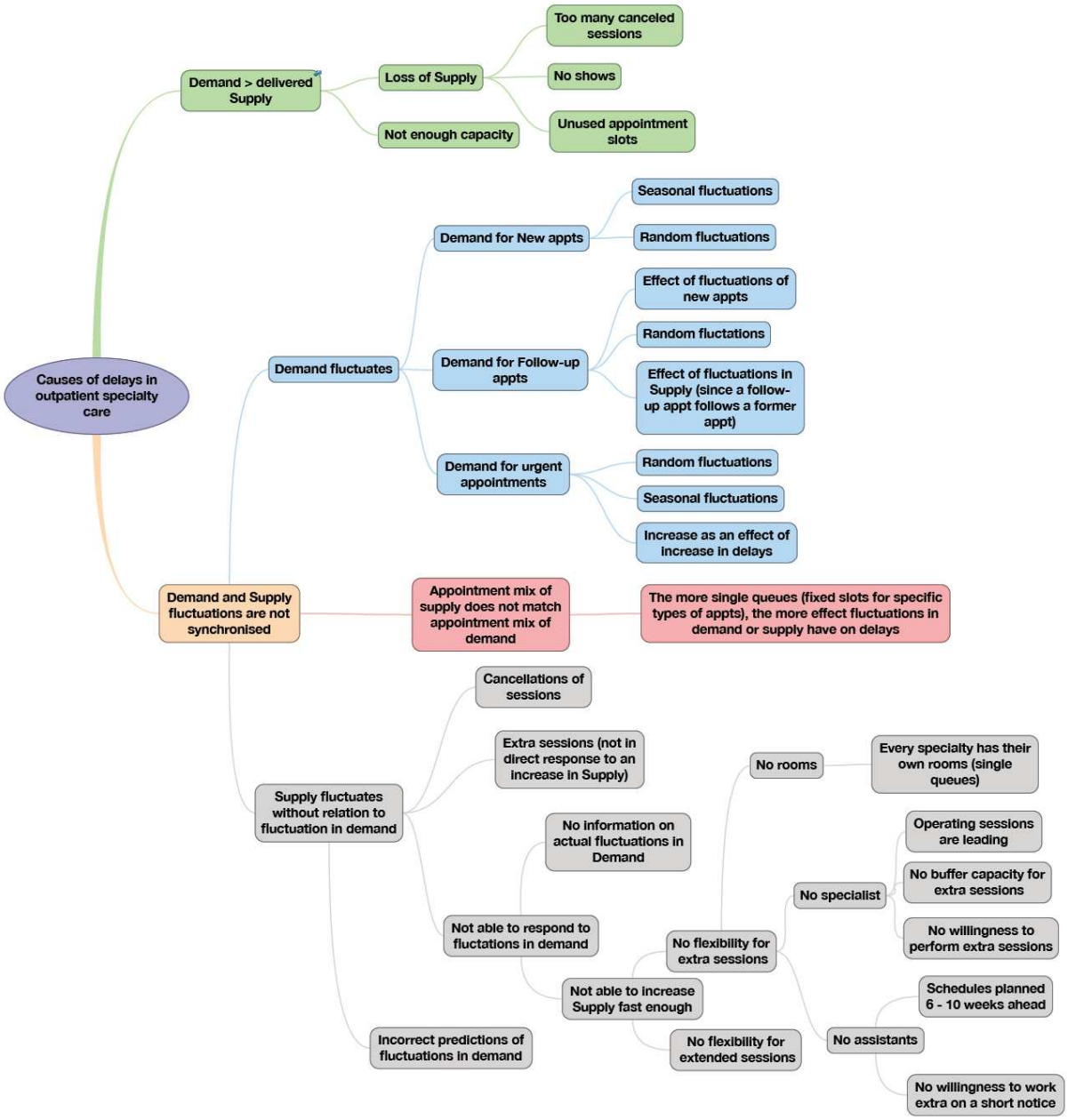


Figure 8.1 Summary of the identified causes of delays in access

3. Identified methods to address the causes of delays

This paragraph provides an overview of methods that are identified in chapter 4, 5 and 6 to address the identified two main causes of delays in the former paragraph.

1. Demand is larger than supply.

If demand is larger than supply, either increasing capacity or reducing the waste can increase supply. We identified three methods to reduce the waste of capacity.

a) Reduce unwanted cancelations of sessions.

Specialists cancel sessions because of holidays, conferences, education or other reasons. Agreements on the spread of sessions per month and the minimum number of sessions per week and monitoring the actual supply of sessions can reduce unwanted cancelations. If enough sessions are planned, but too much cancelled ad hoc, the most applied method is to agree on a 'holy horizon': no cancellations of sessions within usually 6 or 8 weeks from now.

b) Reduce no shows.

The specialist outpatient clinics in chapter 4 reported an average of 8.2% of wasted capacity due to no shows (a patient does not show up for a appointments). Research shows a correlation between longer delays in access and no shows.[1,2] This effect increases if patients have negative emotions to visit the doctor, feel disrespected and if they do no understand the scheduling system. There are methods to remind patients of their appointment, but the specialist outpatient clinics reduced no shows from 8.2% to 4.6% on average, simply as an effect of reducing delays and improving the planning methods.

c) Reduce unused appointment slots.

Unused appointment slots can be the result of working with fixed appointment slots for specific types of appointments that simply are less requested in that period. Some introduced 'Emmenthaler screening' methods where assistants scan the upcoming sessions for empty slots and pro-actively fill them up with other types of appointments. Most specialist outpatient clinics however solved this problem completely by eliminating (almost) all types of appointment slots and create a flexible appointment mix per session.

In conclusion, if demand is larger than supply and supply is wasted, methods to increase supply can solve the problem. If they do not increase supply enough, capacity needs to be increased.

2. Fluctuations in demand and supply are not synchronised.

If there is sufficient supply (including buffer capacity) but still a delay in access, there is an inability to synchronise supply with fluctuations in demand. To address this cause, the specialist outpatient clinics applied methods with three different aims:

a) Increase the ability to notice fluctuations in the volume of demand.
b) Increase flexibility in the volume of appointments.
c) Increase flexibility in the appointment mix.

2) Increase the ability to notice fluctuations in the volume of demand.

None of the specialist outpatient clinics was able to directly monitor actual demand. They did measure weekly the delay in access per specialist for new and follow-up appointments. One specialist outpatient clinic also measured the percentage of free space in the agenda for the next six weeks. This provides a feedforward signal when supply is predicted to be insufficient to meet demand, before delays actually occur. Another reported method is simply to ask on a daily or weekly basis to the assistants whether there are any problems to plan the appointments.

a) Increase flexibility in the volume of appointments.

Some specialist outpatient clinics planned buffer sessions in the agenda of a specialist that are opened when the monitoring of the free agenda space indicates it. Some simply request specialists

whether they are able to perform extra sessions at their convenience, highly depending on the willingness of individual specialists.

One specialist outpatient clinic agreed on a norm of delay in access that triggers the release of extra appointment slots at the end of all the sessions in a week for a specialist. For others it is the individual specialists and the team leader that agree ad hoc on extending sessions or not.

b) Increase flexibility in the appointment mix.
The most applied method to increase flexibility in the mix of appointments is the same as the above described method to reduce unused appointment slots: by stopping the use of fixed appointment slots the mix of supply per type of appointment automatically becomes the same as demand. Some kept fixed slots but created decision rules under which circumstances they can be released for specific other types of appointment slots. None create flexibility in the mix of sessions.

4. Production Planning and Control framework for specialist outpatient clinics

4.1 Introduction

We use the identified variables that the specialist outpatient clinics used to optimise the way supply and demand are matched (chapter 7), the identified causes of delays in access (paragraph 2) and identified methods that the specialist outpatient clinics used to improve access (paragraph 3) to adapt the Production Planning and Control (PPC) framework (chapter 2.5) into a PPC framework for specialist outpatient clinics. The PPC framework relates the identified variables in a logical and hierarchical manner and adds several elements that according to theory should enable specialist outpatient clinics to further optimise delays in access. We use the framework in the next paragraph for our research to identify the most important variables to optimise delays in access. Figure 8.2 illustrates this PPC framework. The three levels 'Strategy', 'Tactics' and 'Control' represent the long to short term planning issues of a specialist outpatient clinic. The rectangular boxes are decisions and the rounded boxes are outcomes. The oval boxes represent inputs that are generated outside this planning hierarchy. Finally, the arrows represent the interdependence of the modules.

4.2 Strategic level

On the strategic level capacities are planned. The capacity planning is based on the 'Production agreement' after 'Negotiations with the insurance companies'. Inputs for the negotiations are 'Forecasting' and 'National agreements' on production growth. The 'Production agreement' is multiplied with the decided 'Buffer percentage' to optimise delays and with the anticipated 'Return rate' (number of follow-up appointments per new appointment) and 'Appointment length' for each type of appointment to determine the number of required sessions, which determines the required 'Specialist capacity'. The 'Specialist capacity planning', 'Facilities planning' and 'Workforce planning' lead to the 'Aggregate planning'. The 'Workforce planning' is also based on the labour policies. The 'Aggregate planning' combines these outcomes with the anticipated seasonal fluctuation in the volume of demand for new appointments, corrected by the anticipated volume of demand for follow-up appointments that can be directed to periods with less demand or with less reduction in supply ('Variables').

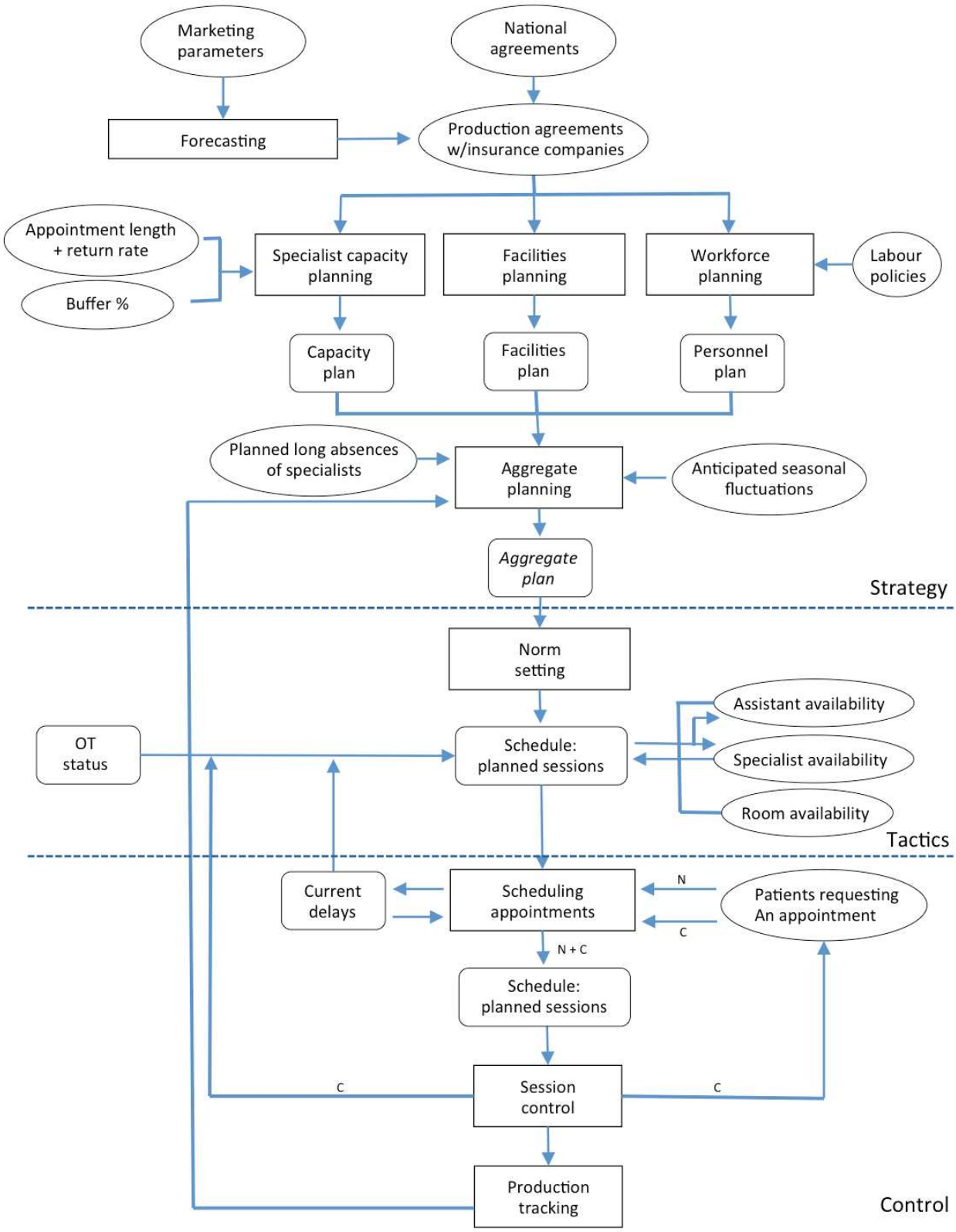


Figure 8.2 Production Planning and Control (PPC) framework for specialist outpatient clinics

The resulting aggregate plan offers four elements:

1. Rough estimate of required sessions per month.
2. Planning horizon for appointments.
3. Cycle horizon (chapter 7.5).
4. Cycle mix (chapter 7.5).

4.3 Tactical level

On the tactical level the release of sessions is based on the strategic plans and in reaction to the actual system status. The 'Aggregate plan' is complemented at 'Norm setting' with norms for the delays in access per type of appointment that can change the decision rules for assistants to schedule appointments and that can trigger a request for additional sessions from 'Specialists availability' and 'Assistants availability'. If supply is insufficient for a specific period the norms trigger temporary adjustments of the cycle mix of sessions and the decision rules to schedule appointments. For example, if too many oncology sessions are cancelled, a general session may be changed into an oncology session. If the total volume of supply is insufficient, a decision rule change can be to postpone more follow-up appointments appointments.

If the specialists also operate in the operating theatre (OT), norms for the delay for surgery are also set with triggers that request extra operating sessions at the expense of outpatient sessions or vice versa. These norms are combined with feedback from 'scheduling appointments' on actual delays to outpatient sessions and from 'OT system status' to surgery. Together with the output from 'Specialist availability' and the feedback from 'session control' where future demand for follow-up appointments is tracked, the release of sessions per specialist is determined for the next cycle.

Figure 8.3 depicts the aggregate planning process from an information flow point of view. The graph on the upside of the figure labelled 'Specialist outpatient clinic' shows how much appointment time has been released on an aggregate level for each week for the next six weeks ('supply'). This is the same as the box 'Schedule: planned sessions' in Figure 8.2. The filling per week shows how much time is released on the operational level as appointments with individual patients. The remaining time is still available. In this example, the delay in access is four weeks, since there is appointment time available in the fifth week. The specialist outpatient clinic releases the sessions seven weeks in advance in this example. The dotted line in week seven shows the time that can become available if the blueprint session mix for that week is released without cancellations of sessions ('capacity'). The decision box concerns the decision whether to release all sessions or part of it. This can be based on information on the availability of resources (specialist, support and facilities), anticipated demand (either based on past experience and / or based on actual developments and their expected impact on demand, and the system status of the clinic itself, in this case the status of how much available time is filled for the next six weeks and what the access is. Figure 8.3 also shows the system status for the 'operating theatre' on the downside of the figure. The upward arrow with the text 'Downstream system status' indicates it is input for the decision how many sessions to release for the specialist outpatient clinic. In this example there is 25% operating time available to plan operations in the next six weeks. If this is considered too low, a decision can be to release an outpatient session less and perform an extra operating session, or vice versa if it is considered more important to offer faster access for outpatient appointments. The norms for these decisions are displayed in Figure 8.2 as 'norm setting'.

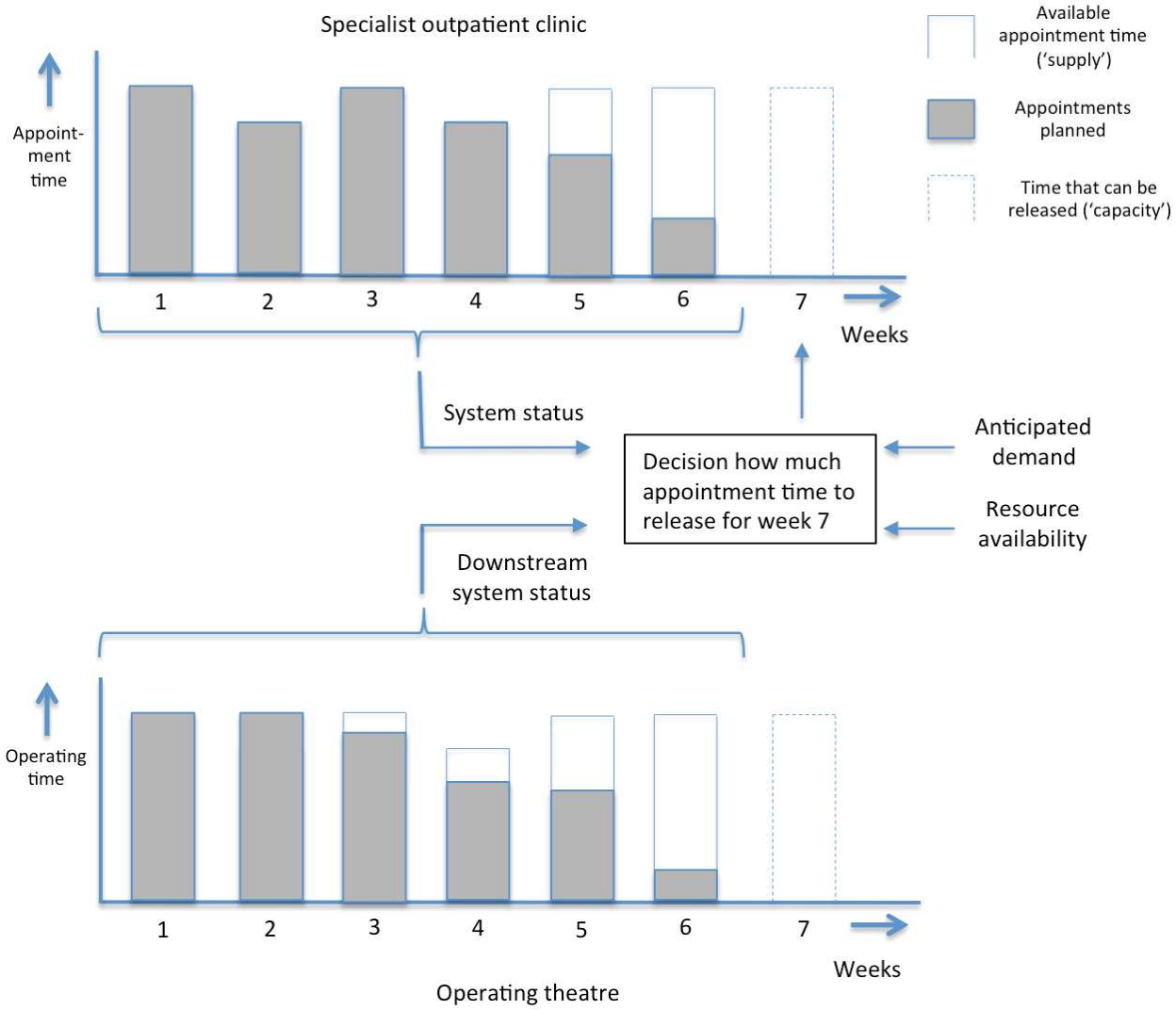


Figure 8.3 Aggregate planning process for a specialist outpatient clinic in coordination with the operating theatre

4.4 Control level

On the control level the requests of patients for an appointment are planned in the sessions. The codes are: N for 'new appointments' and C for 'follow-up appointments'. The applied decision rules depend on the output from 'Schedule: planned sessions' and is based on the 'Current delays' per type of appointment. This results in sessions with appointments that are performed by the specialists and supported by the assistants. 'Session control' is concerned with the day-to-day execution of the sessions and provides feedback to 'Production tracking' on performance measures as 'production', 'no-shows' and 'end time of sessions'. Since appointments often end with a new follow-up appointment, there is a feedback loop to 'Patients requesting an appointment' and there is a second feedback loop that provides this information as anticipated demand for follow-up appointments for 'Aggregate planning'. 'Production tracking' monitors anticipated and actual demand, production volume and system effectiveness and provides feedback to 'Aggregate planning' to consider measures if production is too high or low in relation to the production agreements. It can also lead to a structural adjustment of the required cycle mix if a structural change in demand is detected.

4.5 The PPC framework and the drivers how supply and demand are matched

One finding of our research on sustainability of reduced delays in access is the vital role of clinical leadership of the specialists (chapter 6). This implies that specialists need to be directly involved in the design of a PPC framework for their own practice and that they need to lead or at least be involved in determining and further development of the decisions rules to release capacity on a tactical and on an operational level. It seems important that they connect the decision rules to their vision on quality of care and the priorities they set for it, while understanding the system dynamics (chapter 7.5). Clinical leadership is required of one or several specialists, who are committed, invest time and energy and confront other specialists when they do not uphold proper work practices. For all specialists, it's important to provide accurate feedback to support decisions and direct behaviour to optimise delays in access. If too little feedback is provided it will result in wrong decisions because they are uninformed. If the feedback is too late there is a risk of overreacting. If there is too much feedback it becomes difficult to understand what information to act on. It also becomes difficult to relate the information to the identification of the specialists with delays in access and the feedback becomes noise. The feedback needs to be limited to the required information to interpret problems or developments in relation to delays in access, but also needs to be related to the required information to maintain a positive and cognitive state and what the specialists consider important for the way they want to provide care.

5. Push-pull typology for specialist outpatient clinics

The PPC framework for specialist outpatient clinics (figure 8.2) contains many variables that can be applied in a variety of ways by specialist outpatient clinics. This implies a large number of possible combinations that can result in different delays in access. In this paragraph we present a typology by combining the most important variables. In paragraph 6 this typology is used to analyse the results of the application of WWWL and to determine the most effective typologies. To identify the variables that explain the delays in access we combine the elements of the PPC framework for specialist outpatient clinics, the factors that were identified by the respondents that determine the sustainability of the reduction of delays (chapter 6) and the push – pull description from the Theoretical framework (chapter 2.3 and 2.4). This results in a push – pull typology for specialist outpatient clinics.

Whether a push or pull method is predominant is determined by whether capacity is released based on the actual system status (pull) or that it is released on an exogenous trigger (push). On an aggregate level this concerns the release of sessions in 'Schedule: planned sessions' and on an operational level this concerns the release of appointment slots in 'Schedule: planned appointments'. The aggregate release of capacity is pull based if the actual delays in access, volume of demand and supply of both the specialist outpatient clinic and if relevant the operating theatre determine the number and mix of sessions that are planned. The most important design choice is whether the weekly (or two-weekly) mix of types of sessions is a:

1. Fixed mix of sessions each week (push). Cancellations of sessions do not change the other sessions and extra sessions or extended sessions are not triggered by the system status.
2. Flexible mix (pull). For each week the right mix of sessions is determined using the norms, decision rules and actual volume of supply and demand and delays in access.

3. Fixed mix with flexibility (push-pull hybrid). The flexibility concerns the possibility to add extra appointments slots to sessions or add extra sessions when actual demand exceeds supply.

A related variable in option 2 and 3 is whether the operating theatre is equally flexible to adapt the number of operating sessions in relation to actual circumstanced, versus the pressure to maximally use all allocated operating sessions.

The operational release of capacity is pull based if the date where assistants plan appointments is determined in relation to actual delays in access, volume of demand and supply per type of appointment. The most important design choice is whether sessions have:
1. Fixed slots for each type of appointment (push).
2. Flexible slots where any type of appointment can be booked (pull).
3. Mix forms of fixed and flexible slots (push-pull hybrid)
    a. Partly fixed and partly flexible slots in sessions
    b. Fixed slots that can partly be used flexibly under defined circumstances

The combination of the aggregate and operation levels of push and pull methods in specialist outpatient clinics leads to a typology of nine combinations of push-pull methods, visualised as a matrix in Table 8.1.

Table 8.1 Typology of push-pull methods in specialist outpatient clinics

	1. Flexibility in session mix		
2. Flexibility in appointment mix	Fixed mix (push)	Flexible mix (pull)	Hybrid mix (push-pull)
Fixed mix (push)	(1)	(2)	(3)
Flexible mix (pull)	(4)	(5)	(6)
Hybrid mix (push-pull)	(7)	(8)	(9)

6. Push-pull analysis of the WWWL results

We use the identified typology to analyse the results of the fourteen specialist outpatient clinics that reduced their delays in access with the concept of Working Without a Waiting List (WWWL) (chapter5) and eleven of which were able to sustain their results (chapter 6). The interventions from WWWL that concern the application of push or pull methods (Table 5.1, chapter 5) are listed below. We make a distinction between pull methods and interventions that enable pull methods or minimise the potentially negative effects of push methods on delays in access:

1. Anticipate fluctuations in demand (aggregate planning – push).
2. No cancelling of sessions within six weeks (aggregate planning – minimise negative effects of push method to reduce capacity with no relation to actual system status).
3. Create the possibility of extra sessions on a short notice (aggregate planning – pull, in combination with 4. Weekly measurement of access time).
4. Weekly measurement of access time (aggregate planning – enabler for pull).

5. Short planning horizon (aggregate planning – enabler for pull).
6. Minimize the number of queues (operational planning – enabler for pull).
7. Patient initiated care for follow-up appointments (operational planning – pull).
8. Pro-actively fill empty slots (operational planning – minimise negative effects of push method to fix slots).

The two main strategies of WWWL are to reduce unwanted variation in supply (intervention 1 and 2) and increase the flexibility of supply (intervention 3, 5 and 6). Flexibility in supply can be defined as the extent to which a system is able to create variety in the mix and/or volume of resources in a system, so that the system state can be changed without adding resources or fundamentally changing the resources.[1] Increasing flexibility is part of a strategy to increase the responsiveness of a clinic. The two most important pull method enablers in WWWL are the flexibility in the volume of sessions on a week level and the flexibility in the mix of appointments within sessions. With the data from the results of WWWL we constructed in Table 8.2 an overview which choices the specialist outpatient clinics made on these variables, and place them in the three variations of fixed, flexible and fixed with flexibility.

Table 8.2 Overview of the push-pull methods applied by the WWWL specialist outpatient clinics

		Access	Access	Access	Change	Aggregate: session mix			Operational: appointment mix			Typology	
Clinic		Start project	End project	3 years later	Compared start	Fixed	Flexible	Fixed, with flexible buffers	Fixed	Flexible	Fixed, with flexibility	Before	After
1	Pediatrics	15	1	1	93%	B		A	B	A		1	6
2	Reumatology	43	10	5	88%	B		A	B	A		1	6
3	Dermatology	35	3	7	80%	B		A	B	A	:	1	6
4	Cardiology	56	32	13	77%	B		A	B/A			1	3
5	Orthopaedics	128	81	45	65%	B/A		:	B	A		1	4
6	Reumatology	65	44	23	65%	B/A		:	B		A	1	7
7	Neurology	5	1	2	60%	B/A		:	B		A	1	7
8	Orthopaedics	31	17	13	58%	B		A	B	A		1	6
9	Orthopaedics	63	20	30	52%	B		A	B	A		1	6
10	Urology	21	10	12	43%	B		A	B	A		1	6
11	ENT	38	20	28	26%	B/A			B/A			1	1
12	Cardiology	32	13	26	19%	B/A			B/A	A		1	4
13	Dermatology	33	19	28	15%	B/A			B		A	1	7
14	Neurology	15	8	14	7%	B/A			B/A			1	1
	Average	**41**	**20**	**18**	**57%**								

The first column lists the delay in access to each specialist outpatient clinic at the start of their WWWL project. Access is measured as the third available appointment for new and follow-up appointments, as an average for all specialists. The second column lists access after the application of WWWL and the third column lists access three years later. The fourth column lists the percentage of reduction in access compared to three years after the application of WWWL to access before WWWL. The specialist outpatient clinics are listed in order of that percentage. As described in chapter 6, we use 30% reduction as the norm for substantial impact and therefore the norm for sustainability of the results.

The next columns show the two variables of operational and aggregate flexibility with all three methods of push (fixed), pull (flexible) or push-pull (fixed with flexibility). For each specialist

outpatient clinic is identified what the method was before (B) and after (A) WWWL application. A ':' indicates that it was attempted, but failed. The last column lists which combination of push-pull from the push – pull typology was applied by each specialist outpatient clinic before and three years after applying WWWL. Outpatient specialty clinics 11-14 have less than 30% reduction in delay in access, but as explained in chapter 6, specialist outpatient clinic 12 did sustain their results and 14 claims the same, but it was not possible to confirm their claim and these are therefore excluded from the analysis in this paragraph.

Each specialist outpatient clinic worked push based before WWWL on both the aggregate and operational level (type 1). After WWWL five type of push – pull combinations are visible, four if only the sustainable specialist outpatient clinics are considered. No specialist outpatient clinic has sustained their results without becoming more flexible by applying a pull method.

No specialist outpatient clinic made the session mix per week flexible; they all kept using fixed week schedules. Of the ten sustainable specialist outpatient clinics, seven did introduce a variety of formal methods to create flexibility in the volume of appointments by adding extra sessions or extend sessions. Some did already double book appointments (plan two patients at the same time) before WWWL and some incidentally performed an extra session. These actions were not directed at optimising access or the total system performance however, but to catch up on production parameters or to provide fast access for urgent appointments. The three specialist outpatient clinics that sustained their results with no extra flexibility in volume of supply did formally agree on extra flexibility, but stated that individual specialists simply do not agree on extra or extended sessions. Both specialist outpatient clinics that did not sustain their results (11 and 13) did not create extra flexibility on an aggregate level.

On the operational level of planning appointments, ten out of eleven sustainable outpatient speciality clinics became more flexible. Seven introduced a completely flexible mix of appointments within sessions, three introduced some flexibility. Only outpatient Cardiology clinic number 4, sustained its results without creating operational flexibility. They did create a new schedule, but again fixed appointment slots because they offer patients 'one-stop-shop' visits with diagnostic activities of adjacent departments tightly coupled with the outpatient appointments. They compensate the lack of flexibility in the appointment mix with flexibility in the volume of sessions. They did reduce their number of types of appointments to only two (new and follow-up), which minimised the negative effects of working with fixed. Of the two specialist outpatient clinics that did not sustain their results one uses a fixed mix of appointments with some flexibility and the other uses a fixed mix of appointments.

When the aggregate and operational views are taken together, the specialist outpatient clinics apply five combinations of push-pull methods from the typology of Table 8.1: 1,3,4,6 and 7. In conclusion, all specialist outpatient clinics that sustained their results applied pull methods to increase their flexibility. Most combined pull methods for the aggregate and operational release of capacity (specialist outpatient clinics 1, 2, 3, 8, 9 and 10), but one only used it on an aggregate level (4) while others only on an operational level (5, 6 7 and 12). The three specialist outpatient clinics that had the largest reduction of delay in their access and sustained a delay in access less than one week (1,2 and 3) used both aggregate and operational pull methods. The two specialist outpatient

clinics that did not sustain their results (11 and 13) did not apply pull methods on an aggregate level and only one on an operational level.

The combination of the application of pull methods on an aggregate as well as on an operational level seems most effective, but as specialist outpatient clinic 7 shows, it's also possible to optimise the delay in access to less than a week with only an operational application of pull and specialist outpatient clinic 4 shows a substantial and sustained improvement with only aggregate pull (and minimising the effects of operational push). No specialist outpatient clinics reduced their delay in access without introducing pull.

7. Discussion

7.1 Introduction

In this chapter we provided an overview of the causes of delays in access and the identified countermeasures of these causes. We described a Production Planning and Control (PPC) framework for specialist outpatient clinics with the variables to optimise the way supply and demand are matched. We used the PPC framework to create a push-pull typology of specialist outpatient clinics. With this typology we analysed the outcomes of our research on 14 specialist outpatient clinics that applied Working Without a Waiting List (WWWL). In this paragraph we use the results to reflect on the four hypotheses and answer the research question.

7.2 Reflection on the hypotheses

Hypothesis 1: The way supply and demand are matched is predominantly a push method.

Whether a push or pull method is predominant is determined either capacity released based on the actual system status (pull) or released on an exogenous trigger (push) (chapter 2.3). The analysis of the WWWL results and interventions (paragraph 6) shows that both the weekly release of the schedule with the sessions per specialist and on an operational level the daily release of appointment slots within those sessions for all specialist outpatient clinics were predominantly push based before they applied the concept of WWWL.

The actual system status does sometimes influence the release of capacity. This is ad hoc however and usually not aimed at optimising the system. In conclusion: there are pull methods present, but the push methods are predominant, since the release of most capacity is based on exogenous triggers.

Hypothesis 2: The way specialist outpatient clinics match supply and demand causes delays

To prevent delays, we identified two possible strategies for specialist outpatient clinics:

1. Increase supply.
2. Synchronise supply and demand.

Before they applied WWWL no specialist outpatient clinic offered more appointments than expected demand (strategy 1) and none synchronised supply and demand (strategy 2). The release of sessions was predominantly determined by the availability of the specialist and the release of appointment slots by the availability of fixed appointment slots. The fluctuation in the volume of supply of each type of appointment had no relation to anticipated fluctuations in demand, and it only fluctuates in response to urgent demand. These methods cause avoidable delays in access. This

supports the hypothesis that the way specialist outpatient clinics match supply and demand causes delays.

The hypothesis is further confirmed by the results of the reduction in delays by the specialist outpatient clinics (chapter 5) and the respondents of the interviews confirmed that the sustainability of the results were achieved by improving the way supply and demand are matched (chapter 6).

Hypothesis 3: The specialist outpatient clinics that reduced their delays and sustained the results introduced pull methods

As shown in Table 8.2 nine out of ten specialist outpatient clinics that reduced and sustained their delays introduced pull methods at the aggregate level as well as the operational level and the tenth specialist outpatient clinics introduced the pull method at the aggregate level and minimised the push effects on the operational level. The hypothesis is confirmed.

Hypothesis 4: Other departments that influence the matching of supply and demand of the specialist outpatient clinics are required to be included in the 'pull' system to optimise delays

None of the specialist outpatient clinics that reduced and sustained their delays and none that did not sustain their results included other departments in the pull methods. It might be required to include other departments to further optimise delays, but the results show that without including them, delays can substantially be reduced in a sustainable way. Our research does not support the hypothesis.

Research question: What is required to optimise the way demand and supply are matched in specialist outpatient clinics?

This research question has been answered by identifying the causes of delays and the methods to improve delays, by creating a PPC framework for specialist outpatient clinics, the creation of a push-pull typology and the analysis of the application of WWWL with the push-pull typology. We further answer the research question by comparing the applied methods with WWWL and the PPC framework, by discussing the applied push – pull combinations, the drivers to optimise delays in access and how the optimal delay in access can be determined.

7.3 Comparison of the applied WWWL methods and the PPC framework

The analysis showed that the specialist outpatient clinics were able to substantially reduce delays in access and sustain the results with applying only a part of the elements of the PPC framework. They did:

- Reduce endogenous variation in supply.
- Shift from push to a hybrid combination of push and pull methods on releasing sessions to increase flexibility in the volume of appointments, based on the actual system status.
- Some decided an explicit percentage of buffer capacity to be planned or to be available as optional sessions or prolonged sessions, others simply agreed to be more flexible and depend on the individual specialist and assistants to uphold this agreement.
- Shift from push to pull on releasing appointment slots by creating a flexible appointment mix.

The specialist outpatient clinics did not apply the following elements from the PPC framework:

1. Anticipate seasonal variation in the demand for new appointments and plan availability of capacity accordingly.
2. Keep track of the actual demand for follow-up appointments and plan availability accordingly.
3. Redirect demand for follow-up appointments from anticipated periods with a shortage of supply to other periods (this reduces the impact of the former two elements).
4. Create a flexible session mix per week that offers sufficient slots for each appointment queue.
5. Take the system status on supply and demand for operating sessions into account to determine the release of outpatient sessions (and vice versa).

By applying the other elements of the PPC framework a further reduction of delays in access might be possible.

7.4 Push – pull applications

The WWWL reports and interviews showed that the new methods couldn't be designed in one go. They needed to be tested and adjusted based on experiments, plus later they need to be adapted to changing circumstances. The need for this learning and evolving approach will increase when the described advanced methods are added to the WWWL interventions and need to be combined into a coherent PPC framework for each specialist outpatient clinic.

The specialist outpatient clinics that sustained their results applied four different combinations of push-pull methods. Which combination should be preferred in what situation? None created a flexible mix of sessions. From an Operations Management (OM) point of view this is a preferred method to optimise delays. It requires a substantial investment in time to develop the method for each specialist outpatient clinic however and some understanding of OM principles to learn from the application of the method and develop it further. Considering the major changes that the specialist outpatient clinics already made in their projects, it is not surprising that they did not make their session mix flexible, especially considering that several were able to reduce their delays to less than a week without it. It might however require less buffer capacity if the session mix is flexible because a fixed number of sessions with random cancellations will lead more often to a shortage of a specific type of appointment than a balanced mix (chapter 2.4). For the other specialist outpatient clinics a flexible mix might further optimise their delays.

The more often sessions are cancelled and the less flexibility in the appointment mix, the more important a flexible mix of sessions is in order to optimise delays. For the outpatient surgical clinics the flexible mix ideally should include the mix of operating sessions. Cancellations of either outpatient sessions or operating sessions should lead to a new mix of outpatient sessions and operating sessions, including the consideration of the delays for outpatient specialty appointments ass well as surgery.

For the operational appointment mix, the fewer queues and the fewer fixed slots in the sessions, the shorter the optimal delay in access can be and the less buffer capacity is required. Due to dependencies with other capacities it can be challenging to work with a completely flexible mix and some level of fixed appointments has to be accepted.

In conclusion, it seems that the pull-pull combination (number 5, see paragraph 5) provides the best optimisation for delays and requires the least buffer capacity. If the limitations that the specialist outpatient clinics described is taken as a given, the best mix is the hybrid-pull combination (6) or the hybrid-hybrid combination (9) if part of the appointment mix is partly inflexible due to dependencies on other capacities.

7.5 The drivers to optimise the way demand and supply are matched

To optimise delays in access it seems vital to create the opportunity for the involved specialists to shift their identification with delays in access from 'long delays are an indication of a good doctor' to 'short delays are an indication of a good doctor' in a way that maintains a positive cognitive and affective state, expands their perception of a competent and efficacious doctor and that enables them to connect it to their view of how to deliver patient care in a way that they sense and experience coherence and continuity. The PPC framework can easily create the opposite perception if specialists are treated as resources that need to be planned and controlled. If however the specialists perceive the PPC framework as an enabler of their vision how to deliver quality care, they may drive the decisions to optimise delays in access.

7.6 How to determine the optimal delay in access

From a logistical point of view, determining the optimal delay concerns the balance between minimising the delay and maximising the utilisation of capacity, or more specifically, minimising unused appointment slots in the sessions.. The more effective the methods to create flexibility in the mix and volume of appointments are applied and the more effective the PPC framework is applied to anticipate demand, redirect demand and the more flexibility is created in relation to the operating sessions, the shorter the delay can be without an increase in unused appointment slots. This dynamic is visualised in Figure 8.4 (see for more information chapter 2.2). The upper line represents an outpatient clinic that did not yet apply the methods and the lower line represents the same clinic after applying the methods.

A different point of view that needs to be taken into account is that specialists have expressed their concern that a short delay will lead to an increase of referrals of patients with problems that should be treated by the General practitioner or that will disappear spontaneously if patients have to wait long enough. Both concerns indicate incorrect referrals that should be addressed. This may not be easily possible however and if these concerns exist they should be monitored and taken into account to determine the optimum delay. How many appointments does it concern? At which delay in access do they occur? How much no shows are associated with problems that disappear spontaneously with longer delays (which is inefficient as well)? It should be noted that none of the specialists that achieved and sustained short delays in access raised this concern, which indicates that either they have not experienced it or they consider the benefits bigger than the downside.

The WWWL results show that specialist outpatient clinics have been able to work efficiently with an average delay in access for new and follow-up appointments for all specialists of less than two weeks and some less than one week. The PPC framework adds elements to further improve the way supply and demand are matched and a framework to relate all the elements. The delay in access may substantially be further reduced with these methods with the same utilisation, perhaps to

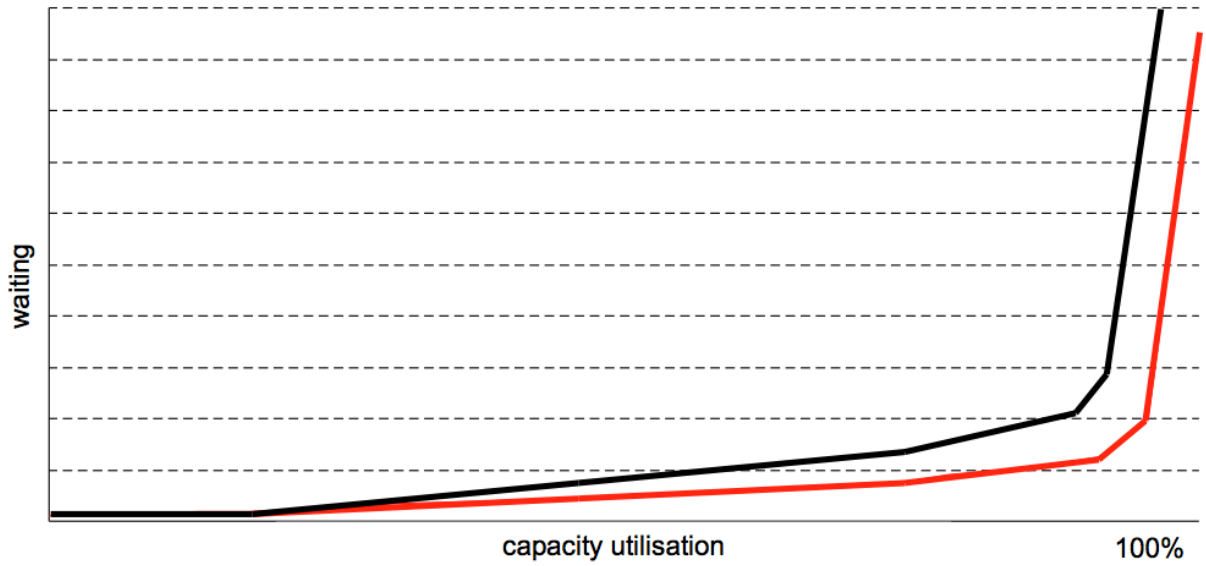


Figure 8.4 The relation between waiting and capacity utilisation with uncertainty in the arrival process and in the service process

several days or even same day access. The only way to discover the optimal access for any specialist outpatient clinic is to experiment and to boldly go beyond the optimum to discover it.

References

1. Gallucci G, Swartz W, Hackerman F. Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center. Psychiatr Serv. 2005 Mar;56(3):344-6.
2. Lacy NL, Paulman A, Reuter MD, Lovejoy B. Why we don't come: patient perceptions on no-shows. Ann Fam Med. 2004 Nov-Dec;2(6):541-5.

CHAPTER 9

General discussion

1. Introduction

Delays in access to specialist outpatient clinics are considered as a commonly prevailing problem that concerns a vital aspect of the quality of care.[1,2,3] Waiting is considered as an unavoidable phenomenon in health care due to the constraints on capacity.[4] Delays used to be dealt with by allocating additional resources, discouraging demand or forcing health providers to treat more patients with the same resources.[5] In the beginning of this century the awareness started to grow that changes in the organisation and logistics might provide new answers to waiting lists and delays.[6] In this thesis we investigate this notion with the following problem statement: 'How can delays in access to specialist outpatient clinics be optimised by improving the way supply and demand are matched?'

This chapter follows the same structure as the thesis. In part I we reflect on the theory. In part II we discuss the main research findings, including methodological considerations. In part III we discuss the main results of the synthesis. The concluding paragraph presents our conclusions on the applicability, recommendations for further research, implications for practice and general conclusions.

2. Part I - Theory

In chapter 2 a theoretical framework is presented on the general science of waiting and specifically on pull methods and the results are applied to the context of specialist outpatient clinics. The theoretical framework ends with the description of a general pull based Production Planning and Control framework. In chapter 3 a literature review is presented on the causes of delays in access to outpatient care and the effectiveness of applied countermeasures. Finally, in chapter 4 an overview is presented of the knowledge on process improvements in hospitals with the pull based 'lean'-philosophy. In this paragraph we reflect on the theory with the findings from our research and results of our synthesis.

There is almost no knowledge available on the specific causes of delays in access to specialist outpatient clinics and very little on the effectiveness of countermeasures. The available knowledge in the field of Operations Management (OM) on the science of waiting is extensive and has been sufficient for this thesis to identify the causes of delays in access to specialist outpatient clinics and sufficient to determine the required countermeasures to reduce delays in access and to create a framework to optimise delays.

The principles of the science of waiting and specifically the pull methods are applicable in specialist outpatient clinics, but the characteristics of the processes in specialist outpatient clinics do require a specific translation. Most important are the causes for variation on the volume of both demand and supply. A specific factor is the degree to which demand for follow-up appointments can be monitored and anticipated and be redirected. Another factor is that the main 'resource', the time of the specialist, is a shared resource that is contested by different departments and that is a substantial part of the time absent.

A different specific characteristic is that most specialist outpatient clinics immediately schedule demand in the 'production' schedule (scheduling an appointment). The consequence is that the operational release of capacity happens real time with demand, while the tactical release of

capacity (scheduling of sessions) has to be released ahead of actual demand. Furthermore, it has to deal with the characteristics described above. Another specific characteristic is that appointments have to be planned at a time that is convenient for the patient and during the production process, the patient can change his mind, require more time to think etcetera. This adds an inherent uncertainty to the volume and timing of 'production' that has to be dealt with.

Altogether, this creates a very specific dynamic to apply push or pull methods to plan and control 'production'. The general pull-based Production Planning and Control (PPC) framework provided the right elements to achieve this, but cannot be copied from the factory environment. It needed to be adapted to specialist outpatient clinics due to the characteristics described above.

We consider pull-methods a requirement to optimise delays in access to specialist outpatient clinics, because they enable hospitals to relate the release of capacity in both specialist outpatient clinics and other departments like the operating to be decided with consideration of the impact on each other's processes. We consider the 'lean-philosophy' a beneficial approach to develop the PPC framework into a pull-based framework because it provides elements that are required to address the problems of delays in specialist outpatient clinics and relates these elements for an integrated optimisation of delays in hospital processes. Due to the increasing complexity when it is applied across departments and resources we expect that other theories will be needed to further study and develop operations management in hospitals. We consider the domains of organisational change and development and sensemaking especially relevant.[7,8]

3. Part II – Research

3.1 Introduction

In part II of this thesis, the interventions and project results of 'Working without a waiting list (WWWL)' are evaluated. WWWL is a practical model to optimise delays in access to specialist outpatient clinics by improving the way supply and demand are matched. The sustainability of WWWL is evaluated by returning to the specialist outpatient clinics three years later with new measurements and interviews of the key actors in order to identify what they consider key determinants of their level of sustainability. Finally, the way supply and demand are matched in specialist outpatient clinics is explored and the variables are identified that determine the mix of push and pull methods to optimise delays in access. In this paragraph the main findings are discussed in relation to the research questions and hypotheses of this thesis. The structure is that we first reflect on the findings on the context of the problem (paragraph 3.2), then on the causes of the problem (paragraph 3.3), the desired state to optimise delays (paragraph 3.4) and finally the evaluation of effectiveness and sustainability of the measures of WWWL to address the causes (paragraph 3.5). Paragraph 3.6 ends with methodological considerations.

3.2 The context of the problem of delays in access

The related research question is:

How are demand and supply matched in specialist outpatient clinics?

The associated hypothesis is:

The way supply and demand are matched is predominantly a push method.

The research findings confirm these two hypotheses: both the release of sessions on an aggregate level and the release of appointments within sessions are predominantly released with push methods. The number of sessions per week or month and the mix of appointments per session are based on a fixed schedule that is repeated every week or two weeks, minus the absence of the specialist. The latter can be due to the specialist needing to perform an extra operating session in the operating theatre, but not vice versa, which is also a characteristic of a push method. As a consequence of the push methods, the variation in demand for follow-up appointments is for a substantial part caused by variation in supply.

The reconstruction of why they designed or changed the way they match supply and demand showed either non-logistical considerations or ad hoc countermeasures that might solve a logistical problem at hand, but that can increase delays in access because of interdependencies.

Contrary to the mission statements of many hospitals that state that patients are placed in the centre of everything they do, this is not evident in the choices how supply and demand are matched. We recognise from our research findings that the intention is present to place the patient in the centre, but that the actual behaviour is based on the needs of the way supply is organised. The choices to release sessions and schedule appointments are supply driven. Fluctuations in demand are mostly ignored, with the exception of urgent appointments.

From an operational management point of view, we do not consider it logical to place the patient in the centre, or in the case of this thesis, the demand for appointments. It makes more sense to emphasise the processes that match supply and demand and release capacity based on the system status. Placing the patient in the centre leads to a similar sub-optimisation as placing the care professional in the centre because the integral system impact is ignored.

We conclude that the planning methods to match supply and demand are aimed at distributing capacity, controlling the (equal division of the) generation of revenues and meeting the practical needs of specialists and assistants. The methods are not aimed at optimising delays in access.

3.3 The causes of delays in access

The next hypothesis is:

The way specialist outpatient clinics match supply and demand causes delays

The hypothesis is confirmed that the way specialist outpatient clinics match supply and demand causes delays. It's possible that in addition to this cause demand can be larger than supply. Our research showed that outpatient specialist clinics have substantial measures at their disposal to reduce the volume of (follow-up) demand, reduce the waste of unused appointments and increase the number of supplied appointments with the same capacity of specialist time. In practice, however, the specialist outpatient clinics increase efficiency to meet production agreements, not to optimise delays.

The specialist outpatient clinics do not anticipate fluctuations in demand and they do not monitor and do not react to fluctuations in demand. There is no strategic buffer capacity; the design of weekly schedules has no flexibility in the mix of sessions; sessions are frequently cancelled in no relation to volume of demand; the design of sessions has a fixed mix of types of appointments with

little flexibility. Altogether, this creates a system that has a small chance to offer the right volume of supply in relation to demand per specific type of appointment on a weekly basis.

We conclude that the way outpatient speciality clinics match supply and demand causes delays because the volume of supply fluctuates out of sync with the fluctuation in the volume of demand. There are no routines to systematically improve the way supply and demand are matched and the available theoretical knowledge on Operations Management (OM) offers the required knowledge to address these causes, but this knowledge is hardly applied in the specialist outpatient clinics.

On a different level, our research findings show that the mind-set and behaviour of specialists are a vital part of the cause of delays and equally for the potential to optimise it. The identity and self-image of specialists is connected with the existence of delays. Long delays can be a symbol for popular and good specialists instead of a specialist that neglects to optimise the matching of supply and demand.

3.4 What is required to optimise delays in access?

The related research question is:

What is required to optimise the way demand and supply are matched?

To optimise delays the specialist outpatient clinics need to anticipate predictable fluctuations in demand, react to actual fluctuations in demand and create flexibility within the mix of appointments to level out fluctuations in supply and in demand for specific types of appointments. When specialists have to divide their time among other departments as well, for example with the operating theatre, the specialist outpatient clinics need to determine the appropriate release of outpatient sessions in relation to the appropriate release of operating sessions to prevent sub-optimisation.

The aggregate release of sessions and the operations release of appointments are the two levels to apply push or pull methods. The operational release of appointments is directly triggered by the actual demand of individual patients and has as a consequence that the aggregate release of sessions cannot be triggered by demand (because it has to be available before individual appointments are planned). This is a characteristic that requires a specific translation to the methods to optimise delays in access.

We believe that a positive identification of specialists with optimising delays is a vital requirement. This requires a positive cognitive and affective state that expands their perception of a competent and efficacious doctor enabling them to connect it to their view of how to deliver patient care in a way that they sense and experience coherence and continuity.

3.5 The effectiveness and sustainability of WWWL

The research questions to determine the effectiveness of WWWL are:

1. *What changed in the way the specialist outpatient clinics match supply and demand?*
2. *What are the results in reduction of delays in access?*

The research questions to determine the effectiveness of WWWL are:

1. *Which changes are still present three years later and what new changes have taken place?*
2. *What is the delay in access three years later?*

3. *Which factors determined whether the improvements in delays in access were sustained according to the involved actors?*

The associated hypotheses are:

- *Applying the WWWL method enables specialist outpatient clinics to reduce delays in access and sustain their results.*
- *The specialist outpatient clinics that reduced and sustained their delay in access introduced pull methods.*
- *Other departments that influence the matching of supply and demand of the outpatient clinics are required to be included in the 'pull' system to optimise delays in access.*

The research findings show that the specialist outpatient clinics substantially reduced their delays with WWWL and that the results were sustained. The findings also confirm the hypothesis that all specialist outpatient clinics that reduced and sustained their delay in access introduced pull methods. They all introduced pull methods on the aggregate level of releasing sessions and almost all did on the operational level of releasing appointments within sessions. The specialist outpatient clinics that did not sustain their results partly introduced pull methods on the operational level of appointments, but none did on the aggregate level of sessions.

The specialist outpatient clinics applied different mixes of interventions, but they all improved efficiency (return rate, no shows, appointment length, unused slots and unwanted cancellations of sessions). They all increased flexibility, either in the volume of sessions, appointments per session or both. They all increased flexibility in the mix of appointments per session and they all created feedback to know when to release extra sessions or appointments or take other actions. Most specialist outpatient clinics combined the reduction of delays with a form of 'patient initiated care': enabling patients to decide themselves if and when to schedule a follow-up appointment, mostly for either chronic diseases or after surgery.

Sustainability

The respondents of the specialist outpatient clinics identified three types of generic factors that they consider crucial for their level of sustainability:

1. Increased responsiveness to better match supply and demand
    a. Making access visible
    b. Flexibility of capacity
    c. Maintaining buffer capacity
2. Clinical leadership and incentives
3. A shared belief that they can and should control access together

Contrary to our third hypothesis, the specialist outpatient clinics were able to substantially reduce and sustain their delays without integrating the release of sessions with the release of capacity for other departments, like the operating theatre. Several outpatient surgical specialties expressed their hope that it will become possible in the future to further optimise delays in both access to their specialist outpatient clinic and the operation theatre.

The specialist outpatient clinics that were not able to sustain the results did not only apply less pull methods. They also experienced less of a shared belief that they can and should control delays in access and less motivation from the involved medical specialists.

From an OM point of view the reduction of unwanted cancellations of sessions and the increase of flexibility of the volume of sessions and the mix of appointments in combination with feedback on the actual system status are the most important explanations why WWWL has proven effective. From the perspective of the actors in the specialist outpatient clinics the motivation of the medical specialists and the shared belief of the teams working in the outpatient specialist clinics are also important explanations why WWWL has proven effective.

Reflection on the WWWL model

The results of our research show that the application of WWWL has been largely effective, substantial and sustainable, however the interpretation of the results raised three questions on the effectiveness of WWWL and two questions on the sustainability:

1) Has the improvement on efficiency been a vital element of the WWWL success?

The specialist outpatient clinics improved the way they match supply and demand without adding capacity, but they all improved efficiency by reducing no shows, unused appointment slots, appointment length, return rate and unwanted cancellations of sessions.[9] WWWL has both improved the utilisation of capacity and reduced delays in access. We think that the efficiency gains have been a prerequisite to optimise delays, because they create the required buffer capacity to become flexible in the volume of supply.

2) How much opportunity for further optimisation of delays in access is possible?

In the WWWL projects the outpatient specialist clinics did not anticipate or influence fluctuations in the volume of demand. They also did not take the system status for related capacities like the operating theatre in consideration and have not been able to create flexibility to replace an operating sessions for an outpatient session. Finally, they have not created flexibility in the mix of sessions per week.

Inclusion of these elements might enable specialist outpatient clinics to further reduce delays with the same utilisation level or increase the utilisation level with the same delay (= with less buffer capacity).

3) Risk of sub-optimisation in the hospital?

The specialist outpatient clinics that applied WWWL optimised the way they match supply and demand within their department. None have included the effect on other departments as part of the matching of supply and demand. They applied pull methods within their department, but from the perspective of downstream departments, the internal demand still follows a push method. The system status of related departments is not taken into account when the specialist outpatient clinics release capacity.

If in a hospital all specialist outpatient clinics optimise delays and respond faster to increases in the volume of demand, this might lead to the increase of fluctuations of demand for the services of other departments, e.g. the radiology department, laboratories and the operating theatre. This optimisation of the outpatient specialist clinics without integration with other related departments

could result in reduction of the performance of the other departments in terms of delays and capacity usage.[10] If all specialist outpatient clinics in a hospital optimise delays, it will require an integrated approach for matching supply and demand across the related departments to prevent sub-optimisation.

4) *How much of the sustainability is the result of the system change versus highly motivated specialists and improvement teams?*

The interviews showed that the participating specialists are highly motivated to optimise delays, that their identification with delays had changed and that this had a strong impact on the teams to create a shared belief that they can control delays. This raises the question to what extent the sustainability of WWWL depends on highly motivated specialists. Will the results sustain if the specialists leave or lose their motivation? Our results show that the system changes are an important part of the explanation of sustainability, but also that a significant portion of the level of sustainability depends on one or two highly motivated specialists within each specialist outpatient clinic.

It's likely that the specialists that chose to participate in the projects were extra motivated, especially regarding the innovative context of the first WWWL collaboratives and the attention the results of their projects received probably further increased the motivation of the specialists to sustain the results. We consider the collaborative form of the projects an essential context for the medical specialists to change their identification and associated behaviour in regard to optimising delays in their practice and for the teams to create a shared belief to cooperate in new ways to optimise delays.

In the years after optimising the way demand and supply are matched it is likely that new changes and problems will emerge that will increase delays in access and need to be addressed. If the non-logistical arguments for change that influence how the involved specialists meet those challenges are not addressed the causes of delays will return. Sustainable change therefore not only requires logistical improvements, but also changes in behaviour, mind-set and perhaps identity construction on the level of individual specialists and on the social psychology of all the people working in the specialist outpatient clinics. To optimise delays in access, specialist outpatient clinics require knowledge from OM as well as a context where sensemaking processes can occur, leading to new forms of improvisation, translation and learning to change the way demand and supply are matched.[7]

The clinical leadership has been vital to sustainable change, but it might also have created vulnerability by making it at least partly dependent on the clinical leader. We believe that long-term sustainability requires that the achieved behavioural change somehow becomes more part of the daily routines of the team working in the specialist outpatient clinic without clinical leadership. There should be an improvement system to adapt the processes and associated behaviour over time in congruence with the required principles to optimise delays in access. We consider the described identity construction of the involved specialists, but also the identity construction of the team as a group vital for this level of sustainability. The surprising finding that senior management is not given any credit for the sustainability of the results adds to the vulnerability of the results since they are likely to play an important role to create an improvement system. It should be noted

that senior management did allow the improvement teams to run experiments and change their practice, which in itself may have been an important factor for change and sustainability.

5) *Does sustainable optimisation of delays in access to specialist outpatient clinics require periodic or continuous change?*

A distinction can be made between episodic, discontinuous and intermittent change and continuous, evolving and incremental change.[8] The WWWL evaluation shows that the high level of sustainability was achieved by sustaining the interventions and that they did not invent new interventions; they only added known interventions shortly after the projects. This indicates that the reduction of delays in access as well as the sustainability of the results is the effect of a periodic change that was achieved during and shortly after the WWWL project. The findings of the in depth research show however that specialist outpatient clinics continuously face small and sometimes large logistical problems that need to be solved. It also showed that decisions usually solve the ad hoc problems, but that these changes can interact and over time lead to a reduction of the effectiveness of the way supply and demand are matched, causing an increase in delays. This suggests that the optimisation of delays in access requires continuous change in response to new problems that continuously emerge over time. The respondents of the WWWL sustainability interviews however, did not mention this requirement to explain their level of sustainability.

The WWWL results and sustainability are the effect of an episodic change that addressed the major causes of delays and that enabled the specialist outpatient clinics to better deal with new problems on a continuous basis. We think that as the respondents of the in depth analysis were not aware of the potential impact of the ad hoc changes over time, the respondents of the WWWL interviews were not aware of the continuous changes they have made since WWWL and that they made better choices than before WWWL to optimise delays. Furthermore, the high level of motivation can have blocked influences that before WWWL led to sub-optimisation. We therefore speculate that either WWWL has created a level of continuous change of which the respondents themselves are not aware, or the future sustainability depends on the motivation of several individuals in each specialist outpatient clinic that were part of the WWWL project.

We conclude that the application of Advanced Access is robust because of the ability to sustain the interventions, but we are not certain that it is resilient because we cannot ascertain the ability to adapt the interventions to changing circumstances. We think that specialist outpatient clinics need a periodic change to reduce delays and need continuous change to sustain and further optimise delays. Generally speaking, most specialist outpatient clinics will need more continuous change than the studied specialist outpatient clinics, because the latter have the advantage of highly motived specialists to sustain the results. Motivated specialists are required to make the needed changes, but if they are not as highly motivated as the ones from the studied specialist outpatient clinics, we think that an improvement system is required to continuously adapt to changing circumstances.

3.6 Methodological considerations

The type of phenomenon of this research is difficult to study, requiring a mix of methods and uncontrolled settings. Several issues should be considered when interpreting our results.

The qualitative results are partially based on interviews that are inevitably biased. The results enhance our understanding of the studied phenomenon, but especially the identified factors for sustainability are likely to be incomplete. With longitudinal and intensive observational studies more factors might be identified which the respondents are unaware of or reluctant to state.

The specialist outpatient clinics that participated in the WWWL project collaboratives had problematic delays but no extreme delays. Outpatient specialty clinics with more extreme delays may deal with a situation where the volume of demand is so much larger than supply that they are not able to improve the situation enough to optimise delays in access. Furthermore, there might be specialist outpatient clinics with short delays that have a different approach that is effective and also worth to investigate.

The number of 20 investigated specialist outpatient clinics is relatively small. A larger group can give a more comprehensive representation of all Dutch specialist outpatient clinics. On the other hand, the complexity of the studied phenomenon required us to go into depth and create a rich understanding, which would not have been possible with a large group. The number is large enough to include a rich diversity within the characteristics of specialist outpatient clinics enabling us to thoroughly understand the phenomenon of delays in access and to create a comprehensive framework.

Each specialist outpatient clinic simultaneously applied several interventions and each applied a different set of interventions. The interventions interact with each other and it's likely that it's the combination of interventions leading to effects that would not be reached if they were applied separately. Our conclusions are valid for the combined interventions, but we cannot be certain what the effects are of individual interventions.

There were no data available on the changes in access for all Dutch specialist outpatient clinics to compare the WWWL results with. The WWWL results are substantial, can be explained by the interventions, this relation is confirmed by the interviews and furthermore the demand has not decreased. This all makes it unlikely that the effects can be explained by general changes for specialist outpatient clinics, but it would have been preferable to have this confirmed.

The investigated specialist outpatient clinics are Dutch. The basic logistical characteristics of specialist outpatient clinics are probably universal, but some aspects can be very different between countries. See Introduction for the Dutch characteristics.

4. Part III – Synthesis

In part III of this thesis, the results of theory and research are used to present an overview of the identified causes of delays and the identified interventions to optimise delays. The identified variables to optimise delays are used to create a Production Planning and Control (PPC) framework for specialist outpatient clinics. With the PPC framework a push-pull typology for specialist outpatient clinics is created. The typology is used to evaluate the WWWL results and ascertain which typologies were successful in reducing delays in access and sustaining the results. Finally, the considerations to determine what the optimum delay in access is are presented. In this paragraph we discuss the findings, overviews and results and reflect on the problem statement of this thesis:

How can delays in access to specialist outpatient clinics be optimised by improving the way supply and demand are matched?'

Specialist outpatient clinics need to both reduce variation and increase flexibility to deal with the remaining variation. WWWL has proven effective to optimise delays in access, but the PPC framework for specialist outpatient clinics shows a more comprehensive and integrated approach to further optimise delays in access. Five elements of the PPC framework have not been applied by the outpatient specialist clinics in the WWWL projects:

1. Anticipate seasonal variation in the demand for new appointments and plan the availability of capacity accordingly.
2. Keep track of the actual demand for follow-up appointments and plan the availability accordingly.
3. Redirect demand for follow-up appointments from anticipated periods with a shortage of supply to other periods (this reduces the impact of variation of the former two elements).
4. Create a mix of sessions per period that offers sufficient slots for each appointment, with consideration of the cancelled or added sessions.
5. Take the system status on supply and demand for operating sessions (or other important related capacities) into account to determine the release of outpatient sessions (and vice versa)

Three of these five elements are part of the WWWL model, but were not applied by the specialist outpatient clinics during the projects and not in the years after the projects. Why not? We speculate four reasons:

1. These interventions require a higher level of OM knowledge and experience than present at the specialist outpatient clinics.
2. The improvement teams underestimated the potential impact of these interventions.
3. The results were already satisfactory and the required extra effort was not considered worth the extra effort.
4. Element '5' requires the cooperation of the operating theatre and perhaps other departments and the outpatient specialist clinics estimated that this was not feasible.

Elements '2' and'4' require an even higher level of OM knowledge and experience than the other three. In conclusion we expect that the potential extra benefits of applying the complete PPC framework outweigh the extra efforts, but it requires a maturity level of applying OM methods and a maturity level of change management that at this time is not common practice in hospitals.

To apply the PPC framework specialist outpatient clinics need to learn how to choose the right cycle-mix and continuously optimise the session mix per multi-weekly period and the appointment mix per week, based on the anticipated fluctuations in demand and supply and the actual system status.

In order to optimise the total logistical system in hospitals an integrated approach is required that WWWL does not offer and that to our knowledge does not yet exist. We think that the more a system to match supply and demand is characterised by complexity and uncertainty, the more important the capability becomes to respond to actual system conditions in relation to the capability to plan, see Figure 9.1. We also think that optimisation requires predominantly pull

methods, but also to some extent push methods. This is necessary because due to some level of natural and random variation in health processes, in hospital systems some capacity needs to be released before actual demand and system status are known. We consider the lean-philosophy promising to create the required integral approach and to create a system and culture for continuous change that enables the system to continuously adapt to changing circumstances and complex new challenges that will keep emerging.

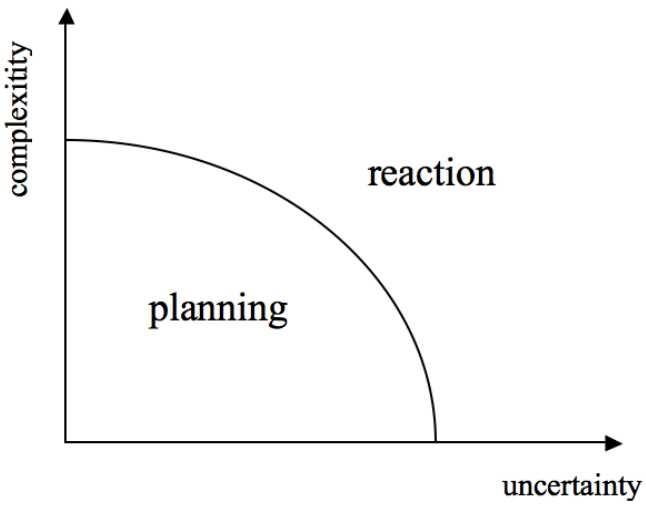


Figure 9.1 Required predominant capabilities

5. Conclusions

5.1 Introduction

In this paragraph we present our conclusions on the applicability of our findings in paragraph 5.2, implications for practice in paragraph 5.3, recommendations for further research in paragraph 5.4, and general conclusions in paragraph 5.5.

5.2 Applicability of our findings

The concept of WWWL and our Production Planning and Control (PPC) framework for specialist outpatient clinics can probably be applied by any specialist outpatient clinic in any country to optimise delays in access. The relevance of our reflections on the importance but also vulnerability of clinical leadership and the importance to identify construction might be different in other countries than the Netherlands.

We expect that the pull methods and the pull based PPC framework to optimise delays in access to specialist outpatient clinics can also be adapted and applied to optimise delays in access to other departments of hospitals. We expect that creating an integral PPC framework for the entire hospital will require additional concepts from the fields of Operations Management (OM) and organisational change that need to be integrated to deal with the increasing complexity and uncertainty. We consider an integral application of the lean-philosophy, similar to the Toyota Production System, a promising approach to achieve this.

Our typology of push-pull methods, our distinction to release capacity on an aggregate and on an operational level and the adaption of the PPC framework to accommodate this can be applied in other industries that also need to release capacity real time with individual actual demand.

5.3 Implications for practice

We believe that every specialist outpatient clinic has a moral responsibility to optimise delays in access and that the knowledge is available to do it and proven to be effective. The optimal delay in access for the studied specialist outpatient clinics is not on the same day. To determine the optimal delay in access the utilisation of capacity forms the most important counterbalance. We recommend to first apply the WWWL model if there is little knowledge of OM available and also if the delay in access has not yet been optimised before. The next step is to test the complete PPC framework to further optimise the delay in access and further increase the utilisation level.

The Dutch Treeknorm for delays in access (80% in three weeks, 100% in four weeks) should be considered as the maximum acceptable delay in access, not the goal. The optimum with WWWL is probably for most specialist outpatient clinics approximately one week.

When delays in access are present no capacity should be added before identified measures to optimise the matching of demand and supply are applied and not before the identified measures to improve efficiency in the delivery of appointments are applied. Increase of capacity should be based on the increase of demand, not on the increase of delays. A buffer capacity is required to optimise delays in access and this should be protected. The more fluctuations in the volume of demand are anticipated and the more demand for follow-up appointments can be directed and the more flexibility is created to react to fluctuations, the less buffer capacity is required. The PPC framework seems to provide the elements to minimise the required buffer capacity.

To optimise delays throughout the hospital system the PPC framework, theoretical framework and identified measures and methods provide the elements to create PPC frameworks for other departments. To continuously optimise delays integrally throughout the hospital, we recommend integrating the PPC frameworks with applying the lean-philosophy integrally, because it combines our identified elements of pull methods, integral system orientation, continuous improvement and personnel management.

To create sustainable change, the system changes need to be accompanied with a change approach that allows for specialists to show clinical leadership and multidisciplinary teams to create a shared belief. We believe that this requires the opportunity for specialists to shift their identification with delays in access in a way that maintains a positive cognitive and affective state, expands their perception of a competent and efficacious doctor enabling them to connect their view how to deliver care in a way that they experience coherence and continuity. This identification from the specialist enables the multidisciplinary team at the specialist outpatient clinic to create a shared identity that enforces the will to optimise delays in access together. In the long term, we believe that this change process requires double loop learning to create continuous improvement routines and problem solving skills that build on the OM principles and that enable teams to learn by identifying causes of problems in the way demand and supply are matched and by testing countermeasures and evaluate the effects from a total system perspective.

Hospitals need to change the stimuli for departments that affect their will to optimise delays and improve the way supply and demand are matched. Especially if an integral system approach is desired the stimuli need to reward behaviour and system changes that optimise the whole system, instead of the own department. Specifically, departments should have stimuli to optimise delays for

the other departments, as well as for their own delays in access. The current stimuli that reward the optimisation of the own department, regardless of the effect on other departments, should be removed. The approach to let the release of capacity of the operating theatre dominate the release of capacity of the specialist outpatient clinics should be replaced with methods to release the capacity of both departments in a way that optimises the total system. It should be possible to exchange operating sessions for outpatient session as much as vice versa.

The alertness of specialists to recognise and act on small signals from patients provides a capability that management should foster and support to optimise the way demand and supply are matched. The role of management is not so much one of personal advocacy, but rather the establishment of supportive structures and processes for the teams to improve and optimise delays in access themselves. A vital element is to provide feedback on the system status that supports decisions to release capacity, prevent the release of capacity or prevent the cancellation of capacity.

5.4 Recommendations for further research

The effectiveness of the WWWL model has been established by our research, but the effectiveness of the PPC framework for specialist outpatient clinics to further optimise delays in access needs to be tested and evaluated in practice.

To adapt and apply the PPC framework for other departments requires research on the causes of delays of those departments and the characteristics of the processes to match supply and demand. To create an integral framework to optimise delays in entire hospital systems additional research is needed to study the relations between the departments and resources of the departments. Analysing data on the variation in demand for new and follow-up appointments and the variation in supply for the specialist outpatient clinics and the related departments will be required to understand how the integral system dynamics can be optimised.

In our literature review on delays in access to specialist outpatient clinics, we concluded that there is almost no research available on the causes, mostly theoretical recommendations for countermeasures with little or no relation to practice There is quite some research on the effectiveness of Advanced Access as a countermeasure (of the unidentified causes) and no research on the sustainability on the effects of countermeasures. Our research included a study of the causes, application of the science of waiting on the causes, evaluation of the application of countermeasures in practice, sustainability of the measures and a synthesis to conclude what has proven effective and what is theoretically possible to further optimise. We recommend that future research on delays or similar types of problems also include all these elements.

We conclude that there is a large body of theoretical knowledge available from the field of OM to improve queuing systems, but there is almost no research available on the translation and application of this knowledge to the practice of specialist outpatient clinics. In line with the conclusions from Singhal and Singhal we recommend that future research in the field of OM will achieve more internal and external validity to facilitate decisions in practice.[11] We further agree with their conclusions that this requires the inclusion of different research paradigms and methods. This can hopefully address the conclusions of researchers in the field of OM that the results of their research on how to optimise delays in access in outpatient clinics is hardly applied in practice.[12-13]

We recommend that application of the knowledge from OM to improve hospital operations needs to be combined with the science of organisational change to create a sustainable impact. Based on the results from our research on the vital role of clinical leadership and the relation with the identity construction of specialists and the teams, we recommend the field of organisational psychology, specifically sensemaking to be combined with research on OM problems to identify the required conditions to change and sustain behaviour to optimise operations.

We recommend research on our new hypothesis that the more a system is characterised by complexity and uncertainty, the more the system requires the capability to react to actual conditions instead of the capability to plan (Figure 9.1).

5.5 Closing remarks

The way supply and demand are matched in specialist outpatient clinics causes delays in access. Working without a waiting list (WWWL) provides an effective and sustainable method to address the causes and optimise delays. Our Production Planning and Control framework for specialist outpatient clinics provides additional elements that might enable specialist outpatient clinics to further optimise delays, in relation to delays in access of other departments, like the operating theatre. WWWL as well as the PPC framework predominantly use pull methods to optimise delays in access and they are expected to be applicable to other departments as well. The lean-philosophy is promising to create an integral approach to optimise delays throughout the hospital system.

Sustainable change can be achieved with a learning environment for teams that enables teams to increase their problem solving skills, study their own practices and how they relate to other departments and by experimenting with changes. The identity construction of specialists and teams seems an important aspect of the change process. To create effective and sustainable change, the question is not whether the specialist, other disciplines or management should be in the lead, but how to create synergy between specialists, other disciplines and management to optimise delays in access as an episodic change followed by continuous change to continuously learn and adapt to new challenges that will keep arising.

References

1. Institute of Medicine. Crossing the Quality Chasm. A new health System for the 21st Century. Washington: National Academy Press, 2001.
2. Department of health (VWS). Waiting times in hospitals (in Dutch). May 2002. http://www.minvws.nl/rapporten/cz/wachttijden_in_ziekenhuizen.asp (accessed 31 Jan 2010).
3. National Institute for Public Health and the Environment (RIVM) How long is being waited on hospital care? (in Dutch). http://www.rivm.nl/vtv/object document/o7864n34418.html (accessed 31 January 2010).
4. Vissers JMH, Van der Bij JDH and Husters RJ. Towards Decision Support for Waiting Lists: an Operations Management View. Health care Management Science 4; 2001, 133-142
5. Hanning, M. Maximum waiting-time guarantee - an attempt to reduce waiting lists in Sweden. Health Policy 36 (1996), 17-35.
6. Smith J. Redesigning health care. BMJ 2001;322;1257-8

7. Weick KE. Sensemaking in organizations. Thousand Oaks: Sage Publications; 1995. xii, 231p.
8. Weick KE, Quinn RE. Organizational change and development. Annu Rev Psychol. 1999;50:361-86.
9. Van der Voort MR, van Merode F, Berden B. Making sense of delays in outpatient specialty care: A system perspective. Health Policy. 2010 Sep;97(1):44-52.
10. Rouppe van der Voort MBV, Van Merode GG, Veraart HGN. Sustainable process improvement with the 'lean'-philosophy ('Duurzame procesverbetering met de 'lean'-filosofie'). Ned Tijdschr Geneeskd. 2013;157:A5541.
11. Singhal K, Singhal J. Imperatives of the science of operations and supply-chain management. J Oper Manag. 2012;30(3):237-44.
12. Gupta D, Denton B. Appointment scheduling in health care: Challenges and opportunities. Iie Trans. 2008;40(9):800-19.
13. Cayirli T, Veral E. Outpatient Scheduling in Health Care: a Review of Literature. Production and Operations Management. 2003;12(4):519-49.

Summary

Samenvatting

Curriculum Vitae

List of publications

Summary

Introduction

This thesis explores the causes of delays in access to specialist outpatient clinics and tests the effectiveness of interventions to optimise delays in access with changes in organisation and logistics. On a daily basis approximately 50,000 people make an appointment at an outpatient speciality clinic in one of the 88 Dutch hospitals for which they will often need to wait 28 days or more. On any given day more than 1.5 million people of the population of 16 million in the Netherlands are waiting for an appointment. The problem of delays, the available theory on waiting, the established effectiveness of applying pull methods in industry to optimise delays and simultaneously improve efficiency, the lack of evidence on the causes of delays in access to specialist outpatient clinics and of approaches to optimise and the success stories on a number of 'Working Without a Waiting List (WWWL)' projects led us to the following problem statement:

> How can delays in access to specialist outpatient clinics be optimised by improving the way supply and demand are matched?

With this thesis we want to design a method to improve the way demand and supply is matched and we want to determine whether and why WWWL is an effective method to optimise delays in access to specialist outpatient clinics.

Theory

Chapter 2 describes the science of waiting from the field of Operations Management and presents a theoretical framework to investigate how delays in access can be optimised. When there is uncertainty in the volumes of supply and demand the delay often rises exponentially when capacity utilisation reaches 100%. The more uncertainty, the earlier the exponential rise of the curve starts. Relatively little buffer capacity is required to keep waiting low, but a lot of buffer capacity is required to reduce waiting further. Since buffer capacity is expensive the challenge is to find solutions that keep waiting to a minimum while requiring minimal buffer capacity. Another challenge is to stay well enough before the steep incline of the line, to avoid an instable situation where waiting can suddenly rise fast. A third challenge is to simultaneously optimise various capacities that are related to each other, for example the operating theatre and the specialist outpatient clinics.

These challenges require the synchronisation of fluctuations in the volume of demand and supply. One strategy is to increase the ability to plan; another strategy is to increase flexibility. The more flexibility to utilise capacity where demand is, the less buffer capacity is required with the same waiting time. There are three types of flexibility to match supply and demand:

- Mix flexibility: Ability to serve different mixes of appointments in one session and different mixes of sessions within a week.
- Changeover flexibility: ability to deal with changes in the mix of types of appointments over time, for example changes in the ratio between new and follow-up appointments.

- Volume flexibility: ability to easily change the volume of supplied appointments in any given week or month.

The described challenges can be addressed on a 'Strategic', 'Tactical' and 'Control' level, representing long to short term planning issues. Generally speaking, on the strategic level, forecasts are used to predict demand and plan capacity and personnel. On the tactical level, the aggregate plan is transformed to a plan of action that will help prepare for upcoming production. Total volume is determined, but the day-to-day schedule with the exact mix and volume per shift is determined later on the control level. On the control level the actual production progress is monitored against the schedule and used to determine the sequencing and scheduling for the work schedule on the tactical level. On all levels and between the levels feedback information is used to base corrective actions on the difference between the actual and desired performance. Feedforward information can be used to predict future situations.

In specialist outpatient clinics the release of capacity happens consequently on two levels:

1. Aggregate level: the scheduling of sessions based on the availability of the specialist, usually several months ahead.
2. Operational level: the scheduling of appointments with individual patients in those sessions, usually several weeks ahead.

In industry Just-In-Time (JIT) manufacturing is considered as an important collection of techniques to minimise the required capacity, increase the responsiveness of a system and relate the release of capacities to each other. In a specialist outpatient clinic the ultimate application of JIT would be to be able to offer every patient an appointment on the same day. A central principle of JIT is to switch from a 'push' operating system to a 'pull' operating system. For specialist outpatient clinics push and pull can be defined as:

> A push system schedules the availability of appointments based on exogenous factors, while a pull system authorises the release of appointments based on system status.

A specialist outpatient clinic with a push system uses exogenous factors (= 'originated outside the system') like predicted demand or information about the availability of the specialist to determine the number of sessions. This method has no relation to actual system status and no relation with actual demand, which results in longer delays in access or unused capacity if actual demand is different than the number of available appointments. A clinic with a pull system lets the actual system status determine the schedule of sessions and mix of appointments within the sessions. When the specialist outpatient clinic acts as an independent system, the system status for pull can be restricted to, for example, the delay in access for new and follow-up appointments. If the matching of supply and demand however also depends on other departments like the operating theatre, they should be included in the system to optimise.

To minimise deviations in actual production from the plan 'cycle mix' planning can be beneficial. The cycle mix is the required mix of types of appointments and types of sessions to deliver those appointments. Changes in the volume of demand for specific types of appointments or the cancelling of sessions can lead to recalculation of the cycle mix for the next period.

In **chapter 3** we complement the theoretical framework with a literature review on what is known about delays in access to specialist outpatient clinics. There is little known about the extent of the problem of delays in access to (specialist) outpatient clinics, but there are many indications that at least in the US, the UK, Canada and the Netherlands a large number of specialty practices deal with longer delays in access than generally desired from the point of view of the patient. There is little scientific insight in the way (specialist) outpatient clinics match supply and demand and there is only some scientific insight in the possible causes of delays. The most important seems variability in both demand and supply and the inability to deal with this. The most important cause for that inability may be the common practice that clinical preferences rather than the needs of patients determine the availability of supply on a daily basis.

Knowledge on possible interventions to optimise delays in access seems divided in two separate worlds. On the one hand there is a large body of knowledge available from the discipline of Operations Management (OM) to improve queuing systems like outpatient clinics. The application of this knowledge has mostly been directed at reducing waiting in the office and reducing no-shows, not at reducing delays in access. There is almost no scientific knowledge on the application of this body of knowledge in real life.

On the other hand there is a large body of knowledge on applying the concept of Advanced Access that is aimed at 'redesigning the clinical office', mostly in general practices in the US and the UK, but also some US specialty practices. The evidence shows substantial improvements, but also shows considerable issues. The biggest concerns relate to a fear of loss of autonomy under physicians and a loss of quality of care due to less continuity in the care between the patient and the physician. This is remarkable, since the concept of Advanced Access recommends maintaining continuity. There is some, but scarce evidence of sustainability of the results, mostly one or two years after implementation and there is also some, but equally scarce evidence of decay.

In **chapter 4** we further complement the theoretical framework with a literature review on how processes are improved by applying the 'lean'-philosophy, a comprehensive concept that is relevant because the concepts of Advanced Access and Working Without a Waiting List (WWWL) that are part of this research are partly based on the general lean-philosophy and more specifically the pull method.

Process improvement occurs more and more frequently, especially with the help of the 'lean'-philosophy. This management philosophy improves quality by continuously removing 'waste'. As a consequence of the interdependence of processes local improvements can cause negative effects elsewhere. An integral system approach is required to prevent this. Several hospitals claim that they are able to achieve this with Lean. Research on process improvement by applying the 'lean'-philosophy reports many positive outcomes defined as increased safety, quality and efficiency. Methodological shortcomings and lack of rigorous evaluations make it impossible however to ascertain the impact. Clear is that the investigated applications are fragmented with an overly focus on the instrumental aspects of the philosophy, a lack of integration in the total system and a lack of attention for the human dimensions. Process improvement is needed to improve both quality and efficiency of health care. It requires that hospitals develop integral systems that combine methods

for process design with continual improvement of processes and associated forms of people management. Vital is that physicians take the lead to manage and improve processes integrally.

Research

In **chapter 5** we assess whether delays to specialist outpatient clinics can be solved by improving the way supply and demand are matched, without adding capacity. We used a systematic review of the interventions applied by eighteen specialist outpatient clinics using the model of Working Without a Waiting List (WWWL), and a statistical analysis of the effects of the interventions on their delays.

The specialist outpatient clinics applied different combinations of interventions aimed at improving the way they match supply and demand, improving the efficiency of the way supply is organised and at reducing unnecessary demand. Fourteen clinics show statistically significant improvements. Two probably significantly improved and two clinics did not. Their access reduced on average 55%, from 47 to 21 days.

It seems that delays in outpatient specialty care can be solved to a large extend by improving the way supply and demand are matched. Policy makers should analyse whether delays are caused by capacity problems or matching problems. For the latter, it appears more effective to invest in the ability to react with the use of pull methods then the ability to plan. Policy makers should create incentives for clinics to keep access short and remove incentives that stimulate delays.

Chapter 6 investigates the sustainability of the results of the WWWL application. Our aim was to determine whether initial improvements are sustained and to identify the factors that influence sustainability according to the involved actors. We performed qualitative case studies in fourteen specialist outpatient clinics. We compared access measurements at the start, finish and three years after the project. Sustained and new interventions were analysed. Interviews with 52 practitioners were analysed with the constant comparative method to identify general factors that influence sustainability.

Eleven out of fourteen clinics were able to sustain or further improve their reduced delays; two did not and for one it is uncertain. The specialist outpatient clinics maintained the majority of the interventions and all introduced new interventions. Three generic factors emerged that influenced their ability to sustain the results: increased responsiveness to better match supply and demand; clinical leadership and incentives; a shared belief that they can and should control access together.

Reduction of delays in access can be sustained if the way of thinking and the planning system becomes demand driven and flexible, and care providers experience benefits. Unlike previous studies, senior management support and formal training were not relevant though clinical leadership and informal socialisation was. Making multidisciplinary teams responsible for the improvement process appears to be vital.

Chapter 7 presents the findings of an in-depth field research that identifies the characteristics how specialist outpatient clinics match supply and demand. We used field observations, interviews,

documents and data from a dermatology outpatient clinic with low dependency on other departments to match supply and demand and an orthopaedic outpatient clinic that has a high dependency on the operating theatre.

It is likely that the way both specialist outpatient clinics match supply and demand causes unnecessary delays in access. This concerns mainly the lack of anticipating fluctuations in demand, limited flexibility to increase volume in response to actual shortage of supply and the lack of flexibility to release the cycle mix of sessions depending on actual conditions. One specialist outpatient clinic furthermore lacks flexibility in the appointment mix. Furthermore, the lack of flexibility to exchange outpatient sessions and operating sessions in the operating theatre limits their ability to optimise delays. Both clinics use dominantly push methods, though some pull methods to let the actual system status trigger actions are also applied.

To optimise delays in access it seems vital for the involved specialists to shift their identification with delays in access from 'long delays are an indication of a good doctor' to 'short delays are an indication of a good doctor'.

Synthesis and discussion

Chapter 8 creates a synthesis by combining the theoretical framework and the research results into a new framework to optimise delays in access to specialist outpatient clinics. When demand is larger than supply it first needs to be balanced. We identified three forms of wasted capacity that can be reduced by improving the way supply and demand are matched: unwanted cancellations of sessions, no shows and unused appointment slots. When delays in access exist while there is enough supply to meet demand, three methods can be applied to improve the synchronisation in the volume of demand and supply: increase the ability to notice fluctuations in the volume of demand, increase flexibility in the volume of appointments and increase flexibility in the mix of appointments. We used the theoretical framework and the results of our research to create a pull based Production Planning and Control (PPC) framework for specialist outpatient clinics to realise this. The PPC framework uses the system status to trigger release sessions on an aggregate level and appointments on an operational level.

The specialist outpatient clinics that reduced and sustained their delays introduced pull methods at the aggregate level as well as the operational level. None included other departments in the pull methods. It might be required to further optimise delays but, contrary to our expectations, delays can substantially be reduced in a sustainable way without including other departments.

The analysis showed that the specialist outpatient clinics were able to substantially reduce delays in access and sustain the results with applying the PPC framework partially. They did:

- Reduce endogenous variation in supply.
- Shift from push to a hybrid combination of push and pull methods on releasing sessions to increase flexibility in the volume of appointments, based on the actual system status.
- Create explicit or implicit buffer capacity and develop methods to use this efficiently.
- Shift from push to pull on releasing appointment slots by creating a flexible appointment mix.

The specialist outpatient clinics did not apply the following elements from the PPC framework:

- Anticipate seasonal variation in the demand for new appointments.
- Keep track of actual demand for future follow-up appointments.
- Redirect demand for follow-up appointments.
- Create a flexible session mix per cycle.
- Take the system status on supply and demand for operating sessions or other departments into account to determine the release of outpatient sessions (and vice versa).

To optimise delays in access it seems vital that the application of the PPC framework supports the opportunity for the involved specialists to shift their identification with delays in access in a way that maintains a positive cognitive and affective state, expands their perception of a competent and efficacious doctor and that enables them to connect it to their view of how to deliver patient care in a way that they sense and experience coherence and continuity. The PPC framework can easily create the opposite perception if specialists are treated as resources that need to be planned and controlled. If however the specialists perceive the PPC framework as an enabler of their vision how to deliver quality care, they may drive the decisions to optimise delays in access.

In **chapter 9** we discuss our main findings. We conclude that the way outpatient speciality clinics match supply and demand causes delays because the methods they use are aimed at other goals. The available theoretical knowledge on Operations Management (OM) offers the required knowledge to address the causes, but is hardly applied. Furthermore our research findings show that the mind-set, identity and behaviour of specialists are a vital part of the causes of delays.

The research findings show that the specialist outpatient clinics substantially reduced their delays with WWWL and that the results were sustained. They introduced pull methods on the aggregate level of releasing sessions and the operational level of releasing appointments. From an OM point of view the reduction of unwanted cancellations of sessions and the increase of flexibility of the volume of sessions and the mix of appointments in combination with feedback on the actual system status are the most important explanations why WWWL has proven effective. We think that the efficiency gains have been a prerequisite to optimise delays, because they create the required buffer capacity to become flexible in the volume of supply. From the perspective of the actors in the specialist outpatient clinics the motivation of the medical specialists and the shared belief of the teams working in the outpatient specialist clinics have also been important.

In chapter 8 we described which elements of the PPC framework the WWWL projects did not address. Inclusion of these elements might enable specialist outpatient clinics to further reduce delays with the same utilisation level or increase the utilisation level with the same delay (= with less buffer capacity).

The clinical leadership has been vital to sustainable change, but it might also have created vulnerability by making it at least partly dependent on the clinical leader. We believe that long-term sustainability requires that the achieved behavioural change somehow becomes more part of the daily routines of the team working in the specialist outpatient clinic without clinical leadership.

We conclude that the application of Advanced Access is robust because of the ability to sustain the interventions, but we are not certain that it is resilient because we cannot ascertain the ability to adapt the interventions to changing circumstances. We think that specialist outpatient clinics need a periodic change to reduce delays and need continuous change to sustain and further optimise delays.

Main methodological shortcomings: the results enhance our understanding of the studied phenomenon, but especially the identified factors for sustainability are likely to be incomplete. The number of eighteen investigated specialist outpatient clinics is relatively small. Our conclusions are valid for the combined interventions, but we cannot be certain what the effects are of individual interventions.

The pull-based Production Planning and Control (PPC) framework provides the elements to optimise delays in access to specialist outpatient clinics. To optimise delays specialist outpatient clinics need to anticipate predictable fluctuations in demand, react to actual fluctuations in demand and create flexibility within the mix of appointments to level out fluctuations in supply and in demand for specific types of appointments on both the aggregate release of sessions and the operations release of appointments. To apply the PPC framework specialist outpatient clinics need to learn how to choose the right cycle-mix and continuously optimise the session mix per multi-weekly period and the appointment mix per week, based on the anticipated fluctuations in demand and supply and the actual system status. In order to optimise the total logistical system in hospitals an integrated approach is required that WWWL does not offer and that to our knowledge does not yet exist. The WWWL clinics optimised their delays alone in their hospital. If all outpatient specialist clinics in a hospital reduce delays in access without integration with other related departments it could result in a decrease of performance of the other departments in terms of delays or capacity usage.

We think that the more a system to match supply and demand is characterised by complexity and uncertainty, the more important the capability becomes to respond to actual system conditions. We also think that optimisation requires predominantly pull methods, but also to some extent push methods because in hospital systems some capacity needs to be released before actual demand and system status are known. We consider the lean-philosophy promising to create the required integral approach based on pull methods and to create a system and culture for continuous change that enables the system to continuously adapt to changing circumstances.

To determine the optimal delay in access the utilisation of capacity forms the most important counterbalance. We recommend to first apply the WWWL model if there is little knowledge of OM available and also if the delay in access has not yet been optimised before. The next step is to test the complete PPC framework to further optimise the delay in access and further increase the utilisation level. The Dutch Treeknorm for delays in access (80% in three weeks, 100% in four weeks) should be considered as the maximum acceptable delay in access, not the goal. The optimum with WWWL is probably for most specialist outpatient clinics approximately one week.

To create sustainable change, an approach is required that allows for specialists to show clinical leadership and multidisciplinary teams to create a shared belief. Departments should have stimuli

to optimise delays for the other departments, as well as for their own delays in access. The alertness of specialists to recognise and act on small signals from patients provides a capability that management should foster and support to optimise the way demand and supply are matched. The role of management is not so much one of personal advocacy, but rather the establishment of supportive structures and processes for the teams to improve and optimise delays in access themselves. A vital element is to provide feedback on the system status that supports decisions to release capacity, prevent the release of capacity or prevent the cancellation of capacity.

We recommend that application of the knowledge from OM to improve hospital operations needs to be combined with the science of organisational change to create a sustainable impact. Based on the results from our research on the vital role of clinical leadership and the relation with the identity construction of specialists and the teams, we recommend the field of organisational psychology and sensemaking to be combined with research on OM problems to identify the required conditions to change and sustain behaviour to optimise operations. To create effective and sustainable change, the question is not whether the specialist, other disciplines or management should be in the lead, but how to create synergy between specialists, other disciplines and management to optimise delays in access as an episodic change followed by continuous change to continuously learn and adapt to new challenges that will keep arising.

Samenvatting

Inleiding

Dit proefschrift onderzoekt de oorzaken van vertragingen in de toegangstijd tot poliklinieken en test de effectiviteit van interventies om de toegangstijd te optimaliseren door veranderingen in de organisatie en logistiek. Dagelijks maken ongeveer 50.000 mensen een afspraak op een polikliniek in een van de 88 Nederlandse ziekenhuis waar ze vervolgens vaak 28 of meer dagen op moeten wachten. Op een willekeurige dag wachten meer dan 1.5 miljoen mensen op een poliklinische afspraak op een populatie van 16 miljoen Nederlanders. Het probleem van toegangstijden, de beschikbare theorie over wachten, de vastgestelde effectiviteit van 'pull-' methoden in de industrie om wachten te optimaliseren en tegelijkertijd efficiëntie te vergroten, het gebrek aan wetenschappelijke inzichten in de oorzaken van toegangstijden tot poliklinieken en in de effectiviteit van benaderingen om het te optimaliseren, plus de succesverhalen over een aantal 'Werken zonder wachtlijst (WZW)' projecten bracht ons tot de volgende onderzoeksvraag:

> Hoe kan de toegangstijd tot poliklinieken geoptimaliseerd worden door de wijze waarop vraag en aanbod afgestemd wordt te verbeteren?

Met dit proefschrift willen we een methode ontwerpen waarmee de toegangstijd geoptimaliseerd kan worden en we willen vaststellen of WZW een effectieve methode is om toegangstijden tot poliklinieken te verkorten.

Theorie

Hoofdstuk 2 beschrijft de wetenschap van wachten vanuit het vakgebied Operations Management en presenteert een theoretisch raamwerk hoe vertragingen in toegangstijd geoptimaliseerd kunnen worden. Als er onzekerheid is in de volume van vraag en aanbod stijgt de vertraging vaak exponentieel zodra de capaciteitsbenutting 100% benadert. Hoe meer onzekerheid, hoe eerder de exponentiële curve snel stijgt. Relatief weinig buffercapaciteit is nodig om wachten kort te houden, maar veel buffercapaciteit is nodig om het nog verder terug te brengen. Aangezien buffercapaciteit duur is, is het de uitdaging om oplossingen te vinden die wachten minimaliseren met zo min mogelijk buffercapaciteit. Een andere uitdaging is om ruim voor de steile curve te blijven, om een instabiele situatie te voorkomen waarbij de toegangstijd plotseling snel kan stijgen. Een derde uitdaging is om deze uitdagingen te optimaliseren in samenhang met het optimaliseren van andere capaciteiten die met elkaar gerelateerd zijn, zoals tussen de operatiekamers en poliklinieken.

Deze uitdagingen vereisen de synchronisatie van fluctuaties in de volume van vraag en aanbod. Een strategie is om het vermogen om vooruit te plannen te vergroten; een andere strategie is om de flexibiliteit te vergroten. Hoe meer flexibiliteit om capaciteit in de zetten waar de vraag zich voordoet, hoe minder buffercapaciteit nodig is met dezelfde wachttijd. Er zijn drie soorten flexibiliteit om vraag en aanbod af te stemmen:

- Mix flexibiliteit: mate waarin verschillende combinaties van afspraken in een spreekuur mogelijk zijn en de mate waarin verschillende combinaties van spreekuren in een week mogelijk zijn
- Omschakel flexibiliteit: mate waarin veranderingen in de mix van type afspraken over tijd opgevangen kunnen worden, bijvoorbeeld veranderingen in de ratio tussen nieuwe en controle afspraken.
- Volume flexibiliteit: de mate waarin de volume van aangeboden afspraken in een week of maand veranderd kan worden.

De beschreven uitdagingen kunnen geadresseerd worden op een 'Strategisch', 'Tactisch' en 'Operationeel' niveau, om lange tot korte termijn planningsvraagstukken te adresseren. In algemene zin wordt op het strategisch niveau voorspellingen gebruikt om vraag te voorspellen en capaciteit en personeel te plannen. Op het tactisch niveau wordt het geaggregeerde plan omgezet in een actieplan die helpt om de komende productie voor te bereiden. Totale volume wordt vastgesteld, maar de dagelijkse planning met exacte mix en volume per dienst wordt pas later op het operationele niveau vastgesteld. Op het operationele niveau wordt de daadwerkelijke productie gemonitord in vergelijking met de planning en worden bevindingen gebruikt om de volgorde en planning van het plan op het tactisch niveau te bepalen. Op alle niveaus wordt 'feedback' informatie gebruikt om acties bij te stellen op basis van het verschil tussen werkelijke en wenselijke prestaties. 'Feedforward' informatie kan gebruikt worden om toekomstige situaties te anticiperen.

Poliklinieken geven capaciteit volgtijdelijk op twee niveaus vrij:

1. Geaggregeerd niveau: het plannen van spreekuren op basis van de aanwezigheid van de specialist, meestal enkele maanden vooruit.
2. Operationeel niveau: het plannen van afspraken in de spreekuren met individuele patiënten, meestal enkele weken vooruit.

In de industrie wordt 'Just-In-Time(JIT) manufacturing' beschouwd als een belangrijke verzameling technieken om de benodigde capaciteit te minimaliseren, responsiviteit van een systeem te vergroten en om het vrijgeven van capaciteiten aan elkaar te koppelen. In een polikliniek zou een ultieme toepassing van JIT het mogelijk moeten maken om elke patiënt dezelfde dag een afspraak te kunnen bieden. Een centraal principe van JIT is om van 'push' naar 'pull' om te schakelen als 'productiesysteem'. Voor poliklinieken kan push en pull gedefinieerd worden als:

> Een push systeem plant de beschikbaarheid van afspraken op basis van exogene factoren, terwijl een pull systeem afspraken vrijgeeft op basis van de status van het systeem.

Een polikliniek met een push systeem gebruikt exogene factoren (= 'vinden hun oorsprong buiten het systeem') zoals voorspelde vraag of de beschikbaarheid van een specialist om het aantal spreekuren vast te stellen Deze benadering heeft geen relatie met de status van het systeem en geen relatie met de werkelijke vraag, wat resulteert in langere toegangstijden of onbenutte capaciteit zodra de werkelijke vraag afwijkt van het beschikbare aantal afspraken. Een polikliniek met een pull systeem laat de status van het systeem de planning van spreekuren en de mix van afspraken in de spreekuren bepalen. Als een polikliniek als een onafhankelijk systeem werkt kan de gebruikte systeemstatus beperkt worden tot, bijvoorbeeld, de toegangstijd voor nieuwe en controle

afspraken. Als het afstemmen van vraag en aanbod echter afhankelijk is van andere afdelingen zoals de operatiekamers, dan moet de status van die capaciteiten geïncludeerd worden in het te optimaliseren systeem.

Om de werkelijke productie minimaal te laten afwijkingen van het plan kan 'cycle mix' planning behulpzaam zijn. De cycle mix is de benodigde mix van type afspraken en type spreekuren. Veranderingen in de volume van de vraag naar specifieke type afspraken of het annuleren van spreekuren leidt tot een herberekening van de cycle mix voor de komende periode.

In **hoofdstuk 3** vullen we het theoretisch raamwerk aan met een literatuuronderzoek naar de beschikbare kennis over toegangstijden tot poliklinieken. Er is weinig wetenschappelijk bekend over de omvang van het probleem van toegangstijden tot poliklinieken, maar er zijn veel aanwijzingen dat in ieder geval in de VS, Engeland, Canada en Nederland de toegangstijd tot een groot aantal poliklinieken langer is dan wenselijk vanuit het perspectief van de patiënt. Er is weinig wetenschappelijk inzicht in de wijze waarop poliklinieken vraag en aanbod afstemmen en er is beperkt inzicht in de oorzaken van toegangstijden. De belangrijkste oorzaak lijkt variabiliteit in de volume van zowel vraag als aanbod en het onvermogen om daarmee om te gaan. De belangrijkste oorzaak voor dat onvermogen is mogelijk dat niet de behoeftes van patiënten, maar de voorkeuren van de poliklinieken de beschikbaarheid van het aanbod bepalen.

Kennis over mogelijke interventies om toegangstijden te optimaliseren lijken verdeeld over twee gescheiden werelden. Enerzijds is er een grote hoeveelheid kennis beschikbaar vanuit het vakgebied Operations Management (OM) om 'wachtrijsystemen' te optimaliseren. De toepassing van deze kennis is voornamelijk gericht op het reduceren van wachttijd in de wachtkamer en het verminderen van 'no-shows', niet het reduceren van de toegangstijd. Er is vrijwel geen wetenschappelijke kennis over de toepassing van deze kennis in de praktijk.

Anderzijds is er een grote hoeveelheid wetenschappelijke kennis beschikbaar over de toepassing van het concept 'Advanced Access' om poliklinieken te herontwerpen, vooral in de eerstelijns zorg in de VS en Engeland, maar in de VS ook enkele poliklinieken. Onderzoeken tonen substantiële verbeteringen, maar laten ook wezenlijke vraagstukken zien. De belangrijkste zijn een angst om autonomie te verliezen bij artsen en kwaliteitsverlies door minder continuïteit van zorg. Het laatste is opvallend, aangezien Advanced Access juist continuïteit als uitgangspunt neemt. Er is enige, maar schaars, bewijs van duurzaamheid van de resultaten, meestal een of twee jaar na implementatie en er is enige, maar even schaars, bewijs van terugval van resultaten.

In **hoofdstuk 4** vullen we het theoretisch raamwerk verder aan met een literatuuronderzoek naar hoe processen verbeterd kunnen worden met de 'lean'-filosofie, een veelomvattende concept die relevant is omdat de concept Advanced Access en Werken Zonder Wachtlijst (WZW) die onderdeel zijn van dit onderzoek gedeeltelijk gebaseerd zijn op de algemene lean-filosofie en meer specifiek gebruik maken van de pull methode.

Procesverbeteringen vinden steeds meer plaats in de zorg, met name met behulp van de 'lean'-filosofie. Deze management filosofie verbetert kwaliteit door continu 'verspilling' te verwijderen. Door de onderlinge samenhang van processen kunnen lokale verbetering negatieve effecten elders veroorzaken. Een integrale systeembenadering is nodig om dit te voorkomen. Enkele ziekenhuizen

claimen dat hen dit lukt met behulp van lean. Onderzoek naar procesverbetering met behulp van de 'lean'-filosofie toont veel positieve effecten op het vlak van verbeterde veiligheid, kwaliteit en efficiëntie. Methodologische tekortkomingen en een gebrek aan rigoureuze evaluaties maken het echter onmogelijk om de impact vast te stellen. Duidelijk is dat de onderzochte toepassingen gefragmenteerd zijn met een overmatige focus op de instrumentele toepassing van de filosofie, een gebrek aan integratie in het totale systeem en een gebrek aan aandacht voor de menselijke dimensies. Procesverbetering is nodig om zowel kwaliteit als efficiëntie in de zorg te verbeteren. Het vergt dat ziekenhuizen integrale systeem ontwikkelen die methoden om processen te ontwerpen combineren met methoden om processen continu te verbeteren en met de gerelateerde vormen van personeelsmanagement. Vitaal is dat artsen het voortouw nemen om processen integraal aan te sturen en te verbeteren.

Onderzoek

In **hoofdstuk 5** evalueren we of toegangstijden opgelost kunnen worden door de wijze waarop vraag en aanbod afgestemd worden te verbeteren met WZW, zonder capaciteit toe te voegen. We evalueerden systematisch van de interventies van achttien poliklinieken die het WZW model toegepast hebben en we voerden een statistische analyse op de effecten van de interventies op hun toegangstijden.

De poliklinieken pasten verschillende combinaties van interventies toe om de wijze waarop ze vraag en aanbod afstemden te verbeteren, de efficiëntie van hoe hun aanbod is georganiseerd te verbeteren en om onnodige vraag naar afspraken te reduceren. Veertien poliklinieken toonden significante verbeteren. Twee hebben waarschijnlijk hun toegangstijd verbeterd en twee is dat niet gelukt. Hun toegangstijd nam gemiddeld met 55% af van 47 naar 21 dagen.

Het lijkt erop dat toegangstijden voor een substantieel deel opgelost kunnen worden door de wijze waarop vraag en aanbod worden afgestemd. Beleidsmaker dienen te analyseren of vertragingen veroorzaakt worden door capaciteitsproblemen of afstemmingsproblemen. Voor de laatste problemen lijkt het effectiever om te investeren in het vergroten van het vermogen om te reageren met behulp van pull methoden, dan het vermogen om te plannen. Beleidsmakers dienen prikkels te creëren om de toegangstijden kort te houden en prikkels te verwijderen die toegangstijden stimuleren.

Hoofdstuk 6 onderzoekt de duurzaamheid van de resultaten van WZW. Ons doel was om vast te stellen of de bereikte verbeteringen vastgehouden zijn en om de factoren te identificeren die de duurzaamheid beïnvloeden vanuit het gezichtspunt van de betrokken actoren. We voerden kwalitatieve case studies uit in veertien poliklinieken. We vergelen hun toegangstijd tijdens de start, op het einde van het project en drie tot vier jaar later. Behouden en nieuwe interventies zijn geanalyseerd. Interviews met 52 betrokkenen zijn geanalyseerd met de 'constant comparative method' om de generieke factoren te identificeren die duurzaamheid beïnvloeden.

Elf van de veertien poliklinieken bleek in staat hun verkorte toegangstijden te behouden of verder te verbeteren; twee lukte dat niet en voor een is het onzeker. De poliklinieken pasten nog steeds de meerderheid van de interventies toe en allen hadden nieuwe interventies geïntroduceerd. Drie generieke factoren kwamen naar voren die de mate van duurzaamheid heeft beïnvloed:

toegenomen responsiviteit om vraag en aanbod af te stemmen; klinisch leiderschap en prikkels; een gezamenlijk geloof dat ze de toegangstijd kunnen en moeten beheersen.

Korte toegangstijden kunnen duurzaam behouden blijven als zowel de manier van denken als het planningssysteem vraag gestuurd en flexibel is en als degenen die zorg leveren en voordeel bij ervaren. In tegenstelling tot eerdere onderzoeken bleek senior management ondersteuning en formele training niet relevant, maar klinisch leiderschap en informele socialisatie daarentegen wel. Multidisciplinaire teams verantwoordelijk maken voor het verbeterproces lijkt cruciaal te zijn.

Hoofdstuk 7 presenteert de bevindingen van diepgaand veldonderzoek waarmee karakteristieken geïdentificeerd zijn hoe poliklinieken vraag en aanbod afstemmen. We gebruikten praktijk observaties, documenten en data van een dermatologische polikliniek met weinig afhankelijkheid van andere afdelingen en van een orthopedische polikliniek met een hoge afhankelijkheid van de operatiekamers.

Het is waarschijnlijk dat de wijze waarop beide poliklinieken vraag en aanbod afstemmen onnodige vertragingen in de toegangstijd veroorzaken. Dit heeft te maken met een gebrek aan anticiperen van fluctuaties in de vraag, beperkte flexibiliteit om de volume van het aantal afspraken te vergroten in reactie op tekorten en een gebrek aan flexibiliteit om de cycle mix van spreekuren aan te passen naar gelang omstandigheden dat vergen. Een polikliniek ontbreekt tevens de benodigde flexibiliteit in de mix van afspraken in spreekuren. De beperkte flexibiliteit om spreekuren en operatie sessies uit te wisselen is een verder beperkende factor om toegangstijden te optimaliseren. Beide poliklinieken gebruiken overwegend push methodes, al worden wel degelijk enkele pull methoden gebruikt om op basis van de actuele toestand van het systeem acties te bepalen.

Om toegangstijden te optimaliseren lijkt het vitaal te zijn dat de betrokken specialisten hun identificatie met toegangstijden verschuiven van 'lange toegangstijden zijn een teken van een goede dokter' naar 'korte toegangstijden zijn een teken van een goede dokter'.

Synthese en discussie

Hoofdstuk 8 creëert synthese door het theoretisch raamwerk en de onderzoeksresultaten te combineren in een nieuw raamwerk om de toegangstijd tot poliklinieken te optimaliseren. Als de vraag groter is dan het aanbod moet dit eerst in balans worden gebracht. We identificeerden drie vormen van verspilde capaciteit die gereduceerd kunnen worden: ongewenste annuleringen van spreekuren, no shows en onbenutte afspraken. Als toegangstijden ondanks dat er genoeg aanbod is kunnen drie methodes toegepast worden om vraag en aanbod beter te synchroniseren: vergroot het vermogen om fluctuaties in de vraag waar te nemen, vergroot de flexibiliteit in de volume van afspraken en vergroot de flexibiliteit in de mix van afspraken. We gebruikten het theoretisch raamwerk en de onderzoeksresultaten om een op pull gebaseerde Productie Planning en Controle (PPC) raamwerk te creëren waarmee poliklinieken dit kunnen vormgeven. Dit PPC raamwerk gebruikt de systeem status voor de vrijgave van spreekuren op een geaggregeerd niveau en de vrijgave van afspraken op een operationeel niveau.

De poliklinieken die hun toegangstijd duurzaam verkort hebben introduceerden pull methoden op zowel het geaggregeerde als het operationele niveau. Niet een includeerde andere afdelingen in de pull methoden. Dit is wellicht wel nodig om de toegangstijden verder te optimaliseren maar blijkt, in tegenstelling tot onze verwachtingen, niet noodzakelijk om toegangstijden duurzaam te verkorten.

De analyse toont dat de poliklinieken in staat waren toegangstijden duurzaam te verkorten door een deel van het PPC raamwerk toe te passen. Ze hebben:

- Endogene variatie in aanbod gereduceerd.
- De wijze van spreekuur planning veranderd van push naar een hybride mix van push en pull methoden.
- Expliciete of impliciete buffercapaciteit gecreëerd en methodes ontwikkeld om die efficiënt in te zetten.
- De wijze van afspraken plannen van push naar pull verschoven door een flexibele afspraken mix mogelijk te maken

De poliklinieken hebben de volgende elementen van het PPC raamwerk niet toegepast:

- Seizoensgebonden variatie in de vraag anticiperen voor nieuwe afspraken.
- Rekening houden met de actuele vraag naar toekomstige controle afspraken.
- Toekomstige vraag naar controle afspraken sturen naar gunstigere periodes.
- Een flexibele mix van spreekuren per periode mogelijk maken.
- De systeem status voor operatie sessies en eventueel andere afdelingen meewegen in het bepalen van het aantal spreekuren die gepland worden (en vice versa).

Om toegangstijden te optimaliseren lijkt het cruciaal dat de toepassing van het PPC raamwerk specialisten de mogelijkheid biedt hun identificatie met vertragingen in de toegangstijd te verschuiven op een wijze die een positieve cognitieve en affectieve gesteldheid handhaaft, hun begrip van een competente en effectieve dokter verruimd en hen in staat stelt hun visie op hoe zorg verleend dient te worden met korte toegangstijden te verbinden zodat ze als coherent en consistent ervaren. Het PPC raamwerk kan makkelijk het tegenovergestelde effect bereiken indien specialisten als middelen benaderd worden die gepland en beheerst moeten worden. Als specialisten het raamwerk echter ervaren als een middel om hun visie op kwaliteit van zorg te realiseren, zullen ze mogelijk de beslissingen leiden om toegangstijden te optimaliseren.

In **hoofdstuk 9** bediscussiëren we onze belangrijkste bevindingen. We concluderen dat de wijze waarop poliklinieken vraag en aanbod afstemmen omdat de methoden die gebruikt worden andere doelstellingen dienen. De beschikbare kennis van Operations Management (OM) biedt de benodigde kennis om de oorzaken van toegangstijden te adresseren, maar het wordt nauwelijks toegepast. Verder tonen onze bevindingen dat de manier van denken, de identiteit en het bijbehorende gedrag van specialisten een cruciaal onderdeel uitmaken van de oorzaken van vertragingen in de toegangstijd.

De onderzoeksresultaten tonen aan de poliklinieken hun toegangstijden substantieel gereduceerd hebben met WZW en dat de resultaten duurzaam zijn. Ze hebben pull methoden geïntroduceerd op het geaggregeerde niveau waar spreekuren gepland worden en op het operationele niveau waar afspraken gepland worden. Vanuit het gezichtspunt van OM zijn de reductie van ongewenste

annuleringen van spreekuren, de toegenomen flexibiliteit in het aantal spreekuren en van de mix van afspraken in spreekuren in combinatie met feedback op de actuele systeemtoestand de belangrijkste verklaringen waarom WZW effectief is gebleken. We denken dat de efficiëntie slagen een voorwaarde zijn geweest voor het optimaliseren van de toegangstijden, omdat zij de benodigde buffercapaciteit gecreëerd hebben om flexibel te worden in het aantal spreekuren. Vanuit het perspectief van de betrokken actoren in de poliklinieken zijn de motivatie van de specialisten en het gezamenlijke geloof in de teams ook van belang geweest.

In hoofdstuk 8 beschreven we welke elementen van het PPC raamwerk door WZW niet geadresseerd zijn. Inclusie van deze factoren kan poliklinieken mogelijk in staat stellen hun toegangstijd verder te optimaliseren met dezelfde of om een betere benuttingsgraad van de capaciteit te bereiken met dezelfde toegangstijd (= met minder buffercapaciteit).

Klinisch leiderschap is cruciaal geweest om duurzame verandering te bereiken, maar kan ook een kwetsbaarheid gecreëerd hebben omdat de resultaten in ieder geval deels afhankelijk zijn van de klinisch leider. We denken dat lange termijn duurzaamheid vergt dat de bereikte gedragsverandering onderdeel wordt van de dagelijkse routines van het team in de poliklinieken zonder klinisch leiderschap.

We concluderen dat de toepassing van WZW robuust is vanwege het vermogen om de interventies te continueren. We zijn echter niet zeker of het ook veerkrachtig is omdat we niet hebben kunnen vaststellen of de poliklinieken de interventies kunnen aanpassen naar gelang gewijzigde omstandigheden. We denken dat poliklinieken periodieke, incidentele verandering nodig hebben om toegangstijden te reduceren en continu veranderprocessen nodig hebben om de resultaten duurzaam te laten zijn en om toegangstijden verder te optimaliseren.

De belangrijkste methodologische tekortkomingen: de resultaten vergroten ons begrip van de bestudeerde fenomenen, maar met name de geïdentificeerde factoren die de duurzaamheid beïnvloeden zijn waarschijnlijk incompleet. Het aantal van achttien onderzochte poliklinieken is relatief klein. Onze conclusies zijn valide voor de combinatie van interventies, maar we kunnen niet vaststellen wat de effecten zijn van individuele interventies.

Het op pull gebaseerde PPC raamwerk biedt de elementen om toegangstijden tot poliklinieken te optimaliseren. Poliklinieken dienen voorspelbare fluctuaties te anticiperen, te reageren op daadwerkelijke fluctuaties in de vraag en ze dienen flexibiliteit te creëren in de mix van spreekuren en van de mix van afspraken binnen spreekuren om fluctuaties in vraag en aanbod gelijkmatig te kunnen opvangen. Om het PPC raamwerk toe te passen dienen poliklinieken te leren hoe ze een cycle mix inrichten en de mix van spreekuren per periode continu te optimaliseren en idem voor de mix van afspraken binnen spreekuren, gebaseerd op geanticipeerde fluctuaties in vraag en aanbod op de actuele systeem status. Om het totale logistieke systeem van een ziekenhuis te optimaliseren is een geïntegreerde benadering nodig wat WZW niet biedt en die voor zover wij kunnen overzien nog niet bestaat. De WZW poliklinieken optimaliseerden hun toegangstijd in hun eentje. Indien alle poliklinieken hun toegangstijd optimaliseren in een ziekenhuis zonder integratie met de andere afdelingen kan dit resulteren in een achteruitgang van de resultaten van de andere afdelingen in termen van toegangstijden en benutting van de capaciteiten.

We denken dat hoe meer de wijze waarop een systeem vraag en aanbod afstemt gekarakteriseerd wordt door complexiteit en onzekerheid, hoe belangrijker het vermogen wordt om te kunnen reageren op de actuele systeem toestand. We denken verder dat optimalisatie overwegend pull methoden vergt, al zullen push methoden nodig blijven omdat in een ziekenhuis altijd capaciteit vrij gegeven zal moeten worden voordat de werkelijke vraag en systeemtoestand bekend zijn. We beschouwen de lean filosofie als veelbelovend om de benodigde integrale aanpak te ontwikkelen die gebaseerd is op pull methoden, om een systeem en cultuur te ontwikkelen die door continu verandering het systeem in staat stelt zich continu aan te passen aan veranderende omstandigheden.

Om de optimale toegangstijd te bepalen is de benutting van capaciteit de belangrijkste balansmaat. We raden aan om eerst WZW toe te passen als er nog weinig kennis en ervaring is met OM en indien de toegangstijd nog niet eerder geoptimaliseerd is. De volgende stap is om het complete PPC raamwerk toe te passen om de toegangstijden verder te optimaliseren en om de benuttingsgraad van de capaciteit te verhogen. De Treeknorm voor toegangstijden (80% in drie weken, 100% in vier weken) dient daarbij beschouwd te worden als de ondergrens van acceptabele toegangstijden, niet het doel. Het optimum met behulp van WZW is waarschijnlijk voor de meeste poliklinieken ongeveer een week.

Om duurzame verandering te realiseren, moeten de systeem veranderingen gepaard gaan met een veranderaanpak die specialisten in staat stelt klinisch leiderschap te tonen en die multidisciplinaire teams in staat stelt een gezamenlijk vertrouwen te ontwikkelen dat ze het kunnen. Afdelingen hebben prikkels nodig om de toegangstijden voor andere afdelingen te optimaliseren in samenhang met die van henzelf. De alertheid van specialisten om kleine signalen van patiënten te herkennen en ernaar te handelen biedt een vermogen die het management dient te koesteren en op bouwen om de wijze waarop vraag en aanbod afgestemd wordt te optimaliseren. De rol van management is niet zozeer persoonlijk te pleiten voor de veranderingen, maar om een ondersteuning te bieden door structuren en hulp aan de teams zodat zij zelf hun toegangstijden kunnen optimaliseren. Een cruciaal element is feedback op basis van de systeem status te organiseren die beslissingen ondersteunt om capaciteit vrij te geven, juist niet vrij te geven of te voorkomen dat het ongewenst geannuleerd wordt.

We bevelen aan dat de kennis van OM om processen te verbeteren gecombineerd wordt met het vakgebied van organisatieverandering om een duurzame impact te bereiken. Gebaseerd op de resultaten van ons onderzoek over het belang van klinisch leiderschap en de relatie met de constructie van de identiteit van specialisten en teams, bevelen we aan om ook het veld van organisatie psychologie en betekenisgeving te combineren met onderzoek naar OM vraagstukken om de benodigde condities te identificeren voor verandering van processen en duurzame gedragsverandering om processen te optimaliseren. Om effectieve en duurzame verandering te bereiken is de vraag niet of de specialist, andere disciplines of het management in de lead moeten zijn, maar hoe synergie gecreëerd kan worden tussen de betrokkenen om toegangstijden te optimaliseren. Eerst in een periodieke, incidentele verandering gevolgd door een continu veranderen om doorlopend te leren en aan te passen aan nieuwe uitdagingen die zich zullen blijven voordoen.

Dankwoord

In 2005 vroeg Frits van Merode aan mij of ik wel eens een promotieonderzoek had overwogen aangezien ik met het Doorbraak programma Werken zonder wachtlijst op het (toenmalige) Kwaliteitsinstituut CBO veel ervaring en data had verzameld rondom het optimaliseren van patiëntenlogistiek. Frits was op dat moment als hoogleraar Logistiek en Operationeel Management expert in het programma. In diezelfde periode sprak ik Bart Berden als bestuurder van het St. Elisabeth Ziekenhuis over de mogelijkheid in het Elisabeth te komen werken en hij suggereerde mijn werk te koppelen aan een promotieonderzoek om mij op inhoud verder te ontwikkelen. Toen wist ik het wel: dicht op de praktijk kunnen werken in combinatie met conceptuele verdieping en die twee werelden verbinden geeft me enorm veel voldoening. Frits was bereid mijn promotor te worden en toen Bart later ook hoogleraar werd, was hij bereid mijn tweede promotor te worden. Ik wil beide bedanken voor hun vertrouwen in mij.; toen ik begon aan mijn promotieonderzoek besefte ik nog niet hoeveel die rol van jullie zou gaan vergen. Van de vele gesprekken met Frits heb ik het meest genoten van de verbindingen die we gelegd hebben met filosofie, sociale psychologie, veranderkunde en systeemdenken en in het bijzonder de zuiverheid van wetenschappelijk denken die je mij aangeleerd hebt. Van Bart heb ik het meest geleerd over logisch redeneren en daar consequent in zijn en het meest genoten heb ik van de geur van versgebakken brood. Beide ben ik immens dankbaar voor de tijd, geduld, denkkracht en inspiratie die jullie mij geboden hebben.

Het St. Elisabeth Ziekenhuis heeft mij in staat gesteld het promotieonderzoek te combineren met mijn werk, zonder die ruimte zou het mij niet gelukt zijn. Tijdens het laatste jaar van mijn onderzoek ben ik gaan werken voor het St. Antonius Ziekenhuis waar ik dezelfde ruimte heb gekregen, waar ik in het bijzonder Dirk Schraven dankbaar voor ben, alsmede het MO.

Mijn onderzoek is voortgekomen uit alles wat ik geleerd heb van het helpen verbeteren van patiëntenlogistiek met vele poliklinieken in vele ziekenhuizen. Dat heb ik samen met Stan Janssen mogen doen, die mij heeft doen beseffen wat synergie betekent (‘die tafels gaan door elkaar’). Wim Schellekens was degene die me uitgedaagd heeft hierin te ontwikkelen (‘zou dat in Nederland lukken?’). Mirella Minkman zorgde ervoor dat ik als jonkie die ruimte kreeg van het management (‘waarom zo ouderwets altijd de senior?’). Veel inspiratie heb ik opgedaan van de medisch specialisten die de lead namen in de projecten, in het bijzonder Jan van Bodegom (‘We annuleren voortaan binnen zes weken alleen nog een spreekuur in verband met een begrafenis, bij voorkeur die van jezelf’), Ina Kuper (‘We zijn een team’), Ineke Blauw (‘patiëntgestuurde zorg’), Aike Kruize (‘We zijn slechts een gast in het leven van een patiënt’), Tonnis Keijer (‘Ik kom wel op de fiets’), Eping Hu (‘de essentie is niet iets wat je hoeft te zoeken’) en Rob Blanken (‘Wat, de toegangstijd is al verdwenen!?!’) en Frans Boereboom (‘Het werkt nog niet, maar ik ga verder). Nòg een stap verder ging de samenwerking, doorzettingsvermogen, chemie en lol met Jacob Caron (‘Lean en Lief’) en Marcel Boonen (‘logica is het begin van de verwarring’) tijdens het project en vervolgens nog meer in de jaren erna. Zij en anders zoals zij vormen de innovatiemotor in de zorg om de kwaliteit van processen wezenlijk en duurzaam te verbeteren door tegen de stroom in te zwemmen en ik beschouw het als een eer met hen te kunnen samenwerken.

De 20 poliklinieken dank ik voor hun medewerking door data ter beschikking te stellen, zich te laten observeren en te laten interviewen. Stan Janssen en Jeroen van Wijngaarden dank ik voor de diepgaande, volhardende samenwerking tijdens ons onderzoek naar de duurzaamheid van Werken zonder wachtlijst. Henk Veraart dank ik voor het gezamenlijke artikel over Lean in de zorg en de vele inspirerende discussies daarover. Ik dank ook de studenten die hebben geholpen bij het verzamelen van data en het uitvoeren van interviews: Angélique van der Veer, Daphne Metaal, Linda de Jong, Paulien Schoneveld, Anita de Nooijer en Claire Hostmann. Voor de steun tijdens mijn onderzoek ben ik dankbaar aan de onderzoeksgroep CAPHRI, in het bijzonder Siebren Groothuis. De wetenschappelijke leergemeenschap Clinamen heeft me scherp gehouden mijn onderzoek in een bredere maatschappelijke context te plaatsen, in het bijzonder Ben van Lier, Arnold Roozendaal en Everard van Kemenade. De uitwisseling met de andere promovendi van Bart heeft me gestimuleerd. Idem voor het bestuur, de raad van toezicht en de promovendi van Lidz.

Dank ben ik verschuldigd voor het meedenken en becommentariëren door Peter Kabel, Jeff Kaas, Jos Benders, Kees Ahaus en Yvonne de Lau. Hollandse Hoogte dank ik graag voor het ter beschikking stellen van de illustratie op de voorpagina van het proefschrift.

Zover was ik echter überhaupt niet gekomen zonder het mentorschap van Teun Hardjono die mij geholpen heeft mijn kompas te richten, Peter Lemaire die mij geholpen heeft mijn kompas te kalibreren en Marius Buiting die me geleerd heeft vanuit mijn intuïtie die kompas te volgen en mij open heeft gesteld dat in de zorg te doen.

Mijn familie en vrienden dank ik voor de wijze waarop ze mijn onderzoek als een echte Trixhot hebben doen ervaren.

Al mijn dankbaarheid voor iedereen die mij gesteund heeft verbleekt echter bij mijn dankbaarheid voor de steun van Caroline, die niet alleen mij recht in de ogen keek toen we samen besloten dit te starten, maar die bovendien tijdens het hele avontuur naast mij in dezelfde richting heeft gekeken. Misschien ben ik door jou nog het meest gegroeid tijdens mijn onderzoek.

In de periode van het onderzoek zijn mijn drie dochters Amélie, Noémie en Eloïse ter wereld gekomen. Een intens genot.

Curriculum Vitae

Marc Rouppe van der Voort was born in Kanshasa (Zaïre) on January 7, **1974**. He finished Grammar School in Oss in **1993** and started his study of **Business Administration** at the Erasmus University Rotterdam where he specialised in Change Management. He worked from **1999** for two years as a trainer and junior consultant at **KDI** (Quality, Sustainability and Innovation), part of Schouten & Nelissen where he became fascinated with the practical challenges of quality management, process improvement and strategy deployment and change management in various sectors. In **2001** he became passionate at helping hospitals to improve processes as an advisor and later senior advisor and program manager at the Dutch Institute for Quality Improvement **CBO**. At CBO he was responsible for the Breakthrough Program Working Without a Waiting List (WWWL) ('Werken Zonder Wachtlijst (WZW)') that later became the starting point of his PhD research.

In **2006** he started to work as manager innovation at the **St. Elisabeth Hospital** Tilburg where he became responsible for developing the principles and methods of process improvement to enable all people of the hospital to improve a process every day with the principles of Lean. He combined this work with his PhD research at CAPHRI, **Maastricht University** in the same year. In **2009** he started a weekly blog on Lean in Healthcare, he co-published two books on the application of Lean in healthcare and in 2011 he became chair of the Dutch network for Lean in Healthcare. In **2010** he was installed by the Dutch Health Ministry as part of the Committee of experts for the program towards transparency in healthcare. In **2013** he started to work as manager Lean and patient logistics at the **St. Antonius Hospital** in Utrecht / Nieuwegein where he combines the knowledge he gained from his PhD research and the experience from his previous work to create pull, improve in small steps together while enabling people to grow.

In **2004** he married Caroline and since 2006 they live together in Tilburg with their three daughters Amélie, Noémie and Eloise.

List of publications

Peer reviewed journals

Marc Rouppe van der Voort, Jeroen van Wijngaarden, Stan Janssen, Bart Berden and Frits Van Merode. *Sustainability of improvements in access to outpatient specialist care in the Netherlands.* Journal of Health Services, Research & Policy. 2014 Vol 19(2) 94-101.

Marc Rouppe van der Voort, Frits Van Merode and Henk Veraart. *Duurzaam procesverbetering met de 'lean'-filosofie*. Ned Tijdschr Geneeskd. 2013;157:A5541.

Roemeling, O., Land, M.J., Ahaus, C.T.B. and Rouppe van der Voort, M. (2013, June). *The impact of lean interventions on buffer reduction in a hospital setting.* Paper presented at the 20th EurOMA conference: Operations management at the hearth of the recovery, Dublin.

M. Rouppe van der Voort, B. Berden and F. Van Merode. *Making sense of delays in outpatient specialty care.* Health Policy, August 2010;97;44-52.

Books

M. Rouppe van der Voort. *Duurzaam processen verbeteren lukt alleen als mensen groeien.* Chapter in: Perspectieven op Kwaliteit.nl, Teun Hardjono et al. NNK, Enschede, 2012.

M. Rouppe van der Voort and J. Benders. *Lean in de zorg: de praktijk van continu verbeteren.* Lemma, Den Haag, January 2012.

J. Benders, M. Rouppe van der Voort and B. Berden. *Lean denken en doen in de zorg: acht verhalen uit de praktijk.* Lemma, Den Haag, January 2010.

P. van Splunteren et al. *Doorbreken met resultaten. Verbeteringen van de patiëntenzorg met de Doorbraakmethode*. Van Gorcum, Assen, 2003.

P. Lemaire and M. Rouppe van der Voort. *Kluwer Handboek Integrale Kwaliteitszorg*, 2002; Chapter 1.6 *Organisatiekunde en kwaliteitszorg* and Chapter 1.8 *Kwaliteitsborging.*

Journals

Eveline van Herwaarden. *Lean in de zorg*. Interview in Kwaliteit in Bedrijf. 2013;Aug.

Redactie KiZ. *Lean en de ijzeren discipline om te blijven verbeteren.* Interview Marc Rouppe van der Voort and Jannes Slomp. Kwaliteit in Zorg. 2013;1.

R. Notten, *Blijf successen vieren*. Interview in Arts en Auto. *January 2013.*

M. Rouppe van der Voort, Revalidatiesector: leer lean denken! *Standpunt in Revalidatiemagazine, December 2012.*

S. Lagewaardt, Lean, de weg van continu verbeteren*. Interview in De Specialist, September 2012.*

M. Rouppe van der Voort. Wat is nodig om lean duurzaam toe te passen? *Zorgmarkt, 2011;3, p.42.*

B. van der Linden. Lean werken in de zorg: twee ervaringsdeskundigen aan het woord. Interview Marc Rouppe van der Voort and Ronald Aalbersberg*. Kwaliteit in Zorg (KiZ), 2011;3, p.10-12.*

Ziekenhuis opereert zoals Toyota produceert*, Brabants dagblad, 16 October 2010.* Interview Marcel Visser and Marc Rouppe van der Voort.

M. Stoffer, M. Rouppe van der Voort en S. Janssen. *Processen herinrichten blijkt inherent onvoorspelbaar proces*. Best Practice, April 2009;2.

M. Rouppe van der Voort. *Wat voor soort bedrijf willen we zijn? Het voorspellen van onvoorspelbare zorg.* Interview Mark van Houdenhoven, Zorgmarkt, March 2008.

M. Hoogendijk. *Duurzame innovatie in het ziekenhuis. Wat bezielt Marc Rouppe van der Voort?* Interview in Sigma, nr.1, February 2008.

M. Rouppe van der Voort, B. Berden and F. van Merode. *De mythe van de beheersbaarheid*, Medisch Contact 2006; 37; 1444-6.

H. Backes, R. Hermans, Janssen, S and M. Rouppe van der Voort, *Een uitgekiende omwenteling. Radiotherapiekliniek weet met vernuft de doorstroomtijden te verkorten,* Medisch Contact, 2004;17.

M. Rouppe van der Voort, S. Janssen, M. Minkman, S. Slot and W. Schellekens. *De balans opgemaakt: Doorbraak-aanpak succesvol om toegangstijden tot een week te bekorten*, Medisch Contact; 2004; 14; 546-49.

J. van Bodegom, M. Rouppe van der Voort and F. van Merode. *Controle over vraag en aanbod,* Medisch Contact; 2004; 12; 469-71.

I. Blaauw, E. Ton, S. Simons, B. Raadgever and M. Rouppe van der Voort. *Doel gemist, missie geslaagd,* Medisch Contact; 2004; 13; 500-3.

I. Kuper, M. Kruisen, V. Sauerwald and M. Rouppe van der Voort. *De winst van gedeelde zorg,* Medisch Contact; 2004; 10; 370-72.

W. Schellekens, M. Rouppe van der Voort and P. van Splunteren*. Steen in de vijver: ziekenhuizen stimuleren om bewezen verbeteringen in te voeren;* Medisch Contact; 2003; 35; 1302-04*.*

M. Rouppe van der Voort. *Leadership & revolutionary breakthroughs in Healthcare*; 2003; executive summary of theme 4.4 of the 47th EOQ congress.

M. Minkman, M. Rouppe van der Voort, S. Janssen and W. Schellekens. *Werken zonder wachtlijst, minder stress op de polikliniek*, Medisch Contact; 2003; 11; 416-18.

M. Rouppe van der Voort. *Leiderschap in de gezondheidszorg*; Sigma; 2002; 6; 42-44.

M. Rouppe van der Voort. *De BenchmarkIndex: Wordt u al geïnspireerd door uw Europese Collegae?;* Specifiek; 2000, 192.

T. Hardjono and M. Rouppe van der Voort*. Naar een Europese Visie op Kwaliteit;* Sigma; 2000; 4.