

Implementation of local quality improvement collaboratives to improve
test ordering and prescribing performance of general practitioners



Jasper Trietsch

© **Jasper Trietsch**, Maastricht 2016

No part of this book may be reproduced or transmitted in any form or by any means, without prior permission in writing by the author, or when appropriate, by the publishers of the publications.

Layout: Tiny Wouters

Cover design: Ubbo Noordhof

Production: Datawyse | Universitaire Pers Maastricht



ISBN: 978 94 6159 588 1

The research presented in this thesis was conducted at the School for Public Health and Primary Care: CAPHRI, department of Family Practice, of Maastricht University. CAPHRI participates in the Netherlands School of Primary Care Research: CaRe.

The studies in this thesis were supported by unrestricted grants from the Netherlands Organisation for Health Research and Development (ZonMw) and health insurance company CZ.

Implementation of local quality improvement collaboratives to improve test ordering and prescribing performance of general practitioners

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
vrijdag 21 oktober 2016 om 12.00 uur

door

Jasper Trietsch

Promotor

Prof. dr. G.D.E.M. van der Weijden

Prof. dr. J.F.M. Metsemakers

Prof. dr. mr. R.P.G. Grol (Radboud UMC Nijmegen)

Beoordelingscommissie

Prof. dr. G.J. Dinant (voorzitter)

Dr. L. van Bokhoven

Prof. dr. F.L.V.M. Buntinx

Dr. J. Burgers (NHG)

Prof. dr. M.J.P. Wensing (Radboud UMC Nijmegen)

Contents

Chapter 1	General introduction	7
Chapter 2	A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: Study protocol. <i>[Published as: Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijters P, Winkens R, van Steenkiste B, Grol R, Metsemakers J Implement Sci 2009, 4:6.]</i>	23
Chapter 3	The challenge of transferring an implementation strategy from academia to the field: A process evaluation of local quality improvement collaboratives in Dutch primary care using the normalization process theory <i>[Published as: Trietsch J, van Steenkiste B, Hobma S, Frericks A, Grol R, Metsemakers J, van der Weijden T. J Eval Clin Pract 2014, 20(6):1162-1171.]</i>	53
Chapter 4	Effect of audit and feedback with peer review on general practitioners' prescribing and test ordering performance: A cluster-randomized controlled trial <i>[Submitted: Trietsch J, van Steenkiste B, Grol R, Winkens B, Ulenkate H, Metsemakers J, van der Weijden T]</i>	73
Chapter 5	The balanced incomplete block design is not suitable for the evaluation of complex interventions <i>[Published as: Trietsch J, Leffers P, van Steenkiste B, Grol R, van der Weijden T: J Clin Epidemiol 2014, 67(12):1295-1298.]</i>	149
Chapter 6	General discussion and Summary	157
	Summary	177
	Samenvatting	185
	Valorisatie	193
	Dankwoord	201
	Curriculum vitae	207



Chapter 1

General introduction

Introduction

In this introduction chapter the background of the subject of this thesis will be described. Dutch general practice faces a problem of prescribing and test ordering, requiring the need for strategies to influence professionals' behaviour and to implement potential effective strategies for quality improvement. At the end of this chapter the research questions, the objectives for this thesis and an outline of this thesis will be presented.

Healthcare workers and policy makers worldwide are searching for ways to deal with the problems they are facing in healthcare. For policy makers the increase in spending on healthcare is a major concern. It is estimated that in the U.S.A. 25.4 billion dollars annually could be saved on unnecessary or inappropriate test ordering, prescribing and radiology orders by investing a fraction of this amount into health information technology.¹ In the years 2004-2011 the average growth in prescriptions in the Netherlands was 5.7% annually. The growth of the national income of the Netherlands was smaller than the growth of the budget of healthcare year after year.^{2,3} If nothing is done to reduce this spiraling spending on healthcare, it is feared that industrialized countries such as the Netherlands won't be able to pay the bill for healthcare anymore in the long term. The growing use of medical services by the public, the growing burden of chronic disease in elderly, expanding biomedical possibilities, increase of administration, the shift towards more defensive medicine and the availability of very expensive treatments are some of the causes of this growth. The problem for physicians is that they face challenges such as how to keep up with the ever-growing volume of evidence from medical research. It is almost impossible for them to remain fully up to date because of the volume of evidence from research published daily.⁴ Also the policy makers, and with them the general public demand more transparency of healthcare workers and institutions in order to value their contribution to healthcare and to differentiate between necessary and potential unnecessary care.

In the Netherlands a solid health care system has evolved over the years with the general practitioner (GP) providing efficient and comprehensive care. The GP acts as a guardian against unwarranted use of drugs, diagnostic facilities and as a gatekeeper to medical specialist care. Although this system has often been referred to as being an important factor in controlling the growth of expenditure on healthcare in the Netherlands, room for improvement remains.⁵⁻⁷ The Dutch College of General Practitioners (DCGP) recognized quickly after the introduction of evidence based medicine in the late 1970s that reviews and guidelines could help physicians by summarizing large quantities of information and making information more easily available to field workers. The DCGP started in 1985 with a project designed to translate evidence to the field. In 1989 this resulted in the first clinical guideline, entirely focused on the care provided by general practitioners in The Netherlands. In 1995 Dutch GPs had access to 50 evidence based guidelines on diseases and illnesses

relevant to them, increasing to 90 guidelines in 2009.⁸ Nowadays over 100 guidelines have been published by the DCGP.

Adherence to aspects of these guidelines proves to be difficult; even with general adherence to these guidelines being approximately 70%, the inter-physician or practice variation remains large as is shown in Table 1.1.⁹⁻¹² There may be sensible and solid patient or professional-related reasons to deviate from guidelines, such as multimorbidity in a single patient, demographic differences or patients' preferences, or a physician's level of uncertainty tolerance; However an important part of the difference remains unaccounted for and can be regarded as unwarranted practice variation.^{13,14}

The inter-physician variation leads to inappropriate care when it is a symptom of underdiagnosing or undertreating one group of patients and at the same time overdiagnosing and overtreating another group and should as such be perceived as detrimental.¹⁵ The inter-physician variation in general practice with regard to test ordering and prescribing is considerable.^{9,16-18} As 80% of all drugs prescribed in the Netherlands are prescribed by general practitioners they could be a valuable target to start intervening on these prescribing patterns. The questions are how unnecessary care and practice variation can be addressed effectively and how physicians' behaviour can be changed.

Table 1.1 Adherence to Dutch guidelines: means in % and range between indicators.

	% Adherence	Range between indicators (%)
Diagnostics, 11 indicators	65	13-96
Imaging techniques	76	13-96
Laboratory testing	53	32-78
Drugs, 44 indicators	68	10-99
Don'ts	78	33-99
dos	62	10-99
Referral primary and secondary care, 25 indicators	89	41-100
Total	74	10-100

Source: Braspenning et. al. 2006.¹²

How to influence test ordering and prescribing effectively?

Many studies have tried to find evidence for effective strategies to improve quality of care. In Northern America and Europe small group quality improvement collaboratives are widely implemented in general practice. In these small groups healthcare professionals gather on a regular basis to discuss current issues and gain new insights on diagnostics and treatments. These so-called peer review groups of healthcare professionals are attractive for interventions aimed at changing professional behavior.¹⁹⁻²³

This thesis is in line with earlier work from our research group on influencing professional clinical behavior. In a preceding robust trial Verstappen et al. have shown the beneficial effects of these peer review meetings on test ordering behavior concerning various clinical topics. The peer review strategy that was evaluated by Verstappen showed a relevant reduction in volumes of tests ordered over a period of six months, with absolute differences ranging from 8-12%.^{24,25} This strategy tested audit, comparative feedback and group discussion with peers on six topics for continuing professional development focussing on test ordering by general practitioners. Lagerlov et al. showed that an intervention with much resemblance to the intervention as performed by Verstappen, improved appropriate drug treatment of asthma patients by 21% and urinary tract infections by 108%.²⁶

Similar effects on volumes of tests ordered as found in the Verstappen study have been found for more or less similar multifaceted implementation strategies.^{27,28} The search for the components of an intervention that do have an impact on professional behaviour has led to a shortlist of strategies that seem more effective than others. Some of these components are audit combined with feedback, guideline dissemination, small group work, standardizing diagnostic order forms, outreaching visits from opinion leaders or academic specialists to healthcare workers in the field and financial incentives. The Cochrane Effective Practice and Organization of Care group (EPOC) systematically reviews studies on educational, behavioural, financial, regulatory and organisational strategies to improve quality of care. Their work has led to the insight that multifaceted strategies are in general more effective than single faceted interventions,^{29,30} although Grimshaw et al. did not entirely confirm this.³¹ Multifaceted strategies are those strategies that combine several components into a single intervention. The hopes are that exposing professionals to a multifaceted strategy will increase the odds of having beneficial effects because of synergetic effects.

Another prevailing insight is that the possible effect of an intervention is larger when strategies are tailored to the needs and beliefs of the targeted population and when barriers and facilitators of change are addressed. Passive dissemination of guidelines and recommendations alone has a small effect on professional behaviour.²⁹ Audit and feedback have often been tested and showed mostly a modest effect in terms of both reducing volumes of tests ordered or drugs prescribed and practice variation. The effect of audit and feedback on adherence to desired practice ranged from -10% to +68% (median +16%).^{30,32-35} In other studies the sole introduction of a problem-based test ordering form proved to improve test ordering at low costs.^{24,25,31,36-39} Also there is evidence suggesting that only the provision of information on test fees to physicians already reduces the volume of tests ordered.⁴⁰ Small group peer review using direct individual feedback seemed to reduce inappropriate prescribing as well.⁴¹⁻⁴³

Thus a multifaceted strategy comprising different elements, including audit and feedback and small peer group work reinforced with opinion leaders seems to combine to the most effective intervention. Financial incentives and the provision of data on

fees could be added to this as they also show promising results. What remains unclear however is what the long term effects of these interventions are once they have been implemented.

Changing professional behavior applied to the Dutch situation in general practice

In our study we more or less copied the Verstappen strategy but added the component of feedback on prescription rates. Moreover we planned this intervention to be situated in regular pharmacotherapeutic audit meetings (PTAM) in general practice (see box 1.1). We invited not only the early innovators but offered this intervention to all PTAM groups in our research area. Hereby we wanted to test whether the results from previous work would hold once implemented in normal quality improvement (QI) settings. Because of the nature and stability of these PTAM groups they provide an excellent and safe environment for peer review. We expected that this existing system of PTAM groups could ensure sustainability of the peer review strategy implementation. The intervention has changed the nature of the PTAM groups into local quality improvement collaboratives (LQICs) with a much broader perspective on quality of care. We expected that the effects found in earlier controlled trials to be confirmed after transfer ring the strategy to the setting of normal QI-practice. By aiming at both test ordering and drug prescribing, our combination strategy could lead to an even larger effect than the Verstappen trial did because of synergy.

Box 1.1 PTAMs in the Netherlands

In the Netherlands, existing networks of pharmacotherapeutic audit meetings was used to disseminate and implement guidelines on test ordering and prescribing. The goal of setting up these meetings by general practice providers always has been to improve the quality of their prescribing behaviour.²⁵ These groups usually consist of six to ten GPs with affiliated community pharmacists.²⁶ During meetings, they discuss the choice of drugs in the context of a specific illness or disease. In recent decades, this form of continuing professional development has gained widespread acceptance amongst GPs and policymakers in the Netherlands. However, they tend to offer little or no room for discussions on test ordering. Because no other system of regular meetings exists, the possible underuse, overuse, and misuse of diagnostic services is not discussed by GPs on a regular basis. The Dutch Institute for the Rational Use of Medicine (IVM) supports and initiates local or regional implementation of quality improvement projects on the use of drugs and supports local PTAM groups with information and educational materials. Performance levels of PTAM groups are assessed once a year and rated on the basis of four levels, level one being the

poorest level of performance and level four the highest. Participation in PTAM groups by GPs is facilitated by national and regional support organizations for primary care, as well as by the government and through incentives by insurance companies. Attendance at PTAM meetings is rewarded by accreditation. Currently, approximately 50% of the group meetings reach the desired level of performance described by policymakers. To reach this level, groups must at least use quantitative feedback data on prescribing, create working agreements, discuss barriers to change, and evaluate working agreements. Most groups are stable and remain together for 10 years or more, with members only being replaced gradually.⁴⁴

Challenges in implementation science

Implementation is the work that has to be done to effectively change usual practice or procedures. It follows on preliminary thinking on how to put change into effect. The work that comes with implementing change is dependent on the context the change is effectuated in. Translating knowledge to the field often proves to be bothersome and elaborate. Also much is unclear on how to implement effectively, therefore worldwide organisations were instituted to find answers to these questions.

In the Netherlands Grol identified six stages in quality of care improvement in his model of effective implementation.⁴⁵ In the first stage, new research findings, new guidelines, experienced weaknesses, or best practices create an opportunity for quality improvement. In the second stage, after the initial implementation process has been planned, targets for improvement or change are set. Prior to the actual implementation, the performance, target group, and setting are analysed. In the fourth stage, the strategies that are to be used are identified and tested. The implementation is developed, tested, and executed. Finally, the implementation is evaluated and adapted, if necessary.⁴⁶ There are many more implementation models available to researchers and policy makers. Most are partly based on the work of Rogers, the PRECEDE-PROCEED model of Green et al and the Plan-Do-Study-Act cycles as defined by Deming.^{47,48} Most models have in common that they describe influencing contextual factors such as setting, economic aspects, professional attitudes and beliefs as well as attitudes and behaviour of patients or professionals. Next to the contextual factors often intervention factors are addressed. These include amongst others the knowledge to practice gap (the wider the gap the greater the difficulties), the resources available and the support provided. May et al. developed a theory that focuses on the process of transferring an intervention to the field, the Normalization Process Theory (NPT). This theory focuses on factors that result from co-operative and collective activities, but are experienced and accounted for by individuals.⁴⁹⁻⁵¹ These factors relate to the interactions between the people and organizations involved in the intervention, to the organizational contexts in which these interactions take place, to the internal and

external coherence of the intervention, to the work that comes with building and sustaining the context of the intervention and to the work of evaluating the intervention.^{52,53} Since our goal was normalization of the strategy, defined as sustainable incorporation into the existing LQICs we found the NPT the best suitable model available.

The explanatory – pragmatic continuum

Most studies performed have been conducted in an explanatory setting with a focus on a single or limited number of clinical topics; meaning that the researchers have described their intervention thoroughly in a protocol that leaves not much room for the healthcare providers to adapt the intervention to their own situation or needs. The main reason is that it enhances the internal validity. In order to facilitate the analysis of the trial often a very limited set of clinical topics is offered to participants thus diminishing the external validity of these results. On the other end of the research design spectrum there are the pragmatic trials that use the context of everyday practice. In these trials the internal validity is challenged but the potential gain is in a higher external validity.^{54,55} Little research has been done so far on the incorporation of quality improvement strategies in usual care. This requires a truly pragmatic trial in which the internal validity is as high as possible without compromising the pragmatic nature of the trial. Most studies on implementation strategies actively involve the researchers as they guide the professionals through the intervention. Follow-up time is generally very short, and often relatively simple behavioral changes are aimed for.^{29,30,43,56-58} Therefore, insight is lacking in whether and how the beneficial effects can be transferred to normal practice. However, there is some evidence from other research groups that the findings from our earlier research can be transferred successfully. But, the level of control of the academic researchers in these trial interventions seemed, at closer look, higher than we anticipated on in our pragmatic trial.^{59,60} It therefore remains unclear what the effects of our approach are without academia monitoring and adjusting processes. With much research being conducted in an explanatory way much resources and time put into the research will not necessarily return in savings or health improvement. The resources available for pragmatic research with a focus on implementation is limited, partly because research is often funded or initiated by the pharmaceutical companies that do not have much interest in this type of research.⁶¹⁻⁶³ Also pragmatic research is, because it takes place in widely varying contexts, time consuming and imposes great difficulties on the data analysis and interpretation because of its inherited flaw of internal validity. On the contrary classical explanatory research with a focus on a limited set of clinical problems and well described populations will lead to results that are much quicker available.

We designed this trial to determine the effects of feedback and peer review within normal practice settings, without external support from an academic research group. Therefore we use the existing LQIC structure as a backbone for the implementation of

our strategy comprising effective elements from previous work. We adapted the feedback and peer review strategy from Verstappen on three elements. First we added prescribing behavior as a target for feedback and peer review, which seemed logical as it had been proven effective in separate trials.^{43,57,58,60,64-66} The second new element was that we provided distance support only to key stakeholders, such as pharmacists and laboratory specialists, instead of direct support to all the LQICs. This included that LQICs were not to be recruited by the research group but by the regional health officers or laboratory specialists themselves. Thirdly, we allowed the LQIC groups freedom of choice regarding clinical topics. The last two elements are indispensable in a truly pragmatic trial.^{22,55,67,68}

The use of the balanced incomplete block design

A design sometimes used in implementation science research is “the balanced incomplete block (BIB) design”. We planned to use this design also for our intervention. Study of the literature on the BIB design raised questions on our choice for this design. It could very well be that the design we chose is not that suitable for use in research with humans. An early description of the BIB design did not reassure us that our choice was a wise one.⁶⁹ A true BIB design is based on a set of assumptions that are rarely met in research with humans. The methods needed to evaluate the effects of an intervention that is designed according to the BIB design has to take these assumptions into account. The question therefore remains whether this type of research design is the same as used before and as we planned to use. If not, then what should we call the type of design we and others before us used and is it even suitable for use in research with humans? When it is suitable, did others before us take the set of assumptions into account when analyzing their work?

Aim and research questions

The general aim of this PhD thesis was to evaluate the effects on volumes of tests ordered and drugs prescribed in general practice of a large scale implementation of a multifaceted intervention using audit, feedback and peer review embedded in existing local quality collaboratives (LQICs) with minimal steering from academia. Therefore this thesis focuses on the following research questions:

Process

- Was the strategy implemented as planned?
- What were the barriers and facilitators of the implementation of the strategy?
- Has the level of group performance improved in the participating groups?

Effect

- Do the volumes of tests ordered and drugs prescribed change in the preferred direction, as described in the working agreements of the LQICs, compared to baseline?
- What is the effect of this strategy on the inter-physician variation of the volumes of tests ordered and drugs prescribed by GPs for specific clinical topics, compared to GPs exposed to the same strategy but for other clinical topics?

Methodology

- Is the balanced incomplete block design suitable for use in implementation science?

Outline of this thesis

The goal of this study was to evaluate the effects of an intervention consisting of audit and feedback embedded in local quality improvement collaboratives after implementation in general practice in the South of the Netherlands.

In the last decade much effort is put into researching the gap between research and practice. Policy makers, governmental institutions, healthcare professionals as well as researchers strive to translate new findings into practice as quick as possible to improve quality of care and reduce costs or waste. In the Netherlands general practitioners take part in structural meetings on pharmacotherapeutic choices in everyday practice. These structures were used to implement an effective intervention to change professional behavior on pharmacotherapy and test ordering.

Chapter 2 describes the protocol of the trial. What was known at the start of the trial on influencing test ordering, prescribing behavior and implementation science is outlined. Both the qualitative and quantitative research questions are described that

helps us to gain insight in how to effectively implement the strategy and its effects after implementation. In chapter 3 the results of the qualitative process evaluation study are reported. The reasons for GPs to enter this trial as well as the successes and failures of the implementation process are presented. The normalization process model is used to organize these factors. The effects on the volumes of tests ordered and drugs prescribed are reported in chapter 4. First the results of a before-after analysis are provided and then the results of the per-protocol analysis. Chapter 5 is a reflection on the methodology often used in implementation science. We show that the design that is often used, the balanced incomplete block design, is actually not suitable for this type of research. Moreover we point out where and why the use of the term balanced incomplete block design went wrong. In chapter 6 the main results and conclusions of the thesis are presented and discussed.

References

1. Bentley TG, Effros RM, Palar K, Keeler EB. Waste in the U.S. Health care system: a conceptual framework. *Milbank Q* 2008;86(4):629-659.
2. Huisarts - Prescripties naar leeftijd [GP: prescriptions per agegroup] [www.nivel.nl/node/3099]
3. Westert G, van den Berg M, Zwakhals S, Heijink R, de Jong J, Verkleij H. Zorgbalans [balanced care], vol. 2014: Bohn Stafleu Van Loghum; 2010.
4. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7(9):e1000326.
5. Delnoij D, Van Merode G, Paulus A, Groenewegen P. Does general practitioner gatekeeping curb health care expenditure? *J Health Serv Res Policy* 2000;5(1):22-26.
6. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Q* 2005;83(3):457-502.
7. Cardol M, van Dijk L, de Jong J, de Bakker D, Westert G. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Huisartsenzorg: wat doet de poortwachter? [Second National Survey on diseases and procedures in the general practice. Primary care: what is the effect of the gatekeeper?]. In. Utrecht/Bilthoven: NIVEL/RIVM; 2004.
8. History of the Dutch College of General Practitioners (NHG) [www.nhg.org/mijlpalen]
9. Braspenning JCC, Schellevis FG, Grol R. Kwaliteit huisartsenzorg belicht [Quality in primary care reviewed]: Nivel; 2004.
10. Burgers JS, Grol RPTM, Zaat JOM, Spies TH, van der Bij AK, Mookink HGA. Characteristics of effective clinical guidelines for general practice. *Br J Gen Pract* 2003;53(486):15-19.
11. Muijters PE, Grol RP, Sijbrandij J, Janknegt R, Knottnerus JA. Differences in prescribing between GPs. Impact of the cooperation with pharmacists and impact of visits from pharmaceutical industry representatives. *Fam Pract* 2005;22(6):624-630.
12. Braspenning J, Schellevis F, Grol R. Assessment of primary care by clinical quality indicators. In: *Morbidity, performance and quality in primary care: Dutch general practice on stage*. edn. Edited by Westert GP, Jabaaij L, Schellevis FG. Oxon: Radcliffe Publishing; 2006:195-204.
13. van den Berg MJ, de Bakker DH, Spreeuwenberg P, Westert GP, Braspenning JC, van der Zee J, Groenewegen PP. Labour intensity of guidelines may have a greater effect on adherence than GPs' workload. *BMC Fam Pract* 2009;10:74.
14. de Jong JD, Groenewegen PP, Spreeuwenberg P, Westert GP, de Bakker DH. Do decision support systems influence variation in prescription? *BMC Health Serv Res* 2009;9:20.
15. Harteloh PPM. kwaliteit van zorg: van zorginhoudelijke benadering naar bedrijfskundige aanpak [Quality of care: from a care standpoint towards a business management standpoint], 4 edn. Maarssen: Elsevier/ De Tijdstroom; 2001.
16. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21(4):387-395.
17. Martens JD, van-der-Weijden T, Severens JL, de-Clercq PA, de-Bruijn DP, Kester AD, Winkens RA. The effect of computer reminders on GPs' prescribing behaviour: A cluster-randomised trial. *International Journal Medical Informatics* 2007;76(suppl 3):S403-416.
18. Berg MJ van den, de Bakker DH, van Roosmalen M, BJ. De staat van de huisartsenzorg [The state of primary care]. In. Utrecht; 2005.
19. Verstappen WHJM, van Merode F, Grimshaw J, Dubois WI, Grol RPTM, van Der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care* 2004;16(5):391-398.
20. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ* 2008;336(7659):1491-1494.
21. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Niessen LW, Grol RP: Short- and long-term effects of a quality improvement collaborative on diabetes management. *Implement Sci* 2010;5:94.

22. Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijters P, Winkens R, van Steenkiste B, Grol R, Metsemakers J: A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: Study Protocol. *Implement Sci* 2009;4:6.
23. Beyer M, Gerlach FM, Flies U, Grol R, Krol Z, Munck A, Olesen F, O'Riordan M, Seuntjens L, Szecsenyi J. The development of quality circles/peer review groups as a method of quality improvement in Europe. Results of a survey in 26 European countries. *Fam Pract* 2003;20(4):443-451.
24. Verstappen WHJM, van der Weijden T, Dubois WJ, Smeele I, Hermesen J, Tan FES, Grol RPTM. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med* 2004;2(6):569-575.
25. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermesen J, Grimshaw J, Grol RPTM. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289(18):2407-2412.
26. Lagerlof P, Loeb M, Andrew M, Hjortdahl P. Improving doctors' prescribing behaviour through reflection on guidelines and prescription feedback: a randomised controlled study. *Qual Saf Health Care* 2000;9(3):159-165.
27. Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, Soutter J, Wilsdon J, Matowe L, Needham G, Gilbert F *et al*. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet* 2001;357(9266):1406-1409.
28. Bunting PS, Van Walraven C: Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem* 2004;50(2):321-326.
29. Freemantle N, Harvey EL, Wolf F, Grimshaw JM, Grilli R, Bero LA. Printed educational materials: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2000(2):CD000172.
30. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006(2):CD000259.
31. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L *et al*. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8(6):iii-iv, 1-72.
32. Baker R, Falconer Smith J, Lambert PC. Randomized controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scan J Prim Health Care* 2003;21(4):219-223.
33. Sondergaard J, Andersen M, Stovring H, Kragstrup J. Mailed prescriber feedback in addition to a clinical guideline has no impact: a randomised, controlled trial. *Scand J Prim Health Care* 2003;21(1):47 - 51.
34. Martens JD, Winkens RA, van der Weijden T, de Bruyn D, Severens JL. Does a joint development and dissemination of multidisciplinary guidelines improve prescribing behaviour: a pre/post study with concurrent control group and a randomised trial. *BMC Health Serv Res* 2006;6:145.
35. Winkens RA, Pop P, Bugter-Maessen AM, Grol RP, Kester AD, Beusmans GH, Knottnerus JA. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet* 1995;345(8948):498-502.
36. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet* 2006;367(9527):1990-1996.
37. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to Improve Physicians' Use of Diagnostic Tests: A New Conceptual Framework. *JAMA* 1998;280(23):2020-2027.
38. Grol RPTM, grimshaw JM. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;362(9391):1225-1230.
39. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. *Med Care* 1993;31(9):784-794.
40. Feldman LS, Shihab HM, Thiemann D, Yeh HC, Ardolino M, Mandell S, Brotman DJ. Impact of providing fee data on laboratory test ordering: a controlled clinical trial. *JAMA Int Med* 2013;173(10):903-908.
41. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O'Brien MA, Johansen M, Grimshaw J, Oxman AD. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012, 6:CD000259.

42. Grimshaw JMMP, Shirran LMAM, Thomas RB, Mowatt GMAMBA, Fraser CMA, Bero LP, Grilli RMD, Harvey EB, Oxman AMDA, O'Brien MAM. Changing Provider Behavior: An Overview of Systematic Reviews of Interventions. *Medical Care August 2001*;39(8):II-45.
43. Arnold SR, Straus SE. Interventions to improve antibiotic prescribing practices in ambulatory care. *Cochrane Database Syst Rev 2005*(4):CD003539.
44. Eimers M, de Groot J. FTO peiling 2005, kwaliteit van farmacotherapieoverleg in Nederland in beeld [PTAM audit 2005, quality of pharmacotherapeutical audit meetings in the Netherlands visualised]. In.: DGV Nederlands Instituut voor Verantwoord Medicijngebruik [DGV the Dutch Institute for Rational Use of Medicine]; 2006.
45. Grol R. implementing guidelines in general practice care. *quality in health care 1992*;1:184-191.
46. Grol R, Wensing M, Eccles M. improving patient care. The implementation of change in clinical practice: Elsevier Limited; 2005.
47. Rogers E: Diffusion of Innovations. 2003.
48. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q 2004*;82(4):581-629.
49. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, Gask L, MacFarlane A, Murray E, Rapley T *et al.* Understanding the implementation of complex interventions in health care: the normalization process model. *BMC Health Serv Res 2007*;7:148.
50. May CR, Finch T, Ballini L, MacFarlane A, Mair F, Murray E, Treweek S, Rapley T. Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. *BMC Health Serv Res 2011*;11:245.
51. Normalization Process Theory On-line Users' Manual and Toolkit. [<http://www.normalizationprocess.org>]
52. Elwyn G, Legare F, Weijden T, Edwards A, May C. Arduous implementation: Does the Normalisation Process Model explain why it's so difficult to embed decision support technologies for patients in routine clinical practice. *Implement Sci 2008*;3(57):57.
53. Morrison D, Mair FS. Telehealth in practice: using Normalisation Process Theory to bridge the translational gap. *Prim Care Respir J 2011*;20(4):351-352.
54. Glasgow RE. RE-AIMing research for application: ways to improve evidence for family medicine. *J Am Board Fam Med 2006*;19(1):11-19.
55. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furlong CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ *et al.* A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol 2009*;62(5):464-475.
56. Fihn SD. Moving implementation science forward. *J Gen Intern Med 2006*;21 Suppl 2:S65-66.
57. van Driel ML, Coenen S, Dirven K, Lobbastael J, Janssens I, Van Royen P, Haaijer-Ruskamp FM, De Meyere M, De Maeseneer J, Christiaens T. What is the role of quality circles in strategies to optimise antibiotic prescribing? A pragmatic cluster-randomised controlled trial in primary care. *Qual Saf Health Care 2007*;16(3):197-202.
58. Riou F, Piette C, Durand G, Chaperon J. Results of a 12-month quality-circle prescribing improvement programme for GPs. *Br J Gen Pract 2007*;57(540):574-576.
59. Wensing M, Broge B, Riens B, Kaufmann-Kolle P, Akkermans R, Grol R, Szecsenyi J. Quality circles to improve prescribing of primary care physicians. Three comparative studies. *Pharmacoepidemiology and Drug Safety 2009*;18(9):763-769.
60. Wensing M, Broge B, Kaufmann-Kolle P, Andres E, Szecsenyi J. Quality circles to improve prescribing patterns in primary medical care: what is their actual impact? *J Eval Clin Pract 2004*;10(3):457-466.
61. Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care 2005*;43(6):551-557.
62. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA 2003*;290(12):1624-1632.
63. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet 2005*;365(9453):82-93.
64. Veninga CC, Denig P, Zwaagstra R, Haaijer-Ruskamp FM. Improving drug treatment in general practice. *J Clin Epidemiol 2000*;53(7):762-772.

65. Welschen I, Kuyvenhoven MM, Hoes AW, Verheij TJM. Effectiveness of a multiple intervention to reduce antibiotic prescribing for respiratory tract symptoms in primary care: randomised controlled trial. *BMJ* 2004;329:431-433.
66. Gjelstad S, Høye S, Straand J, Brekke M, Dalen I, Lindbaek M. Improving antibiotic prescribing in acute respiratory tract infections: cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ* 2013;347:f4403.
67. Sackett D. Explanatory versus management trials. In: *Clinical epidemiology: how to do clinical practice research*. edn. Edited by Haynes R, Sackett D, Guyatt G, Tugwell P. Philadelphia: Lippincott Williams and Wilkins; 2006:183-188.
68. Meeuwse EJ, German P, Melis RJ, Adang EM, Goluke-Willems GA, Krabbe PF, de Leest BJ, van Raak FH, Scholzel-Dorenbos CJ, Visser MC *et al*. Cost-effectiveness of post-diagnosis treatment in dementia coordinated by Multidisciplinary Memory Clinics in comparison to treatment coordinated by general practitioners: an example of a pragmatic trial. *J Nutr Health Aging* 2009;13(3):242-248.
69. Cochran WG, Cox GM. Experimental designs. New York: Wiley; 1950.



Chapter 2

**A cluster randomized controlled trial
aimed at implementation of local quality
improvement collaboratives to improve
prescribing and test ordering performance
of general practitioners: Study Protocol**

Jasper Trietsch, Trudy van der Weijden, Wim Verstappen, Rob Janknegt, Paul Muijers,
Ron Winkens, Ben van Steenkiste, Richard Grol, Job Metsemakers

Implement Sci. 2009;4:6.

Abstract

Background

The use of guidelines in general practice is not optimal. Although evidence-based methods to improve guideline adherence are available, variation in physician adherence to general practice guidelines remains relatively high. The objective for this study is to transfer a quality improvement strategy based on audit, feedback, educational materials, and peer group discussion moderated by local opinion leaders to the field. The research questions are: is the multifaceted strategy implemented on a large scale as planned?; what is the effect on general practitioners' (GPs) test ordering and prescribing behaviour?; and what are the costs of implementing the strategy?

Methods

In order to evaluate the effects, costs and feasibility of this new strategy we plan a multi-centre cluster randomized controlled trial (RCT) with a balanced incomplete block design. Local GP groups in the south of the Netherlands already taking part in pharmacotherapeutic audit meeting groups will be recruited by regional health officers. Approximately 50 groups of GPs will be randomly allocated to two arms. These GPs will be offered two different balanced sets of clinical topics. Each GP within a group will receive comparative feedback on test ordering and prescribing performance. The feedback will be discussed in the group and working agreements will be created after discussion of the guidelines and barriers to change. The data for the feedback will be collected from existing and newly formed databases, both at baseline and after one year.

Discussion

We are not aware of published studies on successes and failures of attempts to transfer to the stakeholders in the field a multifaceted strategy aimed at GPs' test ordering and prescribing behaviour. This pragmatic study will focus on compatibility with existing infrastructure, while permitting a certain degree of adaptation to local needs and routines.

Background

With the ever-growing volume of evidence from medical research, it has become impossible for physicians to remain fully up to date. Reviews and guidelines therefore summarize large quantities of information, making it more easily available to field workers. In the Netherlands, general practitioners (GPs) now have access to more than 80 evidence-based medical guidelines developed by the Dutch College of General Practitioners (NHG). Although general adherence to these guidelines is approximately 70%, the inter-physician variation is large, and adherence to certain aspects of these guidelines proves to be difficult.¹⁻³ Although there may be sensible reasons to deviate from guidelines, such as multi-morbidity in a patient, a physician's level of uncertainty tolerance and patients' preferences, there seems to be room for improvement. The inter-physician variation can be regarded as underdiagnosing or undertreating one group of people and at the same time overdiagnosing and overtreating another group, both leading to inappropriate care.⁴ There is considerable inter-physician variation in general practice with regard to test ordering and prescribing.^{5,6}

Many studies have tried to find evidence for effective implementation strategies to improve quality of care. A multifaceted clustered RCT by Verstappen *et al.* aimed at optimizing GPs' test ordering behaviour by means of local quality improvement collaboratives (LQICs), found a decrease of 8 to 12% in test volumes over a period of six months.⁷ This strategy was tested using six topics for continuing medical education (CME). Other studies have tested several implementation strategies to improve test ordering and prescribing behaviour. Passive dissemination of guidelines or recommendations does not seem to influence test ordering behaviour. Audit and feedback have often been used and showed mostly a modest effect in terms of influencing test ordering or prescribing. The effect of audit and feedback on adherence to desired practice ranged from -10% to +68% (median +16%).⁸⁻¹² In other studies, the introduction of a problem-based test ordering form proved to be a promising tool to improve test ordering.^{7,13-18} Similar effects on volumes of tests ordered as those in the Verstappen study have been found for more or less similar multifaceted implementation strategies.^{19,20} Small group peer review using direct individual feedback seemed to reduce inappropriate prescribing.^{12,21,22} Lagerlov found a 6 to 13% improvement in adherence to guidelines for the prescription of anti-asthmatic drugs and antibiotics for urinary tract infections in an RCT using reflection on guidelines and prescription feedback in small groups.²³

The Cochrane Effective Practice and Organization of Care group (EPOC) systematically reviews studies on implementation strategies to improve quality of care. Their work has generated the general insight that multifaceted strategies are usually more effective than single interventions^{12,24}, although this was not entirely confirmed by an NHS HTA

review by Grimshaw *et al.*¹⁶ The prevailing insight is that the effect of an intervention is larger when tailored strategies are used and when barriers to and facilitators of change are addressed.

Grol has identified in his model of effective implementation six stages in quality of care improvement.²⁵ In the first stage, new research findings, new guidelines, experienced weaknesses, or best practices create an opportunity for quality improvement. In the second stage, after the initial implementation process has been planned, targets for improvement or change are set. Prior to the actual implementation, the performance, target group, and setting are analysed. In the fourth stage, the strategies that are to be used are identified and tested. The implementation is developed, tested, and executed. Finally, the implementation is evaluated and adapted, if necessary.²⁵ The present study will deal with the actual sustainable transfer of a successful implementation strategy to the field. We are not aware of published studies testing this process, or whether effects are sustainable when transferred to the field. Nor are we aware of published studies on the implementation of a large-scale strategy aimed at influencing both the test ordering and prescribing behaviour of GPs simultaneously, using peer review and social influencing in primary care collaboratives.

In the Netherlands, existing networks of pharmacotherapeutic audit meetings (PTAM) can be used to disseminate and implement guidelines on test ordering and prescribing. The goal of setting up these meetings by primary care providers was to improve the quality of their prescribing behaviour.²⁶ The local groups usually consist of six to ten GPs with affiliated community pharmacists.²⁷ During the meetings, they discuss the choice of drugs in the context of a specific illness or disease. In recent decades, this form of CME has gained widespread acceptance amongst GPs and policymakers in the Netherlands. However, these sessions tend to offer little or no room for discussions on test ordering. Because no other system of regular meetings exists, the possible underuse, overuse, and misuse of diagnostic services is not discussed by primary care providers on a regular basis.

The Dutch Institute for the Proper Use of Medicine (DGV) supports and initiates local or regional implementation of quality improvement projects on the use of drugs and supports local PTAM groups by supplying them with information and educational materials.²⁸ Performance levels of PTAM groups are assessed once a year and rated on the basis of four levels, level one being the poorest level of performance and level four the highest. We will use this division into levels as a parameter for pre-randomization stratification.

Participation in PTAM groups by GPs is facilitated by national and regional support organizations for primary care, as well as by the government and through incentives by

insurance companies. Attendance at PTAM meetings is rewarded by accreditation. Currently, approximately 50% of the group meetings reach the desired level of performance described by policymakers.²⁷ To reach this level, groups must at least use feedback on prescribing, create working agreements, discuss barriers to change, and evaluate working agreements. Most groups are stable and remain together for 10 years or more, with members mostly being replaced gradually.²⁷ Because of the nature and stability of these groups, they provide an excellent and safe environment for participants to discuss their own behaviour and barriers to change. We expect this existing system of PTAM groups will ensure sustainability of the implementation itself. Therefore, we plan to use these groups in a large pragmatic trial on the implementation of guidelines, using the strategy previously tested by Verstappen *et al.*⁷ However, we will expand the strategy, using social interaction and external influencing as key approaches for establishing behavioural change, to both test ordering and drug prescribing. In our view, the groups will no longer function merely as a PTAM group, but rather begin acting as LQICs. This trial is expected to show whether the effects found in less pragmatic trials can be confirmed. Aiming at both test ordering and drug prescribing, our combination strategy could lead to an even larger effect because of synergy. We will also evaluate the costs of implementing the strategy on a large scale.

Objectives and research questions

Hypotheses

We expect that the transfer of the strategy of LQICs to stakeholders in the field will be feasible. We hereby hope to create a solid basis for continuation after the end of the study.

We also expect that large-scale implementation, giving attention to both test ordering and prescribing behaviour, will lead to similar changes in performance as those found on test ordering in the trial by Verstappen *et al.*⁷

Successful implementation will be positively related to the level of group performance of the groups included, in terms of level of attendance, number of meetings, drawing up working agreements, discussing barriers to change, and evaluating working agreements.

Objectives

1. To implement the LQIC strategy in the south of the Netherlands, stimulating the relevant parties in the field to take the lead.
2. To determine the critical conditions for effective nationwide implementation.
3. To improve the level of group performance in the participating groups.
4. To reduce undesirable physician variation in test ordering and prescribing; and to reduce underuse or overuse of specific tests and drugs.
5. To examine the costs of large-scale implementation of this strategy, and thus to be able to predict future costs for expansion and maintenance of the strategy.

Research questions

Process

1. Was the strategy implemented as planned?
2. What were the barriers to and facilitators of the implementation of the strategy?
3. Has the level of group performance been improved in the participating groups?

Effect

1. Do the volumes of tests ordered and drugs prescribed change in the preferred direction, as described in the working agreements of the LQICs, compared to baseline?
2. What is the effect of this strategy on GPs' test ordering and prescribing behaviour in terms of interphysician variation and total volumes of tests and prescriptions with respect to specific clinical topics, compared to that among GPs exposed to the same strategy but for other topics?
3. Is any gain in the level of group performance predictive of the effect achieved?

Cost

What are the costs of implementing the strategy?

Methods

Design and ethics

This multi-centre study will use a balanced incomplete block design, consisting of two arms (Figure 2.1). LQICs will be allocated at random to one of these two arms. All LQICs

allocated to arm A will receive the intervention with respect to the clinical topics associated with arm A. All LQICs allocated to arm B will receive the same intervention, but with respect to the topics associated with arm B (Table 2.1). Each arm will have five different CME topics to choose from. Each LQIC will choose three different topics for their discussions, and serve as a control for the other arm. The GPs will not be aware of the topics they are serving as controls for, to avoid the Hawthorne effect.²⁹

The Maastricht Medical Research Ethics Committee has approved this study. All participating GPs will be asked to sign a written informed consent form.

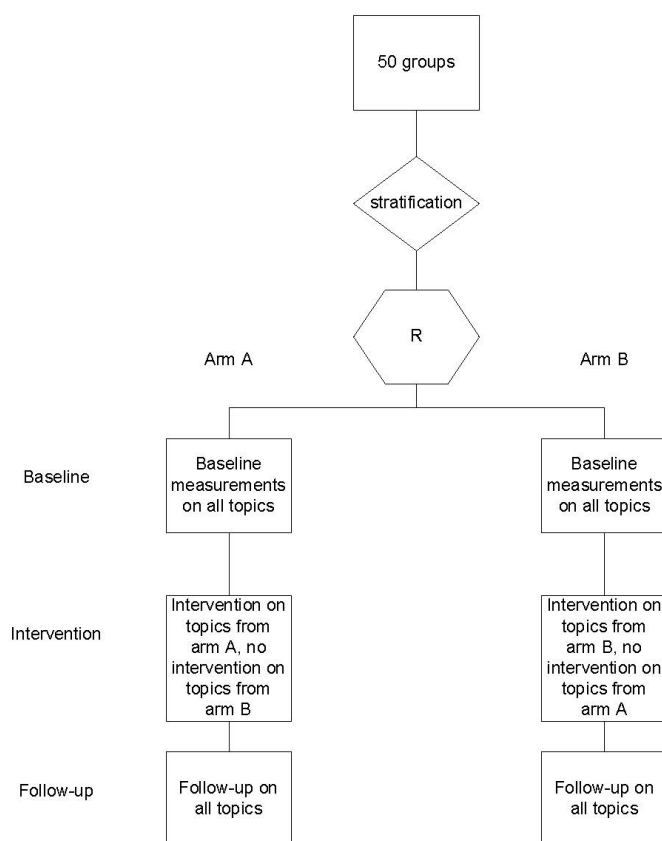


Figure 2.1 Flowchart of randomization and intervention

Table 2.1 Modules and distribution over the research arms. For a complete list of all tests and drugs for the modules [See Appendix 2.2]

	Modules		
	Topic	Examples of tests	Examples of drugs
Arm A	Hypercholesterolaemia	LDL	Statines
	Anaemia	haemoglobin	ferro medication
	Rheumatic complaints	Waller-Rose	NSAIDs
	Urinary tract infections	Urinary cultures	Antibiotics
	Prostate complaints	PSA	α -blockers
Arm B	Type 2 Diabetes Mellitus	HbA1c	Metformin
	Dyspepsia	gastroscopy	proton-pump inhibitors
	Chlamydia infections	chlamydia cultures	Antibiotics
	Thyroid problems	TSH	Levothyroxine
	Perimenopausal complaints	FSH	Estradiol

Population

LQIC groups will be recruited by regional medical coordinators, which are regional health officers or managers often employed by regional hospitals or primary care laboratories. We have identified 24 organizations offering diagnostic facilities in the south of the Netherlands. All organizations will be visited by the researcher and asked to cooperate. Each medical coordinator then will be asked to recruit two to four LQIC groups. They will only be included when all group members consent to participate. The area from which groups can be recruited will be restricted to the three southern provinces of the Netherlands (Limburg, Noord-Brabant, and Zeeland) because these are covered by the insurance companies who provide data for the pharmaceutical database at Maastricht University (UM). A representative with special expertise in and knowledge of diagnostic testing, recruited by the medical coordinator, will attend each LQIC meeting. This representative will receive copies of the feedback forms of all GPs in a LQIC, to enable him or her to prepare the sessions. The representative will act as a moderator during the sessions devoted to diagnostics, after having been trained to do so (see under ‘training’). The medical coordinator will finally also liaise between their diagnostic centre and the research team. Other stakeholders in our strategy include community pharmacists, UM, the DGV, insurance companies, PTAM groups, and individual GPs. Community pharmacists play a major role in PTAMs in the Netherlands, providing expertise and sometimes feedback on prescriptions to the participating GPs. Our intervention will leave the role of the pharmacists more or less unchanged. They provide easily accessible knowledge for GPs, thus breaking down barriers which might be inherent in distance support such as academic detailing. Like the medical coordinator, a pharmacist will function as a moderator in the LQIC. All community pharmacists will receive training prior to the first session, as described above. The

pharmacists will receive copies of the feedback forms of all participating GPs in a group, to enable them to prepare the sessions.

The initiator of this trial is the Department of General Practice of Maastricht University. The design and maintenance of the database on diagnostics and the data gathering process are coordinated by the first author. The Maastricht University Centre for Information and Data Management (MEMIC) will host the diagnostics database, as they already do for the prescriptions database.

Randomization

LQIC groups will be randomized as such (cluster randomisation). The intervention is aimed at these groups. Pre-randomization stratification will be performed on group size and level of group performance using a pre-randomization questionnaire [See Appendix 2.1] prior to the intervention. The levels of group performance are as determined by DGV.²⁸ This level is a known confounder for an effective intervention on medical education among groups of GPs.^{30,31} After stratification, all groups within a stratum will be randomly allocated to either arm A or arm B (Figure 2.1).

Sample size

A sample size calculation is not really possible beforehand, because it is not yet known what working agreements will be created and with respect to what tests or drugs. The specific targets, incorporated in working agreements will probably be based on extreme overuse or underuse of certain tests or drugs by some or all group members. It is possible, for instance, that the group will decide to eliminate a particular obsolete test or drug or create a working agreement to decrease or increase the mean volume of tests ordered or drugs prescribed by 20%, from 35% to 55%.

The sample size calculation used in this trial is as follows: to detect an improvement of 20% in a certain target between groups, assuming an ICC of 0.10,⁵ an alpha of 0.05 and a beta of 0.1 and a mean group size of seven, 44 LQICs would be needed. Anticipating a dropout of 10%, we would need to recruit 50 groups. A population this large would account for approximately 900,000 registered patients.

Intervention

Several theories have been postulated on how change in healthcare can be accomplished, and how effective change strategies can work in implementation of innovation. In cognitive theories, professional behaviour is considered to result from

rational processes and experiences from earlier caseloads. In social interaction theories, change of professional behaviour is thought to be strongly mediated by peers in a group, the strength of inter-individual ties within groups, the existence of opinion leaders, and how much the desired behaviour is consistent with, and fits in, everyday practice. In total quality management theories, the use of systematically gathered data is considered to be crucial to facilitate effective professional development. These data can then be used in plan-do-study-act cycles (PDSA cycles) to provide insight into displayed behaviour and help identify areas where improvement is possible. This leads to the description of targets. These theories may overlap or may be complementary. In implementation science, the use of these theories as a framework is considered obligatory.³² This intervention therefore will be multifaceted and consist of audit, comparative graphical feedback, and small group work with peer review of each other's performance, discussion of barriers to change, reaching agreement on future policy, and testing the agreement. After randomization to arm A or B, each group can choose from the corresponding set of five clinical topics allocated to that arm, to decide which three topics they want to discuss. Two balanced sets of topics, one for each arm, have been defined by the researchers. Each set consists of three major topics, from which the group has to choose two, and two minor topics, one of which has to be chosen. Thus, each LQIC will be asked to complete the entire strategy for three clinical topics of their choice during the intervention period. They are free to schedule extra meetings on topics not included in this trial, but these meetings will not be included in the final analysis. Feedback on the topic under discussion will be sent to the medical coordinator (diagnostic feedback) or local community pharmacists (prescription feedback) two weeks prior to the test ordering or the prescribing session of the LQIC, together with the relevant educational materials (see under 'clinical topics'). The first session, which will last approximately 90 minutes, will address the diagnostic test ordering behaviour of the individual GPs and will have the structure described under 'session structure' (Table 2.2). During this session, the GPs will discuss their diagnostic test ordering patterns and relate them to the guidelines provided. Individual and group working agreements will be created after barriers to change have been discussed. The second session will have the same structure, but the subject for discussion will be physicians' prescribing performance. This session will end by creating group and individual working agreements about preferred medication. Barriers to change from an individual perspective will again have to be discussed. After this first topic has been completed, the cycle will be repeated, for a new topic, as shown in Table 2.3. At the start of this new cycle, the group will reflect on the previous agreements, and revise them if necessary. The working agreements will then be prepared for further dissemination in the practices. Each session will be chaired by a member of the LQIC itself. When test

ordering is discussed, a local representative from the diagnostic centre will be present, while a local community pharmacist will be present when pharmacotherapy is discussed. They will act as moderators, not as chairpersons.

Table 2.2 Session structure.

90 minutes	5 min	Explaining the method / reflection on previous topic
	5 min	Critical look at participants' own feedback
	5 min	Pairwise/group discussion on inter-individual differences
	25 min	Plenary discussion, relating feedback to guidelines
	10 min	Pairwise discussion on barriers to change
	25 min	Plenary discussion on barriers to change, aimed at problem solving
	15 min	Drawing up individual and group working agreements

Table 2.3 Example of a schedule for the intervention.

Topic	GPs	Medical coordinator	Community pharmacist
1. Anaemia	1. Meeting on tests	Moderator	Prepares second session
	2. Meeting on drugs	Prepares third session	moderator
2. Chlamydia infections	3. session on test and drugs (anaemia)	moderator	Present as expert
	1. session on tests (Chlamydia)		

We will test the model and the logistics needed prior to the large-scale implementation. We plan to do this in a small pilot study involving five groups of GPs. This pilot study will run for four months, during which period the participating GP groups will schedule two meetings. Each session will be structured according to the method provided by the researchers. The first session will address test ordering, while the second session will address prescribing. For reasons of efficiency, a set of only three topics will be used for the pilot study. The topics, which have been proposed by the project team members, are anaemia, dyspepsia, and asthma in combination with chronic obstructive pulmonary disease (COPD).

Clinical topics

The set of clinical topics the GP groups can choose from in the main study has been proposed by the authors. After eligible topics were selected and divided over the two trial arms, both arms were balanced in terms of the weight of the topics. The weight depends on the prevalence of the underlying disease and whether the emphasis within the topic is on either the volume of tests ordered or the drugs prescribed. The two sets of topics are also balanced in terms of subjects, emphasising diagnostic or prescribing features (Table 2.1). Each topic includes a number of tests [See Appendix 2.2] and drugs

[See Appendix 2.3] predefined by the project group. For the purpose of feedback and education, these include both well-accepted and commonly not accepted (or even obsolete) tests and drugs. Educational materials on each topic will be based on the relevant national primary care guidelines from the Dutch College of General Practitioners, guidelines from the Dutch Institute for Healthcare Improvement (CBO), and international guidelines if applicable. Guidelines will be read and 'condensed' into short versions called modules. These modules have been drafted by one of the authors (JT) and then commented on by an expert on the topic. Indicative prices for each test and drug will be provided, as well as a short description of its values and drawbacks, given the indication. Each module will consist of a maximum of six easily searchable pages.

Extraction of feedback data

Data on test ordering behaviour will be extracted by the regional coordinators from the various databases available at the participating hospital laboratories or primary care diagnostic centres. Each centre will receive a data fact sheet prescribing the required data format. This format is based on rational criteria for laboratory test registration to facilitate the integration of the individual databases into one main database. All datasets on diagnostics will be combined into one newly formed database, to be maintained by UM (Figure 2.2). Data on prescribing behaviour will be extracted from the databases of health insurers and collected into one database, as has already been done at our institute. This database consists of the reimbursements for prescriptions written by GPs for approximately 5.5 million persons in the south of the Netherlands. Feedback will then be derived from the two main databases and processed into graphical comparative feedback reports. Data will be presented as the volume of tests ordered (*e.g.*, haemoglobin) or defined daily dosages (DDDs) prescribed per 1000 patients per six months. Participating GPs will receive their data as clustered column charts, each cluster presenting the data for the individual GP, the practice in which he or she works, the small group and the wider region. An example of such a graphical feedback report is shown in Figure 2.3.

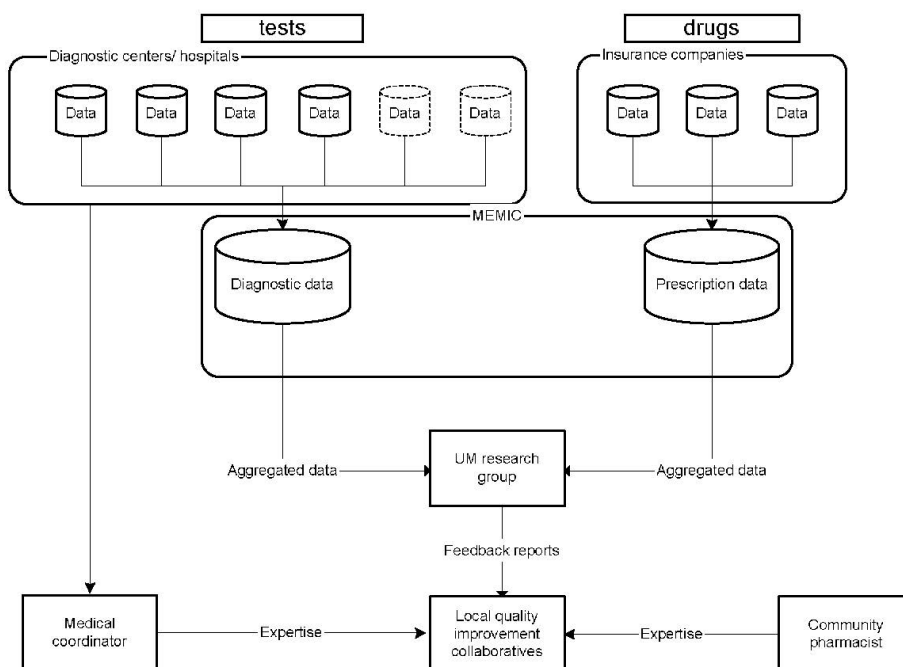


Figure 2.2 Data and Knowledge flowchart.

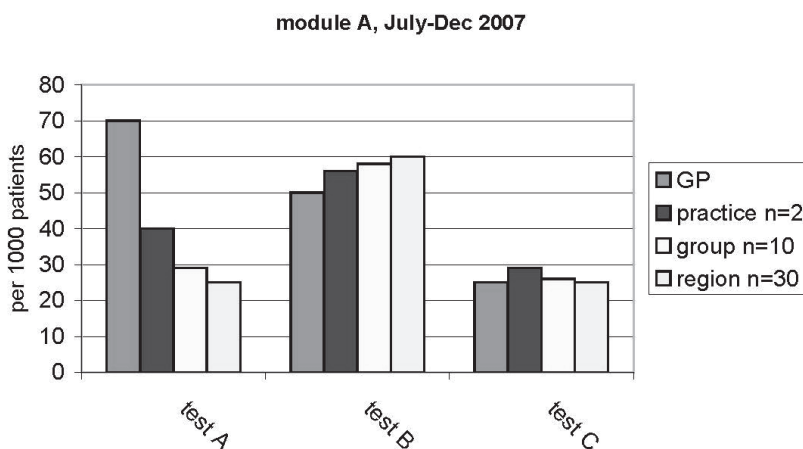


Figure 2.3 Example of a graphical comparative feedback sheet.

LQIC meeting structure

Each meeting will be structured according to a uniform schedule. After participants have studied the feedback in pairs or as a group, they will discuss it. Subsequently, the guidelines as described in the educational materials will be discussed in relation to the feedback. A plan will then be formulated to improve the test ordering or prescribing behaviour. The next step will involve addressing and discussing all the barriers to change at individual and group levels. Finally, working agreements will be created regarding test ordering and prescribing behaviour for the tests and drugs discussed. A standardized group meeting structure card will be provided to each LQIC, showing the structure as recommended by the researchers. However, groups will be free to adapt the structure to their own preferences or needs.

Training

The participating medical coordinators and local community pharmacists will be trained prior to the first LQIC session, in a two- to three-hour standardized training session covering three main subjects. The first subject will involve an explanation of the structure of the trial, the objectives, the development of the outlines, the source of the feedback data, and the process of data gathering. The second subject will be the preferred structure for the meetings, the tools that are to be used, how to read the feedback reports and relate the feedback to the guidelines. The final subject of the training session will be how to act as a moderator instead of a chair during a meeting. Training sessions will partially be constructed like a LQIC meeting, with the trainees acting as GPs and the trainer as the moderator.

Variables

Outcome measures

Process evaluation:

1. The performance level of the small group collaborative.
2. Process data such as attendance at meetings, actually creating working agreements, following the LQIC strategy, the number of groups that complete participation, and the number of regions actually participating.

Effect evaluation:

1. The volumes of particular tests ordered and particular drugs prescribed for which the group has agreed that change, either decrease or increase, would be necessary.
2. The total volumes of tests ordered and drugs prescribed by the participating GPs for the clinical topics chosen.
3. The inter-physician variation in test ordering and prescribing behaviour for the clinical topics chosen.

Cost evaluation

The costs of implementing the LQIC strategy.

Explanatory variables

We will monitor data that are known to moderate quality assurance strategies. Therefore the following data will be gathered prior to the intervention: group size, age and gender of GPs, type of practice, number of patients registered with the practice, number of patients a GP is accountable for, number of working hours a week per GP, number of working hours a week for the group practice as a whole, distance to the hospital / diagnostic centre, responsibility for training GP trainees, total number of GPs collaborating in the practice, whether a GP admits sales representatives from pharmaceutical firms and if so how often, involvement in developing national guidelines, and GPs field(s) of special expertise.

All medical coordinators will be asked if problem-based test ordering forms are used in their region and to send us a copy of such a form.

Measurements

Prior to randomization, the chair of the group will be asked to fill out a short pre-randomization form, with which we will be able to determine the number of GPs in the group and be able to assess the level of group performance [See Appendix 2.1]. Data on test ordering and prescribing behavior will be extracted from the existing databases at baseline (t=0) and t=6 months, t=12 months and t=18 months. The dataset obtained at t=0 and the final set will be used for a before-after analysis. A new questionnaire will be sent to the chair, assessing the level of group performance after the intervention. After each meeting, the chair will be asked to fill out a form with questions about the structure of the session, whether working agreements were created, whether barriers to change were discussed and if so what the nature of these barriers was, what

educational materials were used, and the group members' experiences with the strategy.

The process of implementing the strategy in the south of the Netherlands will be monitored. Participants will be questioned about their experiences with the strategy. Participating GPs will be asked to report their experiences with the strategy, and to provide us with the necessary details on the sessions they have attended. After each session, the targets set by each group will be recorded.

Analysis

Analysis will be based on the intention-to-treat principle. Data on GPs lost to follow-up will be extracted from the various databases if possible.

We will analyse covariance using test and drug volumes during the intervention period as the dependent variables, and the baseline data and the explanatory variables as independent variables. The analysis will be repeated using proportions stemming from prescription performance indicators, if available. The unit of allocation to the trial is the LQIC. In larger practices with more than one GP, not all volumes of tests ordered and drugs prescribed will be traceable to an individual GP. In these cases, the unit of analysis will be the practice as a whole. Because of this unit of analysis error, the data will be analysed using multilevel modelling. Data on drugs and tests will be clustered to individual GPs at level one, the practice at level two, the LQIC at level three and the region at level four.

The nature of this study makes it difficult to blind the participants, except for the tests or drugs serving as controls in the other arm. The data analyst will be blinded for the allocation result. Costs of the intervention will be calculated. A cost-effectiveness analysis will be based on these figures. We will use cost minimization analysis from a societal perspective, assuming that the strategy will reduce redundant testing and prescribing. If there are signs of improvement of care (higher scores on the performance indicators), the impact on health may be estimated by modelling the future gains and benefits. Data include costs of coordinating the strategy by the regional contact group, of preparing feedback reports and of chairing the GP groups. The costs of the entire test ordering strategy by Verstappen were €554.70 per GP per six months (three meetings). The major part of the cost of this strategy consisted of opportunity costs, viz the costs of the GPs' time spent attending the session. Because GPs were already attending these meetings and were financially compensated, it seems fair to ignore the opportunity costs. This results in costs of the test ordering strategy of

€92.70. The gains obtained by improving test ordering behaviour were €301.00 per GP per six months. Introducing the test ordering strategy would save €208.30 (92.70 to 301.00) per GP per six months.³³ Because prescribing costs are higher, the cost reductions gained by reducing superfluous prescribing should also be higher.

Time schedule

The intervention period will start in september 2007 and run through the spring of 2009. Process evaluation will start when all groups are included. During the intervention, new datasets will be obtained every six months in order to keep the databases up-to-date for future use in new sessions.

Discussion

To our knowledge, few studies have been published on the transfer of effective implementation strategies to the field. Our strategy has proved to be effective in an earlier trial on test ordering by GPs in the Netherlands. However, because this strategy was disseminated and controlled by academics, it remains unclear how large its effect will be when transferred to the field. We set up a pragmatic design in order to test this final step in implementation research, giving the diagnostic centres a leading role and leaving GPs much room to adapt and to internalise the strategy. The project team will act as facilitators to these centres, the pharmacists involved, and the LQICs. The strategy is targeted first on test ordering and second on prescribing, which is the natural order followed by GPs when consulted by a patient.

Our strategy is based upon several theories on effective behaviour change and on effective implementation. These theories can be identified at several levels of organisation in our trial. At the level of diagnostic centres and the LQICs, we expect the innovators and early adopters to join the trial, which refers to Roger's innovation-diffusion theory.³⁴ Within groups we expect to see change according to theories such as Ajzen's theory of planned behaviour and the PDSA-cycles.^{25,35} During a meeting, we expect to see the preparation for change based on performance data and actual actions towards change. When new data will be provided to the groups, we expect reflection on the goals previously set. The theory of planned behaviour states that individuals are willing to show change in behaviour dependent on the perceived control over the behaviour itself, the attitude of the individual to the desired behaviour, and the perceived social norms. By providing graphical comparative feedback, we target at these perceived social norms. Comparative feedback sets the norm for a group, and

through the phenomenon that one does not like to be an outlier we expect regression to the mean with regard to the inter-physician variation. The moderator who is also an expert on the subject under discussion is expected to act as opinion leader. Furthermore, even a GP from within the group itself can act as a local opinion leader and thus influence the rest of the group.

The existing PTAM group structure in the Netherland is widespread and functions reasonably well. However the need to improve the functioning of these groups is clearly present. Our strategy is known to improve test ordering behaviour of GPs, but is not used widely. Transferring PTAM groups into LQICs gives us the opportunity to add a discussion on test ordering behaviour to existing discussions on prescribing by GPs in PTAMs. The constitution of LQICs therefore is not 'old wine in new bottles' but a completely new approach within existing structures.

Several methodological challenges were encountered when we designed this trial. First, individual GPs are known to choose topics for CME in which they already show good performance.³⁶ This might result in a 'ceiling effect', meaning that little or no improvement in test ordering or prescribing behaviour would be possible. However, because the LQIC will have to reach consensus on the clinical topics they choose, the risk of such a ceiling effect is probably not very great.

Second, using an implementation strategy on ten different clinical topics from which GPs can choose introduces challenges to the sample size calculation. We chose to leave the LQICs some freedom of choice with regard to the topics. All clinical topics are well-described in the national guidelines for each topic. We will use a set of 204 tests and drugs to generate feedback [See appendix 2.2, and Appendix 2.3]. Because we do not know what agreements local groups will come to, and do not know beforehand what the desired direction for change is, sample size calculation is very difficult. Because we intend to improve care by using the national guidelines, we do not expect to decrease quality of care by this study. However, it is impossible to predict if change will be towards better care.

Third, the databases we use are complex, as are the origins of the data. Most local databases on diagnostics used in this trial are intended primarily for billing purposes. This might create problems when extracting data, reading it into a central database and translating it into feedback. In the past, no significant problems were encountered when extracting data from laboratories (personal communication by Verstappen). Data on tests not performed within a laboratory (*e.g.*, gastroscopy and X-rays), however, are

often stored in separate databases and might not be linked to a GP but to a patient. In these cases, tracing the GP who ordered the test is possible but will require an extra effort from the diagnostic centres. It is possible that recruiting groups, supplying a moderator for the sessions and implementing this time-consuming data extraction process might prove to be too much of an effort for the centres. Most contact persons of the centres, however, have indicated that they were most willing to cooperate and were aware of the opportunities offered by this trial.

Fourth, the database on prescriptions consists of data from the large insurance companies in the south of the Netherlands. Using these records as a basis for feedback might create several problems. Although most inhabitants of the southern provinces are insured by one of these companies, prescription data for patients insured with other companies will not be included in our database on prescriptions. This problem might be solved in the future by adding more insurance companies to the database. Another potential problem may be that recording errors are likely to be present in the databases. Desk staff at local pharmacies often links a prescription to one of the GPs in a practice, and often almost all prescriptions for a practice are thus linked to one physician, even when several physicians collaborate in the practice. This creates an inaccuracy in the database, but only for GPs sharing an office. To solve this problem, we will also aggregate to an extra level in these cases, *viz* the subgroup of GPs sharing an office, thus creating a fourth column on the graphical feedback sheet. The last problem we expect to encounter using a large database on prescription is that we do not know the indication for which medication was prescribed; these indications are not known to pharmacists and thus are not stored in any database. This makes it impossible to trace a prescription back to a specific disease. By building a similar database on tests ordered by GPs, we will encounter this problem as well. We do not however expect this to be a problem because we will use graphical comparative feedback. All data from all participating GPs are expected to be equally affected by this problem and thus the feedback will be comparable.

Fifth, the tests of the diabetes and hypercholesterolemia topics partly overlap. We accepted this, however, because in diabetes, the glucose and HbA1c items are the primary indicators, whereas cholesterol, LDL, HDL, and the ratio are the primary indicators in the hypercholesterolemia topic.

Sixth, we have to be aware of the Hawthorne effect. As discussed above, we chose to use a balanced incomplete block design to overcome this problem. The complexity of the strategy, however, would make it more attractive to use a different design and start

the trial in phases. This would mean that different regions would enrol in the strategy successively, so we could learn from the early regions what the weaknesses of our design were and what we would have to alter. This would create an opportunity to ameliorate the strategy with each new phase. To this end, a dynamic wait-listed design could have been more appropriate and beneficial.³⁷ Conversely, we would then have had to wait after completing enrolment and intervening in one region for new data to be added to the database. The delay would be six months after each region. This left us with no choice but to start with the entire population in the same period. In this situation, we considered the balanced incomplete block design to be most useful.

Finally, GPs and moderators cannot be sufficiently blinded in our present design. However, because GPs do not know what clinical topics are available in the arm they are not allocated to, we do achieve some level of blinding.

Notwithstanding these methodological challenges, there are also opportunities in the Dutch healthcare system that makes it attractive to start this trial now. First, the strategy we intend to use fits in well with the new Dutch healthcare system. After the recent reform, healthcare has turned into a competitive business, in which financial profits and market shares may influence decision-making. Our study might create profiling opportunities for centres, which might bind GPs more tightly to them, and thus might be a way for the centres to improve their chances in this market. Finally, diagnostic centres are under increasing pressure from various parties in the healthcare system to provide feedback to GPs. GPs want feedback to monitor and claim results when treating chronically ill patients (*e.g.*, diabetics), while insurance companies want laboratories to provide feedback in order to influence test ordering behaviour, and primary care organizations need GPs' performance data for various reasons, such as certification.

A preliminary investigation identified 24 eligible diagnostic centres in hospitals, all of which provide diagnostic facilities to GPs. All were contacted and appointments for personal visits were made. Two centres were not interested in participating, and were therefore not visited. Two centres expressed an interest but faced major strategic challenges and found no time to participate. The remaining 20 centres all agreed to participate. One of the participating centres will not be asked to recruit groups, however, because it is not linked to a region like the other centres, which means that knowledge of local PTAM group structures is lacking. This centre will, however, participate in the large database on diagnostics.

In the south of the Netherlands, health insurance is offered predominantly by two companies, which insure the majority of the inhabitants of these provinces. These insurance companies regularly send updated reports on prescription data to UM. These files are and will be combined into one research database on prescriptions, maintained by MEMIC. Because the recent health care reform in the Netherlands, insurance companies have been given a large role in guarding and improving the quality and continuity of care. They promote the existence of PTAM groups in order to improve the quality of care, giving financial incentives to GPs for attending such group meetings. In some cases, extra incentives are given if working agreements are created and adhered to. However, the insurers are unable to evaluate the quality of the group work. The strategy evaluated in the proposed study should provide them with a tool to ensure high quality group meetings.

References

1. Braspenning JCC, Schellevis FG, Grol R. Kwaliteit huisartsenzorg belicht [Quality in primary care revised]. Nivel; 2004.
2. Burgers JS, Grol RPTM, Zaat JOM, Spies TH, van der Bij AK, Mekkink HGA. Characteristics of effective clinical guidelines for general practice. *Br J Gen Pract* 2003;53:15-19.
3. Muijers PE, Grol RP, Sijbrandij J, Janknegt R, Knottnerus JA. Differences in prescribing between GPs. Impact of the cooperation with pharmacists and impact of visits from pharmaceutical industry representatives. *Fam Pract* 2005;22:624-630.
4. Harteloh PPM. kwaliteit van zorg: van zorginhoudelijke benadering naar bedrijfskundige aanpak [Quality of care: from a care standpoint towards a business management standpoint]. 4 edn. Maarssen: Elsevier/ De Tijdstroom; 2001.
5. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21:387-395.
6. Martens JD, van-der-Weijden T, Severens JL, de-Clercq PA, de-Bruyn DP, Kester AD, Winkens RA. The effect of computer reminders on GPs' prescribing behaviour: A cluster-randomised trial. *Int J Med Inform* 2007;76:S403-416.
7. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermesen J, Grimshaw J, Grol RPTM. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289:2407-2412.
8. Baker R, Falconer Smith J, Lambert PC. Randomized controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scan J Prim Health Care* 2003;21:219-223.
9. Sondergaard J, Andersen M, Stovring H, Kragstrup J. Mailed prescriber feedback in addition to a clinical guideline has no impact: a randomised, controlled trial. *Scand J Prim Health Care* 2003;21:47-51.
10. Martens JD, Winkens RA, van der Weijden T, de Bruyn D, Severens JL. Does a joint development and dissemination of multidisciplinary guidelines improve prescribing behaviour: a pre/post study with concurrent control group and a randomised trial. *BMC Health Serv Res* 2006;6:145.
11. Winkens RA, Pop P, Bugter-Maessen AM, Grol RP, Kester AD, Beusmans GH, Knottnerus JA. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet* 1995;345:498-502.
12. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;CD000259.
13. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet* 2006;367:1990-1996.
14. Verstappen WHJM, van der Weijden T, Dubois WI, Smeele I, Hermesen J, Tan FES, Grol RPTM. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med* 2004;2:569-575.
15. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to Improve Physicians' Use of Diagnostic Tests: A New Conceptual Framework. *JAMA* 1998;280:2020-2027.
16. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:iii-iv:1-72.
17. Grol RPTM, grimshaw JM: From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;362:1225-1230.
18. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. *Med Care* 1993;31:784-794.
19. Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, Soutter J, Wilsdon J, Matowe L, Needham G, Gilbert F, Bond S. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet* 2001;357:1406-1409.
20. Bunting PS, Van Walraven C. Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem* 2004;50:321-326.

21. Grimshaw JMMP, Shirran LMAM, Thomas RB, Mowatt GMAMBA, Fraser CMA, Bero LP, Grilli RMD, Harvey EB, Oxman AMDA, O'Brien MAM. Changing Provider Behavior: An Overview of Systematic Reviews of Interventions. *Med Care* 2001;39:II-45.
22. Arnold SR, Straus SE: Interventions to improve antibiotic prescribing practices in ambulatory care. *Cochrane Database Syst Rev* 2005:CD003539.
23. Lagerlov P, Loeb M, Andrew M, Hjortdahl P. Improving doctors' prescribing behaviour through reflection on guidelines and prescription feedback: a randomised controlled study. *Qual Saf Health Care* 2000;9:159-165.
24. Freemantle N, Harvey EL, Wolf F, Grimshaw JM, Grilli R, Bero LA. Printed educational materials: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2000:CD000172.
25. Grol R, Wensing M, Eccles M. improving patient care. The implementation of change in clinical practice. Elsevier Limited; 2005.
26. Paes AH, Cornips PMJM. Ontstaan en verloop van een farmacotherapie-overleg [the emergence and development of pharmacotherapeutic meetings]. *Pharm Weekbl* 1987;122:225-230.
27. Eimers M, de Groot J. FTO peiling 2005,kwaliteit van farmacotherapieoverleg in Nederland in beeld [PTAM audit 2005, quality of pharmacotherapeutical audit meetings in the Netherlands made visible]. DGV Nederlands instituut voor verantwoord medicijngebruik [DGV the Dutch Institute for the Proper Use of Medicine] 2006.
28. DGV Nederlands instituut voor verantwoord medicijngebruik [DGV the Dutch Institute for the Proper Use of Medicine] [www.medicijngebruik.nl]
29. Verstappen WHJM, van der Weijden T, ter Riet G, Grimshaw J, Winkens R, Grol RPTM: Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. *J Clin Epidemiol* 2004;57:1119-1123.
30. Eijk MECv, Paes AHP, Porsius A, Avorn J, Boer Ad. Pre-randomization decisions and group stratification in a randomized controlled trial to improve prescribing. *Pharm World Sci* 2004;V26:227-231.
31. Eimers M, van der Aalst A, Pelzer B, Teichert M, de Wit H. Leidt een goed FTO tot beter voorschrijven? [Does an effective PTAM lead to more appropriate prescribing?] (Dutch). *Huisarts en Wetenschap* 2008; 51:340-345.
32. Grol RPTM, Bosch MC, Hulscher MEJL, Eccles MP, Wensing M. Planning and Studying Improvement in Patient Care: The Use of Theoretical Perspectives. *The Milbank Quarterly* 2007;85:93-138.
33. Verstappen WHJM, van Merode F, Grimshaw J, Dubois WI, Grol RPTM, van Der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care* 2004;16:391-398.
34. Rogers EM. Diffusion of innovations. 4th edn. New York The Free Press; 1995.
35. Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 1991; 50:179-211.
36. Hobma S: Directed self-learning as approach to continuing professional development. Maastricht University, General practice; 2005.
37. Brown CH, Wyman PA, Guo J, Pena J. Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide. *Clin Trials* 2006;3:259-271.

Appendix 2.1

All questions below concern the period from January till December 2007

1. How many meetings were scheduled for your PTAM group in 2007?

..... meetings (when 0 meetings were scheduled, this was the last question)

2. On average, how much time was spent on each meeting?

.....hoursminutes

3. Who usually prepared the meetings in 2007?

Choose the answer that is most appropriate

- ☐ general practitioner and pharmacist together
- ☐ general practitioner or pharmacist
- ☐ general practitioner
- ☐ pharmacist
- ☐ hospital deployed medical specialist
- ☐ no one
- ☐ others, i.e.

4. Did these people discuss the preparation of the PTAM during a pre-PTAM meeting discussion?

- ☐ no, never
- ☐ yes, but for less than 50% of all PTAM meetings
- ☐ yes, for more than 50% of all PTAM meetings, but not for all
- ☐ yes, always

5. Did you use prescribing feedback or benchmark data during PTAM meetings?

- ☐ No, never
- ☐ yes, but in less than 50% of all PTAM meetings
- ☐ yes, in more than 50% of all PTAM meetings, but not in all
- ☐ yes, always

6. Did the group reach agreement on preferred drugs during the PTAM meetings?

- ☐ no, never (if so, this was the last question)
- ☐ yes, on one clinical topic
- ☐ yes, on two or more clinical topics
- ☐ yes, always

7. Were agreements on the prescribing of drugs recorded in a meeting report?

- ☐ no, never
- ☐ yes, some but not all agreements
- ☐ yes, all agreements were recorded in a report

8. Were targets set by the group?

- ☐ no, never
- ☐ yes, but in less than 50% of all PTAM meetings
- ☐ yes, in more than 50% of all PTAM meetings, but not in all
- ☐ yes, always

9. Were the effects of the agreements on prescribing behaviour evaluated in another PTAM meeting?

- ☐ yes
- ☐ no (if no, this was your last question)

10. Were prescribing feedback or benchmark data used in this meeting

- ☐ yes
- ☐ no

Appendix 2.2

module	test	label on feedback form
diabetes type 2	glucose random	glucose f/ nf
	glucose non-fasting	glucose f/ nf
	glucose fasting	glucose f/ nf
	HbA1c	HbA1c
	total cholesterol	T-cholesterol
	HDL	HDL
	LDL	LDL
	triglycerides	triglyc
	creatinine	creat
	albumin concentration, urine	albumin urine
dyspepsia	13C-ureum breath test	breath test
	14C-ureum breath test	breath test
	faeces test HP	faeces test HP
	faeces test HP (PCR)	feces test HP
	helicobacter pylori, IgG	antibodies HP
	helicobacter pylori, any antibodies	antibodies HP
	helicobacter pylori, IgM	antibodies HP
	gastroscopy	gastroscopy
anaemia	X-ray, stomach	X-ray, stomach
	Hb	Hb/ Ht
	Ht	Hb/ Ht
	anaemia diagnostic cascade	anaemia cascade
	ferritin	ferritin
	bilirubin	bilirubin
	folic acid	folic acid
	vitamin B12	vit B12
	reticulocytes	reticulocytes
	serum iron	serum iron
	LDH	LDH
	total iron binding capacity	transferrin/TIBC
chlamydia	transferrin	transferrin/TIBC
	chlamydia DNA test, urine (PCR)	chlamydia urine
	chlamydia DNA test, endocervical (PCR)	chlamydia endocervical/vaginal
	chlamydia DNA test, vaginal swab (PCR)	chlamydia endocervical/vaginal
	antibodies chlamydia trachomatis, IgA	antibodies CT
	antibodies chlamydia trachomatis, IgG	antibodies CT
	antibodies chlamydia trachomatis, any	antibodies CT
	antibodies chlamydia trachomatis, IgM	antibodies CT

module	test	label on feedback form
prostate complaints	creatinine	creat
	PSA	PSA
	PSA complex	PSA
	free/total PSA ratio	F/T-PSA
rheumatic complaints	Waler-Rose	(semi-)qualitative rheumatic disease markers
	Latex test	(semi-)qualitative rheumatic disease markers
	IgM rheumatic disease markers	IgM rheumatic disease markers
	anti-cyclic citrulline peptide antibodies	anti-CCP
	ESR	ESR
	CRP	CRP
	X-ray wrist	X-hand/foot/wrist
	X-ray MTP	X-hand/foot/wrist
thyroid disease	antistreptolysin antibodies titer	AST
	TSH	TSH
	thyroid diagnostics (cascade)	thyroid cascade
	FT4	FT4
	T3	T3
	TSH-receptor stimulating antibodies	TSI
	thyroid peroxidase (TPO) antibodies	TPO-Ab
	ESR	ESR
	leucocytes	luco diff
	leucocyte differentiation	luco diff
	thyroid ultrasound	US thyroid
UTI	urine culture and susceptibility testing	Urine culture
dyslipidaemia	total cholesterol	T-cholesterol
	HDL	HDL
	LDL	LDL
	triglycerides	triglycerides
	total cholesterol/HDL ratio	ratio T-chol/HDL
	homocystein	homocystein
perimenopausal complaints	follicle-stimulating hormone (FSH)	FSH
	luteinizing hormone (LH)	LH
	estradiol	estradiol

Appendix 2.3

Module	Drug	ATC-code	Label on feedback form
diabetes type 2	metformin	A10BA02	metformin
	glicazide	A10BB09	glicazide
	glimepiride	A10BB12	glimepiride
	tolbutamide	A10BB03	tolbutamide
	glibenclamide	A10BB01	glibenclamide
	rosiglitazone	A10BG02	rosiglitazone
	pioglitazone	A10BG03	pioglitazone
	repaglinide	A10BX02	other oral blood glucose lowering drugs
	acarbose	A10BF01	other oral blood glucose lowering drugs
	glimepiride/rosiglitazone	A10BD04	combinations, oral
	metformin/glibenclamide	A10BD02	combinations, oral
	metformin/rosiglitazone	A10BD03	combinations, oral
	insulin, fast-acting	A10AB	insulin, fast-acting
	insulin, intermediate-acting	A10AC	Insulin, intermediate-acting
	insulin, intermediate-acting combined with long-acting	A10AD	insulin mix
	insulin, long-acting	A10AE	insulin, long-acting
dyspepsia	magnesium compounds	A02AA	antacids
	auminium compounds	A02AB	antacids
	combinations and complexes of al- ca- en mg- compounds	A02AD	antacids
	antacids with sodium bicarbonates	A02AH	antacids
	famotidine	A02BA03	other H2-antagonists
	ranitidine	A02BA02	ranitidine
	cimetidine	A02BA01	other H2-antagonists
	nizatidine	A02BA04	other H2-antagonists
	omeprazole	A02BC01	omeprazole
	pantoprazole	A02BC02	pantoprazole
	lansoprazole	A02BC03	lansoprazole
	rabeprazole	A02BC04	rabeprazole
	esomeprazole	A02BC05	esomeprazole
	misoprostol	A02BB01	misoprostol
	Pantopac ®	A02BD04	Pantopac ®
anaemia	ferrous sulfate	B03AA07	ferrous sulfate
	ferrous fumarate	B03AA02	ferrous fumarate
	ferrous gluconate	B03AA03	ferrous gluconate
	ferrous chloride	B03AA05	ferrous chloride
	cyanocobalamin	B03BA01	vitamin B12
	hydroxocobalamine	B03BA03	vitamin B12
	folic acid	B03BB01	folic acid
chlamydia	Doxycycline	J01AA02	doxycycline
	azithromycin	J01FA10	azithromycin
	amoxicillin	J01CA04	amoxicillin
	erythromycin	J01FA01	erythromycin

Module	Drug	ATC-code	Label on feedback form
prostate complaints	alfuzosin	G04CA01	alfuzosin
	tamsulosin	G04CA02	tamsulosin
	doxazosin	C02CA04	other alpha-adrenoreceptor antagonists
	terazosin	G04CA03	other alpha-adrenoreceptor antagonists
	prazosin	C02CA01	other alpha-adrenoreceptor antagonists
	finasteride	G04CB01	finasteride
	dutasteride	G04CB02	dutasteride
rheumatic complaints	phenylbutazone	M01AA01	other NSAIDs
	indometacin	M01AB01	indomethacine
	diclofenac	M01AB05	diclofenac
	sulindac	M01AB02	other NSAIDs
	aceclofenac	M01AB16	other NSAIDs
	meloxicam	M01AC06	preferential cox-2 inhibitors
	piroxicam	M01AC01	other NSAIDs
	tenoxicam	M01AC02	other NSAIDs
	ibuprofen	M01AE01	ibuprofen
	naproxen	M01AE02	naproxen
	ketoprofen	M01AE03	other NSAIDs
	flurbiprofen	M01AE09	other NSAIDs
	tiaprofenic acid	M01AE11	other NSAIDs
	dexibuprofen	M01AE14	other NSAIDs
	dexketoprofen	M01AE17	other NSAIDs
	tolfenamic acid	M01AG02	other NSAIDs
	azapropazone	M01AX04	other NSAIDs
	diflusal	N02BA11	other NSAIDs
	metamizole sodium	N02BB02	other NSAIDs
	celecoxib	M01AH01	coxibs
	etoricoxib	M01AH05	coxibs
	nabumetone	M01AX01	preferential cox-2 inhibitors
	diclofenac + misoprostol	M01AB55	diclofenac+ misoprostol
	sulfasalazine (text no feedback)	A07EC01	DMARDs
	methotrexate (text no feedback)	L01BA01	DMARDs
	leflunomide (text no feedback)	L04AA13	DMARDs
	etanercept (text no feedback)	L04AA11	DMARDs
	infliximab (text no feedback)	L04AA12	DMARDs
	gold preparations (text no feedback)	M01CB	DMARDs
	penicillamine (text no feedback)	M01CC01	DMARDs
	chloroquine (text no feedback)	P01BA01	DMARDs
	azathioprine (text no feedback)	L04AX01	DMARDs
	ciclosporin (text no feedback)	L04AD01	DMARDs
thyroid disease	levothyroxine sodium	H03AA01	levothyroxine
	liothyronine sodium	H03AA02	liothyronine
	thiamazole	H03BB02	thiamazole

Module	Drug	ATC-code	Label on feedback form
UTI	nitrofurantoin	J01XE01	nitrofurantoin
	trimethoprim	J01EA01	trimethoprim
	fosfomycin	J01XX01	fosfomycin
	amoxicillin and enzyme inhibitor	J01CR02	amoxicillin/enzyme inhibitor
	sulfamethoxazole and trimethoprim	J01EE01	co-trimoxazole
	norfloxacin	J01MA06	norfloxacin
	ciprofloxacin	J01MA02	ciprofloxacin
dyslipidaemia	simvastatin	C10AA01	simvastatin
	pravastatin	C10AA03	pravastatin
	atorvastatin	C10AA05	atorvastatin
	fluvastatin	C10AA04	fluvastatin
	rosuvastatin	C10AA07	rosuvastatin
	bezafibrate	C10AB02	fibrates
	gemfibrozil	C10AB04	fibrates
	ciprofibrate	C10AB08	fibrates
	acipimox	C10AD06	nicotinic acid (-analogue)
	ezetimibe	C10AX09	ezetimibe
	xantinol nicotinate	C04AD02	nicotinic acid (analogue)
	nicotinic acid	C10AD02	nicotinic acid (analogue)
	colestyramine	C10AC01	colestyramine
	ezetimibe / simvastatin	C10AX	combinations
perimenopausal complaints	estradiol	G03CA03	estrogen monotherapy
	estriol	G03CA04	estrogen monotherapy
	ethinylestradiol	G03CA01	estrogen monotherapy
	conjugated estrogens	G03CA57	estrogen monotherapy
	prostagens and estrogens, fixed comb.	G03AA	oral contraceptives, >50 yr
	prostagens and estrogens, fixed comb.	G03AB	oral contraceptives, >50 yr
	prostagens and estrogens, fixed comb.	G03AC	oral contraceptives, >50 yr
	prostagens and estrogens, fixed comb.	G03BB01	oral contraceptives, >50 yr
	estrogen/cyproterone (Climene '28' ®)	G03HB01	combinations with estrogens
	estrogen/drospirenone (Angeliq ®)	G03FA17	combinations with estrogens
	estrogen/dydrogesterone, sequential (Femoston®)	G03FB08	combinations with estrogens
	estrogen/dydrogesterone, fixed (Femoston continu ®)	G03FA14	combinations with estrogens
	estrogen/norethisterone, sequential (Trisequens ®)	G03FB05	combinations with estrogens
	estrogen/norethisterone, fixed (Activelle ®, Kliogest ®)	G03FA01	combinations with estrogens
	estrogen/norethisterone, patch (Estracomb TTS ®)	G03FB05	combinations with estrogens
	clonidine	C02AC01	clonidine
	tibolone	G03CX01	tibolone



Chapter 3

The challenge of transferring an implementation strategy from academia to the field

A process evaluation of local quality improvement collaboratives in Dutch primary care using the normalization process theory

Jasper Trietsch, Ben van Steenkiste, Sjoerd Hobma, Arnoud Frericks, Richard Grol, Job Metsemakers, Trudy van der Weijden

J Eval Clin Pract 2014;20:1162-1171

Abstract

Introduction

A quality improvement strategy consisting of comparative feedback and peer review embedded in available local quality improvement collaboratives proved to be effective in changing the test-ordering behaviour of general practitioners. However, implementing this strategy was problematic. We aimed for large-scale implementation of an adapted strategy covering both test ordering and prescribing performance. Since we failed to achieve large-scale implementation the aim of this study was to describe and analyze the challenges of the transferring process.

Methods

In a qualitative study 19 regional health officers, pharmacists, laboratory specialists and GPs were interviewed, within 6 months after the transfer period. The interviews were audiotaped, transcribed, and independently coded by two of the authors. The codes were matched to the dimensions of the Normalization Process Theory.

Results

The general idea of the strategy was widely supported but generating the feedback was more complex than expected and the need for external support after transfer of the strategy remained high because participants did not assume responsibility for the work and the distribution of resources that came with it.

Conclusion

Evidence on effectiveness, a national infrastructure for these collaboratives and a general positive attitude were not sufficient for normalization. Thinking about managing large databases, responsibility for tasks and distribution of resources in large-scale implementation should start as early as possible when planning complex quality improvement strategies. Merely exploring the barriers and facilitators experienced in a preceding trial is not sufficient. Although multifaceted implementation strategies to change professional behaviour are attractive, their inherent complexity is also a pitfall for large-scale implementation.

Introduction

Local quality improvement collaboratives (LQIC's) or quality circles are working groups of physicians that meet to discuss medical issues in order to improve the quality of care. These LQIC's have been implemented throughout many countries in Europe in the past decades.¹ In the Netherlands one or more affiliated community pharmacists (to be named "pharmacists" in the rest of this manuscript) are part of a LQIC and acts as an expert on prescribing. Earlier work by our research group showed that feedback and peer review in LQICs is an effective approach to improving general practitioners' (GPs') test-ordering behaviour. This strategy is based on audit, provision of feedback, the use of educational materials and peer group discussion guided by clinical guidelines, in local groups moderated by local opinion leaders with intensive support from the academic research group. The LQICs met at regular intervals to discuss feedback on their actual test-ordering behaviour in terms of six clinical topics. The feedback was sent by post, accompanied by educational materials, and consisted of graphical reports on standardized volumes of tests ordered by a GP over a 6-month period, compared with the volumes for the group and the region. This multifaceted strategy resulted in an average 8-12% decrease in test volumes over a 6-month period, especially in areas with large interdoctor variation and superfluous testing.²

Since we wanted to build on this success, our objective was large-scale implementation of this strategy in the south of the Netherlands. We designed this trial to determine the effects of feedback and peer review under usual conditions, without external support from an academic research group. Therefore we used the existing LQIC structure as a backbone for the large-scale implementation of what had been proven effective under controlled conditions. We adapted the feedback and peer review strategy on three elements. The first element, adding prescribing behaviour as a target for feedback and peer review, seemed logical as it had been proven effective in separate trials.³⁻⁹ The other two new elements are requisite in a truly pragmatic trial.¹⁰⁻¹³ The second new element was that we provided distance support only to key stakeholders such as pharmacists and laboratory specialists, rather than direct support to all the LQICs. This also meant that LQICs were not to be recruited by the research group but by the regional health officers or laboratory specialists themselves. And, thirdly we allowed the LQIC groups freedom of choice regarding clinical topics. Because of the complexity of the intervention much time was spent on designing and testing the central database. We then pilot tested the strategy in two groups to identify flaws and errors, that were addressed and taken care of.

May et al. developed a theory that focuses on the process of transferring an intervention to the field, the Normalization Process Theory (NPT). This theory focuses

on factors that result from co-operative and collective activities, but are experienced and accounted for by individuals.¹⁴⁻¹⁶ These factors relate to the interactions between the people and organizations involved in the intervention, to the organizational contexts in which these interactions take place, to the internal and external coherence of the intervention, to the work that comes with building and sustaining the context of the intervention and the work of evaluating the intervention.^{17,18} Since our goal was normalization of the strategy, defined as sustainable incorporation into the existing LQICs we found this the best suitable model available. Table 3.1 presents an overview of these constructs as applied to LQIC; this formed the basis for analyzing the results of this process evaluation.

Table 3.1 Normalization Process Theory constructs and their operationalization in the feedback and peer review strategy.

NPT constructs	Operationalization
Coherence <i>The work that comes with designing and operationalizing a complex intervention within its context</i>	<p>Consensus is reached on the beliefs about workload and the values of work.</p> <p>All LQIC members, health officers, pharmacists and laboratory specialists feel the need for improving test ordering and prescribing. Agreement about goals, the importance of the work, the outcomes of the strategy and the need for a safe environment to discuss performance is reached and maintained.</p> <p>There is agreement on the new roles of all participants and their influence on relationships.</p> <p>The ownership and reliability of knowledge is clear and undisputed. All LQIC members agree that the pharmacist and laboratory specialists to be invited are in fact seen as opinion leaders having expert knowledge.</p> <p>The need for knowledge input from the pharmacists and laboratory specialists is recognized and shared by all GPs of a group. Laboratory specialists share the belief that the strategy is beneficial, i.e. positive for customer relations, to them and that the strategy can improve test ordering behaviour of GPs.</p>
Cognitive participation <i>The work that comes with building and sustaining commitment of all stakeholders</i>	<p>All participants actively adopt their roles.</p> <p>Health officers, pharmacists and laboratory specialists take up an active role to support the participants in the process.</p> <p>All recruiters actually start to recruit LQICs, reaching out to GPs in the field and forming new contacts</p> <p>The laboratory specialists and pharmacists attend the train-the-trainer sessions provided, to learn how to work with comparative feedback.</p> <p>All LQIC members commit to taking up an active role concerning peer review during the group session</p>

Table 3.1 (continued)

NPT constructs	Operationalization
Collective action <i>The operational work that comes with the intervention</i>	<p>Adaptation of the intervention to the needs and beliefs of a group is the aim of all group members.</p> <p>Group members together decide about adding a discussion of a case if they feel the meeting benefits from it.</p> <p>Beliefs about the boundaries of responsibilities and the autonomy available to each group are shared.</p> <p>The new role of the pharmacist and the laboratory specialist are discussed openly in the group, consensus is sought.</p> <p>GPs are taught by the pharmacists and laboratory specialists what the new role of moderator of a group is, the group members adopt these functions.</p> <p>The research group provides a train-the-trainer session on working with comparative feedback, discussing the strengths and limitations of it</p> <p>The risks of disturbing the group by reforming the system and patterns of behaviour are minimized.</p> <p>Laboratory staff briefs their database manager of the participation in the trial and the need for timely delivery of the correct data.</p> <p>Data is delivered to the central organization by laboratories and insurance companies</p> <p>Databases on test ordering and prescribing are kept up to date.</p>
Reflexive monitoring <i>The work of monitoring, evaluating and sustaining the intervention</i>	<p>Data gathering is facilitated by means of a standardized format.</p> <p>Allocation of funds, time, expertise and risks associated with the strategy are considered.</p> <p>Sustaining the allocation of funds and formalizing the relevant responsibilities.</p> <p>Stakeholders evaluate the strategy and agree on what group of stakeholders is responsible for sustaining the strategy after implementation.</p> <p>Issues of patient safety and ethics are considered.</p> <p>The research group has carefully watched not to collect any data leading to individual patients. Data of participating GPs is collected only after their written consent. Laboratories and insurance companies are informed formally of the precautions taken.</p> <p>Groups evaluate their meetings in order to adopt the standard format to their own needs.</p>

Little research has been done so far on the normalization of quality improvement strategies. In most studies on implementation strategies the researchers are involved by guiding the professionals through the intervention. Follow-up time is generally very short, and often relatively simple behavioural changes are aimed for.^{3,4,6,19-21} Therefore, we lack insight in whether and how the beneficial effects after primary implementation can be transferred to the field. However, there is some evidence from other research groups that support that the findings from our earlier work can be transferred successfully to a more complex setting. But, the level of control of the academic

researchers in these trial interventions seemed, at closer look higher than we anticipated on in our pragmatic trial.^{9,22} It therefore remains unclear what the effects are without academia monitoring and adjusting processes. We conducted this qualitative study to identify the barriers and facilitators for large-scale transfer of feedback and peer review. Our research question for this study was: what challenges were reported in the process of large-scale transfer of a complex implementation strategy to stakeholders in the primary care field?

Methods

The intervention in the underlying study can be described as follows. A strategy based on audit, provision of feedback, the use of educational materials and peer group discussion guided by clinical guidelines, in local groups moderated by local opinion leaders with support from the academic research group. The LQICs met at regular intervals to discuss feedback on their actual test-ordering behaviour in terms of six clinical topics. Feedback was sent by post, accompanied by educational materials, and consisted of graphical reports on standardized volumes of tests ordered by a GP over a 6-month period, compared with the volumes for the group and the region. Box 3.1 provides a more detailed outline of the adapted strategy, its stakeholders and its organizational levels.

General design

Qualitative methods were used to study the factors that hampered or facilitated the normalization process. We investigated the main factors at different levels by means of in-depth semi-structured interviews. The Maastricht Medical Research Ethics Committee has approved this trial.

Population

Participants recruited for the implementation study¹⁰

We divided the South of the Netherlands into 17 regions with a total of 24, often hospital-based, regional primary care laboratories and one laboratory operating at national level. One of the researchers (JT) visited at least one primary care coordinator in each region, who had access to data on test ordering. A flowchart of the results of the recruitment of laboratories and LQICs is given in Figure 3.1.

Box 3.1 LQIC feedback and peer review implementation strategy, summary.

Different roles can be distinguished at all levels of participation: the role of the key stakeholders in recruiting LQICs and delivering test ordering data; the role of insurance companies in delivering prescription data; the role of the pharmacists and laboratory specialists in moderating the groups; and the role of the GPs in the LQICs. The central level recruited regional health officers and laboratory specialists at regional primary care diagnostic facilities; the intermediate level. They signed a contract stating that (1.) at least two existing LQIC groups were to be recruited to participate in the study; (2) that data on test ordering had to be delivered to the research group in a predefined format and (3) that moderators for the LQICs had to be provided by them. Recruitment of LQICs by them was facilitated by the central level. A training session on how to moderate groups was provided to the pharmacists, laboratory specialists and regional health officers. During the training sessions, the reliability and validity of the data and possible sources of bias were explained and discussed, as well as the whole process from data extraction to processing it into feedback. Because there is no reliable register of LQIC groups in the Netherlands we did not know how many LQICs were available for recruitment. Regional recruitment by the key stakeholders fitted best in our pragmatic approach, also they were more likely to know the local situation concerning the LQICs.

LQIC groups decided on what clinical topics were chosen out of a set of five, 3 topics were scheduled for meetings.

Feedback

- Group members received individual comparative feedback on test ordering and prescribing with a summary of guidelines on the clinical topics chosen. We used national guidelines that were developed by authorised organisations such as the Dutch college of General Practitioners (www.nhg.org).
- Feedback was generated centrally from databases on tests and prescriptions. These databases were fed with data packages from local laboratories (often hospital-based) for test ordering data and from insurance companies for prescription data (see Figure 3.2).
- Feedback was standardized per 1000 patients per GP and per six months.

Train-the-trainer session for the moderators (one 3-hr session)

- Laboratory specialists and pharmacists were trained in providing support prior to the first meeting.
Content:
 - How the feedback was generated and where bias could occur. Examples of bias were given, and how to deal with questions on this.
 - Pharmacists and laboratory specialists were provided with an outline of an optimal meeting schedule, as well as an outline of the minimal process set that has to be completed during a meeting. They were trained to follow the detailed agenda in the first meeting and later actively adapt the agenda to their own needs, but to stick at least to the minimum process set.

Meeting

- Feedback and guidelines were to be discussed in a standardized format in the LQIC meeting.
- To enhance the discussion on test ordering, a specialist on clinical testing (i.e. clinical chemist) was assigned to each LQIC to moderate this discussion. The discussion on prescribing behaviour was held with the pharmacist present, as was common practice for many years.
- General practitioners were encouraged to formulate individual and group agreements on future test ordering and prescribing, based on room for improvement identified from interdoctor variation and recommendations in the guidelines.
- The group drew up an implementation plan after individual barriers and facilitators had been addressed, and group and individual goals were set.
- All meetings came in pairs:
 - The first meeting was on test ordering on a specific clinical topic.
 - The second meeting was on prescribing behaviour on the same clinical topic.
- CME points were granted and recorded by the researchers.

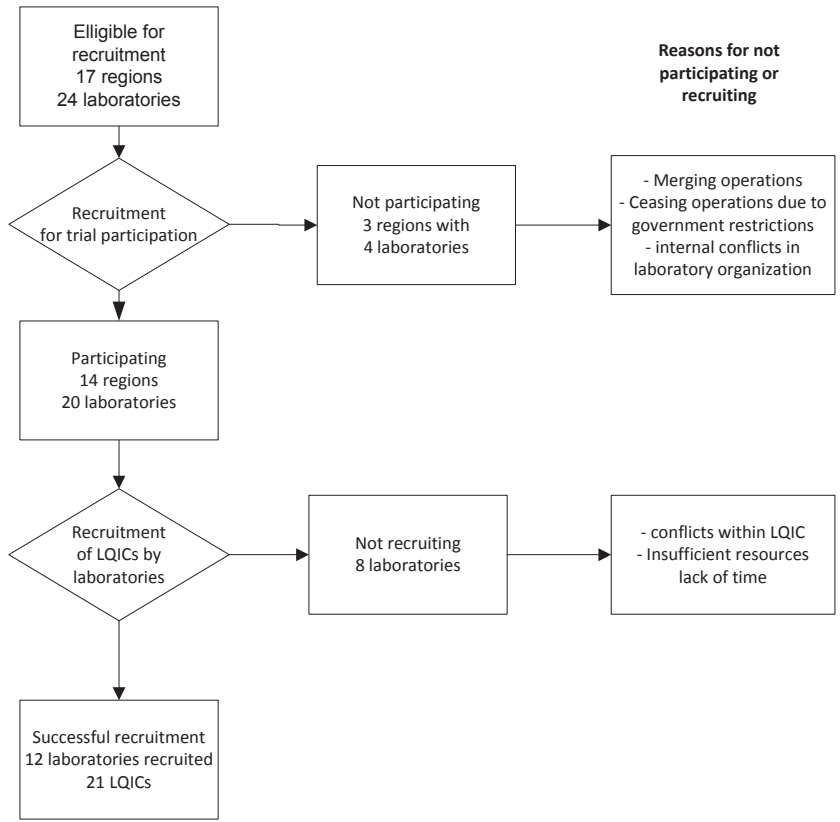


Figure 3.1 Flowchart of recruitment of laboratories, laboratory specialists or regional health officers and their recruitment of LQICs for the underling study.

Participants recruited for the present study (interview study)

We distinguished three levels of organization in the implementation;

1. the central level, consisting of the research group and a data cleaning and handling agency;
2. the intermediate level of the regional health officers, pharmacists and laboratory specialists;
3. the local level of GPs embedded in LQICs.

We purposively sampled candidates for individual interviews with one or more candidates from each of these three levels. We recruited candidates who had normalized the strategy, as well as those who had encountered many barriers when implementing this new strategy or even failed to recruit LQICs. Some of them were

enthusiastic throughout the trial whilst others were less enthusiastic or even negative. The 10 pharmacists, laboratory specialists and health officers were chosen because of their central role within the LQIC and because they acted as a liaison between the research group and the field. The six GPs were picked randomly from the groups but all acted as chair in their group.

Data collection and analysis

The interview outline was guided by the concepts of the Normalization Process Theory (Table 3.1). The interviews were held by an independent researcher (BvS). All interviews were audio taped and fully transcribed.

The transcripts of the interviews were imported in NVivo 7.0 and independently coded (JT, TW, BvS), using conventional content analysis.²³ We started with open coding, agreement on coding was reached by consensus (JT, TW, BvS). After completing the coding process all codes were matched to the NPT.

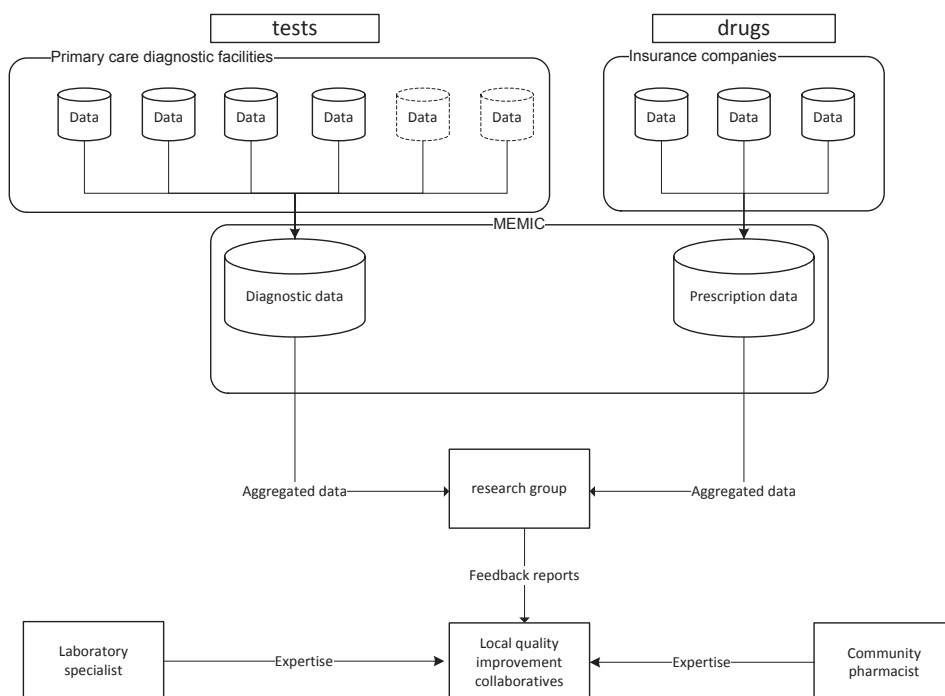


Figure 3.2 Schematic outline of the organization of the databases that generate input for the LQIC feedback and peer review strategy.

Results

Population and data saturation

Nineteen of the 22 pharmacists, laboratory specialists and GPs who were asked to participate in an interview agreed to do so (Table 3.2). The 10 pharmacists, laboratory specialists and health officers were active in 10 LQIC groups in total representing 116 GPs. The LQIC members that were interviewed were chairs of their groups, in total representing 65 GPs. All 19 interviews were transcribed verbatim. After coding and analyzing 14 interviews, no new insights emerged from the interviews, which was confirmed when we coded the remaining 5 interviews, indicating theoretical saturation. All codes were categorized and matched with corresponding items from the Normalization Process Theory; all items were covered by the model, and no items from the model remained unmatched.

Qualitative results from the interviews

The major themes emerging from the interviews (Table 3.3) are discussed below following the structure of the Normalization Process Theory, four constructs dealing with the work that comes with normalizing an innovation, namely coherence, cognitive participation, collective action and reflexive monitoring.

Table 3.2 Level of organization and degree of participation of interviewees for purposeful sampling (n=19).

		Level of organization		
		Central level	Intermediate level	Local level
Degree of participation	Local recruitment successful & Adaptation to LQIC successful	1 IT specialist ¹	3 experts ⁴ 2 health officers	3 LQIC members (GPs)
	Local recruitment successful & Adaptation to LQIC unsuccessful		5 experts ⁴	3 LQIC members (GPs)
	Local recruitment unsuccessful	not applicable ²	2 health officers	not applicable ³

¹ at the central level there is only 1 IT specialist who played a central role in generating the feedback.

² From the central level there is no contact with the groups that were not recruited.

³ there were no LQIC groups included, therefore no LQIC members were available to be interviewed.

⁴ Experts are pharmacists and laboratory specialists

Table 3.3 Factors reported to be barriers to the implementation of the feedback and peer review strategy.

	Level of organization		
	<i>Central level</i>	<i>Intermediate level</i>	<i>Local level</i>
Coherence	The coherence at the central level was obvious	Clarity about everyone's roles Strategy believed to be too complex	Clarity about everyone's roles GP's believed the strategy to be too complex Presentation of case report was desired by GPs
Cognitive participation	The quality of the feedback wasn't as indisputable, at start, as aimed for	Specificity of the feedback was questioned by pharmacists and laboratory specialists	Quality of the feedback was believed not to reflect current practice. Wait and see attitude of group members
Collective action	Data cleaning took much of available time Quality of the data delivered by hospitals and laboratories was insufficient	Lack of skills moderating the groups and leading the groups away from discussing the quality of the data.	Lack of skills of group members when working with comparative feedback or a moderator in a group.
Reflexive monitoring	Gathering data from laboratories and insurance companies proved more time consuming than anticipated on.	Workload heavier than expected for laboratory specialists	Lack of clarity about responsibilities for sustaining the strategy. GPs wanted the intervention to continue but felt not responsible for sustaining the strategy.

Coherence

All interviewees still favoured the idea of using feedback data to set goals for test ordering and prescribing behaviour, as had already been found in the recruitment phase of the underlying trial. Also the provision of comparative feedback, together with an outline of the applicable guidelines in the form of fact sheets, was considered to be adding value to the current level of discussion in the LQICs.

Interviewees reported different expectations of the workload involved in the LQIC intervention. The GPs expected a lower workload than before because they expected to receive a complete CME package similar to what they were used to, without any preparatory work, whilst laboratory specialists expected a heavier workload than before because they had to reach out to GPs.

Quote of a laboratory specialist

"Most GPs are present because of the CME credits, sprawled in a chair, knowing that no input is expected from them. That is illustrating the attitude of many GPs attending the meeting."

The expectation of the practical delivery of the intervention differed between interviewees; pharmacists and GPs expected to receive a complete package for a meeting, including prefab PowerPoints and case reports, while laboratory specialists expected they could start to reach out to LQICs and discuss test ordering based on the feedback. It appeared that everyone's role was not clear to all participants beforehand. This resulted in an unforeseen demand for support by the research group.

Many LQIC members, laboratory specialists and pharmacists reported they routinely used a case report as a starting point for the meeting and the following discussion. They sensed a feeling of frustration because such a case report was not provided in the strategy, hampering a smooth start-up process.

Quote of LQIC chair (GP):

"A case report is essential, if none is available, the whole meeting is of less value, I think. We can all read the guideline and look at our own performance data, but in the end you just need a case to start the discussion"

They also reported that the peer review model was depending too much on self-activation; they were willing to discuss feedback, provided it was embedded in this traditional one-way education.

A lack of in-depth discussion of clinical topics was reported by several GPs; they felt there was too little room for discussion about guidelines, and were not used to the type of discussion on inter-doctor variation that we aimed for. The level of group performance with regard to continuous professional development (CME) before the trial and the amount of room for improvement differed significantly between groups. Groups with a higher performance level prior to the trial seemed to be less content with the strategy compared to those with a lower previous level of group performance.

Quote of a pharmacist of a LQIC with a high performance level

"I feel we did not improve the quality of the pharmacotherapeutic meeting with this [The intervention, JT] compared to what we were used to do. Only the discussion on test ordering data seems to have some added value to the group and was new to us."

Laboratory specialists however reported that the strategy fitted in well with their organizations' strategy to intensify contacts with GPs.

The introduction of a market-driven healthcare system in the Netherlands acted as a facilitator of implementation of the LQIC strategy, because laboratories are encouraged to reach out to GPs and GPs are encouraged to be transparent and prove their performance.

Quote of a laboratory specialist

"....with the introduction of completion in healthcare we've lost 25% of the market to our main competitor. Therefore we are willing to work with GPs. We always have been willing out of professional interest, but now because we need to."

Cognitive participation

The recruitment of LQICs by health officers and laboratory specialists was unsuccessful in eight of the participating organizations, for various reasons (Figure 3.1). The role of recruiter was new to them and often new relationships had to be formed.

The moderation of the meetings by pharmacists and laboratory specialists was new to most of them. The group members were not familiar with this role and expected them to take the lead; this resulted in passive attitudes during meetings, with hesitations to start the discussion.

The quality of the data used for feedback was a major concern in many LQIC groups; especially for data on prescriptions. The data was often perceived as being too old, lacking specificity and being unreliable or insufficiently linked to current guidelines. This often hampered the peer review process and prevented the groups from establishing working agreements.

Quote of a participating GP:

"No, the data were too old; you should receive more recent data to discuss in a group. More recent data have higher appeal to the participants; otherwise the meeting will fail in its goal."

Despite the train-the-trainer session, the pharmacists and laboratory specialists felt insufficiently equipped to lead the group away from a defensive discussion about reliability and validity of the data.

Collective action

The primary care coordinators responsible for data deliveries to the research group as well as the data handling agency reported that the pre-defined data format facilitated data extraction and aggregation in a larger database. With data stemming from several hospitals and insurance companies, however, data cleaning took much of our available time. Also data was often received late or incomplete by the central level (research group) whereby much more time was spent gathering data than anticipated on. As a result, the feedback generated from these data was in some cases delivered only just in time or incomplete.

The process of generating feedback on demand proved to be bothersome and complex. For instance some LQICs demanded from the research group to deliver feedback within a few days. We found out that they had failed to inform the central level in time of their meeting schedule leaving the research team almost no time to generate the feedback and check it. We tried to prevent this from happening again by emphasizing to each new group starting, that all those involved had to receive the schedule in time. This largely solved the problem.

Despite the training in moderating the LQIC that the laboratory specialists and pharmacists received (train-the trainer), different views about the division of roles remained in some groups. They had to get used to working with comparative feedback and had to make their expertise available to the LQICs. The expertise of the pharmacists was not always undisputed. In the new format of the current trial their role was limited, leaving many responsibilities to the group members, who however often failed to adopt to these responsibilities, leaving the pharmacist “in front of the class” without the tools to perform. This sometimes hampered the adoption of their new role. Since the laboratory specialists were new in the groups, they were not confronted with a legacy of old routines. Although the content and goals of the train-the-trainer sessions were clear to most participants, many experienced a lack of skills to support the group dynamics efficiently and to introduce the strategy successful in a group. During the train-the-trainer sessions we provided the pharmacists and laboratory specialists with an outline of the proposed LQIC meeting structure, which proved helpful to them.

The less pragmatic oriented GPs, pharmacists and laboratory specialists reported they found the strategy to be too complex and ambitious. They desired more background information on the strategy and felt puzzled by its complexity.

Laboratory specialists had often met pharmacists and GPs for the first time during this trial, thus expanding and improving their network. This was reported to be of great value. However the workload for the laboratory specialists turned out to be heavier than expected, mostly because they were given a new role in supporting groups.

Quote of a laboratory specialist

“.....each session in a group means one evening away from home. Especially in this region where the distances are large [...] I told them I could run two groups, three would be too much”

Reflexive monitoring

Data provision by the laboratories proved to be difficult despite the standardized format. The database manager at the central level reported he had to put much effort in getting the data from the system administrator from the laboratories because they

did not prioritize data delivery. It was reported by them that saving the data-extraction queries, as the research group suggested, for use in the next time period was increasingly helpful in the course of the implementation period. By fine tuning these queries after each extraction, the quality of the delivered data improved.

The pragmatic design of this trial meant that each group was allowed and encouraged to adapt the intervention to their needs and beliefs. After first acquainting themselves with the strategy, many groups started to adapt it to their own situation and actually normalized the strategy.

Quote of a participating GP

"...off course we could just have worked with the feedback. But we cannot really deal with that. So we decided to add something. We decided to start off with taking a brief look at national guidelines, like an introduction. The second meeting was much more interesting because of this."

However some of the groups that had adapted the strategy to their own needs still hesitated to take responsibility for the organization of subsequent meetings and adopted a wait and see attitude, expecting action from the academic researcher.

Quote of a GP:

"It [the implementation, JT] went wrong because of a lack of support from the research group".

A lack of clarity about responsibilities appeared when asked who should take the lead in sustaining the LQIC infrastructure. Most experts on diagnostics stated that the GPs themselves had to come up with initiatives while the GPs considered the research group and insurance companies to be the designated organizations to sustain this.

The central administration of CME points was evaluated as a very positive aspect of the strategy.

Discussion

The results can be summarized as follows. Although the health officers, pharmacists and laboratory specialists as well as the local GP groups were willing to cooperate in this trial, and it was tailored to the existing Dutch infrastructure of peer review groups, we encountered problems at all stakeholder levels. Despite the lessons learnt from the pilot phase, it appeared to be more complex than expected to generate the feedback data and to produce the feedback reports in time. At the intermediate and local level it was unclear who was responsible for the work (generating test ordering data,

organizing meetings and moderating groups). Group members tended to adopt a passive role. In our extremely pragmatic trial, only an outline of the strategy was given, together with the minimum requirements which had to be met. Many groups were not used to structured discussions about performance on the basis of feedback. Most groups still held on to outdated CME models known to be ineffective.^{24,25} We anticipated on these routines and tendencies by training the LQIC moderators but conclude that this was insufficient, and the moderators in the successful prior study² apparently had been talented early innovators who were able to build on extensive prior experience for this task. In the end the train-the-trainer sessions that were designed to deal with the expected problems proved insufficient.

The interviews suggested that many GPs felt that the existing structure with informal meetings that did not require much preparation was an easy way to earn CME credits. This is in contrast with the new format, where all members have to make a significant contribution by studying their own performance data.

Gathering data for feedback took a lot of time and was not always successful, although many of the problems were gradually solved by redesigning parts of the dataflow. For example, we anticipated on a slow data delivery by laboratories by requesting data well in advance. A more regional focus with generating feedback reports at the hospital, skipping the bothersome task of processing data at a central level, could have facilitated the implementation process. Transferring the feedback-generating technology to all laboratories and decentralizing prescribing data to these laboratories is however a difficult operation with a considerable risk of failure and possibly ethical concerns because individual patients might be identifiable. The GPs were especially critical about the prescribing data. This might partly be caused by distrust towards insurance companies which provided us with this data. These defensive mechanisms towards feedback have also been found in earlier research.²⁶

Participants in our trial were pleased with the central registration of CME points, which was an unexpected facilitator of the process. However, whereas the aim of the study was to transfer responsibilities to the field, this aspect of certification ironically shifted from decentralized to centralized (academic) level.

Strengths and limitations

The purposeful sampling procedure enabled us to gather data from all levels of organization and all degrees of participation within these levels, thus providing us with insights into the process. Because the interviewer was independent, meaning that he was not part of the original research team, we feel they could speak freely about the strategy.

Although we only sampled a few LQICs, and within these LQICs, only one group member, data saturation was reached. By sometimes sampling one person at the intermediate level and one at the local level within one LQIC it may be that there exists an overlap within that group providing a more complete picture of the opinion of that LQIC. Nevertheless, it is not impossible that we have missed some barriers or facilitators that would have been mentioned by members who were not interviewed.

The number of GPs participating in this trial and the difficulties in implementing the strategy led us to the conclusion that scaling up feedback and peer review nationwide is impossible at the moment. We feel that lessons can be learned from our findings for other countries with a structure of LQICs and computerized databases containing data on test ordering and prescribing. Given the exceptional central role GPs in the Netherlands play in containing costs and resources in healthcare, we expect that LQICs in other countries may even have a larger gap to bridge when implementing this strategy. One of the reasons for the lack of success in implementing this strategy might be that we adapted the previous trial too much, thus stretching the capabilities at all levels to the limit.

The reliability of data stemming from large databases was often questioned. Because we used prescription data originating from insurance companies instead of data directly from GP practices or pharmacies the data is less accurate. For instance an error made in the administrative system of the pharmacy results in erroneously labelling the wrong physician to a certain prescription. These errors can be incidental but also structural. Nevertheless we feel the bias in these databases will be distributed evenly over all GPs. Therefore it will effect actual feedback but it won't affect the essence of the data; the comparability between peers.

Recommendations for practice and further research

When designing strategies to sustainably change professional behaviour, it is necessary to identify beforehand all individuals, organizations and structures involved and the work they have to do, as well as how the work changes their interactions. Merely exploring perceptions on barriers and facilitators experienced by early adopters in a preceding trial is not sufficient; influence of future policy and the allocation of resources also have to be clear. The NPT provides a good framework to assess whether all conditions for successful implementation are met beforehand.

We need to learn more on the last phase of implementation, that is, truly pragmatic trials of large-scale implementation where normalization is the goal and external support is limited. These trials should offer more intensive support at the start to overcome teething problems. This should then gradually develop into a form of distance support after the initial phase. From the start, these interventions should

focus on sustainability of the knowledge transfer. This study showed that quality improvement strategies that have shown to be effective may not be ready for scaling up to large-scale implementation at once, even not when a strong structure is present, as the existing LQICs in this trial. Implementation of complex interventions should be carried out gradually, while guiding participants efficiently by thorough train-the-trainer opportunities. This however will take resources, endangering the cost-effectiveness as evaluated in the original small-scale primary implementation exercises.

Conclusions

Even if an implementation strategy is found to be successful upon rigorous testing in a pragmatic cluster RCT on effects and costs, this does not guarantee that the success can be generalized if the strategy is transferred to the field on a large scale. Even in implementation projects, therefore, the degree of pragmatism is often not as high as in real practice. It seems necessary to perform a new diagnostic analysis of barriers and facilitators before every large-scale implementation.

When working with performance feedback for healthcare professionals, much attention should be paid to the underlying dataset. Although multifaceted interventions might be more effective in changing professional behaviour, they also carry a risk, as adding extra elements to an intervention also adds to its complexity. A balance between adding effectiveness, preventing complexity and finding the right scale for the implementation should be sought when designing implementation strategies.

References

1. Beyer M, Gerlach FM, Flies U, Grol R, Krol Z, Munck A, et al. The development of quality circles/peer review groups as a method of quality improvement in Europe. Results of a survey in 26 European countries. *Fam Pract* 2003;20:443-451.
2. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, Grimshaw J, et al. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289:2407-2412.
3. Arnold SR, Straus SE. Interventions to improve antibiotic prescribing practices in ambulatory care. *Cochrane Database Syst Rev*. 2005(4):CD003539.
4. van Driel ML, Coenen S, Dirven K, Lobbestael J, Janssens I, Van Royen P, et al. What is the role of quality circles in strategies to optimise antibiotic prescribing? A pragmatic cluster-randomised controlled trial in primary care. *Qual Saf Health Care* 2007;16:197-202.
5. Veninga CC, Denig P, Zwaagstra R, Haaijer-Ruskamp FM. Improving drug treatment in general practice. *J Clin Epidemiol* 2000;53:762-772.
6. Riou F, Piette C, Durand G, Chaperon J. Results of a 12-month quality-circle prescribing improvement programme for GPs. *Br J Gen Pract* 2007;57:574-576.
7. Welschen I, Kuyvenhoven MM, Hoes AW, Verheij TJM. Effectiveness of a multiple intervention to reduce antibiotic prescribing for respiratory tract symptoms in primary care: randomised controlled trial. *BMJ* 2004;329:431-433.
8. Gjelstad S, Høye S, Straand J, Brekke M, Dalen I, Lindbaek M. Improving antibiotic prescribing in acute respiratory tract infections: cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ* 2013;347:f4403.
9. Wensing M, Broge B, Kaufmann-Kolle P, Andres E, Szecsenyi J. Quality circles to improve prescribing patterns in primary medical care: what is their actual impact? *J Eval Clin Pract* 2004;10:457-466.
10. Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijers P, Winkens R, et al. A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: Study Protocol. *Implement Sci* 2009;4:6.
11. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furlong CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464-75.
12. Sackett D. Explanatory versus management trials. In: Haynes R, DL S, GH G, P T, editors. Clinical epidemiology: how to do clinical practice research. Philadelphia: Lippincott Williams and Wilkins; 2006: 183-188.
13. Meeuwse EJ, German P, Melis RJ, Adang EM, Goluke-Willems GA, Krabbe PF, et al. Cost-effectiveness of post-diagnosis treatment in dementia coordinated by Multidisciplinary Memory Clinics in comparison to treatment coordinated by general practitioners: an example of a pragmatic trial. *J Nutr Health Aging* 2009;13:242-248.
14. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: the normalization process model. *BMC Health Serv Res* 2007;7:148.
15. May CR, Finch T, Ballini L, MacFarlane A, Mair F, Murray E, et al. Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. *BMC Health Serv Res* 2011;11:245.
16. May CR, Murray E, Finch T, Mair F, Treweek S, Ballini L, et al. Normalization Process Theory On-line Users' Manual and Toolkit. 2010 [cited 2012 4th november]; Available from: <http://www.normalizationprocess.org>
17. Elwyn G, Legare F, Weijden T, Edwards A, May C. Arduous implementation: Does the Normalisation Process Model explain why it's so difficult to embed decision support technologies for patients in routine clinical practice. *Implement Sci* 2008;3:57.
18. Morrison D, Mair FS. Telehealth in practice: using Normalisation Process Theory to bridge the translational gap. *Prim Care Respir J* 2011;20:351-352.

19. Freemantle N, Harvey EL, Wolf F, Grimshaw JM, Grilli R, Bero LA. Printed educational materials: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2000(2):CD000172.
20. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2006(2):CD000259.
21. Fihn SD. Moving implementation science forward. *J Gen Intern Med* 2006;21 Suppl 2:S65-66.
22. Wensing M, Broge B, Riens B, Kaufmann-Kolle P, Akkermans R, Grol R, et al. Quality circles to improve prescribing of primary care physicians. Three comparative studies. *Pharmacoepidemiol Drug Saf* 2009;18:763-769.
23. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;15: 1277-1288.
24. Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien Mary A, Wolf F, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews* [serial on the Internet]. 2009; (2): Available from: <http://www.mrw.interscience.wiley.com/cochrane/clsystrev/articles/CD003030/frame.html>.
25. Hobma S. Directed self-learning as approach to continuing professional development. Maastricht: Maastricht University; 2005.
26. Mitchell MW, Fowkes FG. Audit reviewed: does feedback on performance change clinical behaviour? *J R Coll Physicians Lond* 1985;19:251-254.



Chapter 4

**Effect of audit and feedback with peer review on
general practitioners' prescribing and test ordering
performance: A cluster-randomized controlled trial**

Jasper Trietsch, Ben van Steenkiste, Richard Grol, Ron Winkens, H. Ulenkate,
Job Metsemakers, Trudy van der Weijden

Submitted

Abstract

Background

Changing professional behaviour remains challenging. With the ever-growing costs of healthcare and lagging adherence to guidelines, there is a need to find effective interventions that influence professional performance. Research on this topic is mostly performed in a well-controlled research setting. In a previous well-controlled RCT we found beneficial effects of a peer review strategy on test ordering performance. In the present study we tested if a large-scale implementation of the peer review strategy in the normal quality improvement setting in primary care had similar effects, this time to improve test ordering and prescribing performance.

Methods

We planned a cluster-RCT in existing local quality improvement collaboratives (LQICs) in primary care. The study ran from January 2008 to January 2011. LQICs were randomly assigned to one of two trial arms, with each arm receiving audit and feedback combined with peer review, but being presented with a different set of five clinical topics. All LQICs were exposed to the intervention, and acted as blind controls for the other arm. The collaboratives chose three of the five clinical topics presented to them, according to their preference, and agreed on targets for change. We performed an intention-to-treat pre-post analysis of the differences in testing and prescribing rates at the LQIC level, and a per-protocol analysis for the same differences.

Results

Twenty-one LQIC groups, including 197 GPs working in 88 practices, entered the trial. The overall results did not show a difference in the changes in test ordering or prescribing performance between intervention and control groups. It was observed that the groups with the highest baseline test ordering and prescription volumes showed the largest improvements.

Discussion

Our study shows that the beneficial results from earlier work could not be confirmed by our attempt to implement the strategy in the field. Based on our findings, it might be interesting, in addition to focusing on shared working agreements, to develop ways to tailor implementation efforts to those who need it most, viz. practices that show large, unwarranted practice variations.

Introduction

Spiralling healthcare costs are a major concern for policymakers worldwide. Overuse, underuse and misuse of healthcare are estimated to be responsible for 30% of the total spending on healthcare annually. It has been estimated that 7% of the wasted healthcare spending in the US is due to overtreatment, including test ordering and prescribing.¹ In the years 2004-2011, the average annual growth in the number of prescriptions in the Netherlands was 5.7%; in fact, the growth of the national income of the Netherlands has been smaller than the growth of the healthcare budget year after year.^{2,3} If nothing is done to reduce the growth in healthcare spending, it is feared that Western countries will not be able to pay the healthcare bill in the long term. Therefore, physicians are being targeted by policymakers to contribute on reducing waste in healthcare, and are encouraged to alter their habits.

An unsolved problem with changing professional behaviour is the lack of a clear and solid benchmark for the desired behaviour.^{4,5} This can be overcome by using practice variations as a proxy for quality of care. A certain degree of practice variation is clearly warranted, given the unique profiles of individual patients and practice populations. However, when practice variation is caused by underuse or overuse of care, this results in unwarranted variation and thus inappropriate care.⁶ In the Netherlands, general practitioners (GPs) now have access to over 100 evidence-based clinical practice guidelines. These guidelines have been developed by the Dutch College of General Practitioners (NHG) with the aim of reducing unwarranted practice variation and improving the quality of care provided. Although the general adherence to these guidelines seems quite reasonable, viz. approximately 70%, there is considerable practice variation in test ordering and prescribing, indicating room for improvement in poorly performing practices.⁷⁻¹²

In local quality improvement collaboratives (LQICs), general practitioners meet on a regular basis to discuss current issues and gain new insights concerning test ordering and prescribing behaviour. Healthcare organisations and governments promote these meetings as a means to implement guidelines. LQICs are widely implemented in primary care, mainly in Europe and, to a lesser degree, in North America. In the Netherlands, almost 98% percent of all GPs take part in an LQIC on a voluntary basis. These groups have a long history as pharmacotherapeutic audit meetings (PTAMs), having started in the early 1980s (Box 4.1). In the Netherlands, local pharmacists are members of these groups and are well respected for their input and knowledge. Nowadays, these PTAM groups are no longer exclusively focussing on pharmacotherapeutic topics and developing into LQICs working on broad quality improvement agendas,¹³ so we will use the term LQIC instead of PTAM in this article.

Box 4.1 Existing PTAM infrastructure in primary care in the Netherlands.

LQICs in primary care in the Netherlands were formed in the early 1980s. Their goal was to rationalise prescribing behaviour by discussing this behaviour in the presence of fellow general practitioners (GPs) and at least one community pharmacist. The government promoted these groups, the professional association granted continuing medical education credits (CME points) and insurance companies supported the groups financially. However, participation in an LQIC is not obligatory and is therefore a voluntary commitment of GPs and pharmacists to establish agreements on pharmacotherapeutic treatment.

In past decades, this form of CME has gained wide acceptance amongst GPs, and as many as 98 percent of all GPs participate in an LQIC.¹ These groups usually consist of six to ten GPs, and they tend to be very stable over the years.² The quality of the group work is classified into four categories, with level 1 being the lowest. At the highest level, agendas are set prior to a series of meetings, there is a high average attendance, audit data are discussed and group working agreements are established and evaluated afterwards. At the lowest level, groups simply meet and discuss a disease and the pharmacotherapeutic options. These sessions are often facilitated by the pharmaceutical industry, while level 4 meetings are preferably not facilitated by pharmaceutical companies. Most groups reach level 2 or 3, while level 4 is reached in at most 20% of all cases.¹ Even when groups do meet the criteria for level 4, the data from the audit is often flawed and incomparable due to a lack of access to databases containing aggregated prescribing data.

¹ Florentinus SR, van Hulten R, Kloth ME, Heerdink ER, Griens AM, Leufkens HG, Groenewegen PP: The effect of pharmacotherapy audit meetings on early new drug prescribing by general practitioners. *Ann Pharmacother* 2007, 41:319-324.

² Teichert M, van der Aalst A, de Wit H, Stroo M, De Smet PA: How useful are prescribing indicators based on the DU90% method to distinguish the quality of prescribing between pharmacotherapy audit meetings with different levels of functioning? *Eur J Clin Pharmacol* 2007, 63:1171-1177.

These LQICs are an attractive target for interventions aimed at changing professional behaviour both effectively and efficaciously.¹³⁻²⁰ In a robust trial on three clinical topics, Verstappen et al. showed the beneficial effects of a multifaceted strategy involving audit and feedback with peer review in LQICs on test ordering behaviour. They found a reduction in the volumes of tests ordered ranging from 8 to 12% for the various clinical topics.^{21,22} Lagerlov et al. showed that individual feedback embedded in local peer group discussions improved appropriate treatment of asthma patients by 21% and urinary tract infections by 108%, compared to baseline values.²³ There is also evidence suggesting that the mere provision of information on test fees when presented at the time of the order entry reduces the volumes of tests ordered.²⁴

Most of this evidence, however, stems from trials focussing on a single or limited number of clinical topics, and involving a strong influence of the researcher on the participants, e.g. as moderator during sessions. Moreover, in the Verstappen trial, the included groups were selected by the researcher and can be regarded as innovator groups. We wanted to build on the experiences from the work by Verstappen et al. and undertake a large-scale implementation of the multifaceted strategy, in a pragmatic trial with much room for the LQICs to adapt the strategy to their own needs and

without any researchers being present. The trial was pragmatic because it studied the effects of the intervention in the normal quality improvement setting in primary care with minimal support from the researchers.

We expected that the beneficial effects found in earlier, more controlled trials, would be confirmed. We hypothesized that our multifaceted intervention using auditing and feedback with peer review, including the establishment of shared working agreements, and moderated by local opinion leaders within existing LQICs, would reduce inappropriate testing and prescribing behaviour. In addition to the feedback data, the LQIC participants received educational materials comprising summarised guideline recommendations and information on the costs of tests and drugs. We even expected a synergetic effect by aiming at both test ordering and drug prescribing, as each topic was to be discussed twice in separate meetings, the first on test ordering and the second on prescribing. Our research question was therefore: What is the effect of audit and feedback with peer review on general practitioners' prescribing and test ordering performance?

We also report the sum scores of volumes of tests and prescriptions in a per-protocol analysis. This analysis was not planned in the study protocol[13], but we decided to add it as the process evaluation of the study revealed that the uptake of the strategy was much lower than expected.²⁵

Methods

Design

We conducted a two-arm cluster-randomised trial with the LQIC as the unit of randomisation and with central allocation. Core elements of the intervention are audit and comparative feedback on test ordering and prescribing volumes, dissemination of guidelines and peer review in quality improvement collaboratives moderated by local opinion leaders.¹³

The intervention started in January 2008 and was completed as planned at the end of December 2010. We measured baseline performance during the six months before the intervention, and follow-up performance during the six months after the intervention. The design of this intervention is described in more detail in the trial protocol.¹³ Approval for this trial was obtained from the Maastricht Medical Research Ethics Committee. All participating GPs were asked to sign informed consent prior to randomisation.

Setting and participants

Recruitment was restricted to the south of the Netherlands, because of our access to prescribing data of GPs working in this area. First, the regional health officers or key laboratory specialists for all 24 primary care diagnostic facilities in the south of the Netherlands were identified and recruited by the first author. They were trained in a three hour session in their region by the researchers. Object of this training was to transfer knowledge on effectively discussing test ordering and prescribing behaviour, setting working agreements, on how to effectively moderate meetings and how to deal with questions on the validity of the feedback data or other aspects of the intervention. Also written and digital materials were made available to enable them to facilitate recruitment of LQIC groups. The routines of the LQICs were deliberately left unchanged as they represented normal quality improvement routines in primary care in the Netherlands. Only when test ordering was discussed a laboratory specialist from the diagnostic facility moderated the group discussion.

Intervention

In this trial with audit and feedback with peer review in LQICs we wanted to test the results on test ordering behaviour and prescribing behaviour of the strategy. Aggregated comparative feedback was provided on tests ordered or drugs prescribed in the period of six months before each meeting in which it was discussed. Feedback was sent to the moderator for that session (the local pharmacist or laboratory specialist). At the start of each meeting, each GP received feedback report on their own performance. The feedback was adjusted for practice size and compared with the aggregated results from their practice, their LQIC group and neighbouring groups (Figure 4.1), as well as an outline of the recommendations from the guidelines, validated by clinical experts (Appendix 4.1). To mimic the normal situation in self-directing LQIC groups, the groups in both arms were allowed to choose three clinical topics out of a set of five presented to them. The set of five topics differed between the two arms (Table 4.1). Each group planned two paired meetings for each topic, one on test ordering and one on prescribing making it a total of six meetings. Each meeting lasted between 90 and 120 minutes as was usual before depending on the intensity of the discussion. Groups were encouraged by the trained moderator (see under “setting and participants”) to establish working agreements to improve their performance, and to discuss barriers to change. The LQICs were allowed to adapt the format of the meeting to their own needs and routines, as long as peer review and working agreements were included. At the end of each meeting groups were asked to fill out a form stating what working agreements and goals were set.

Feedback reports were generated from two main databases, one on diagnostic tests and one on prescriptions, with data originating from primary care diagnostic facilities and the two dominant insurance companies in the region. The databases contained data on the specific test or drug, the date it was ordered or prescribed, the practice in which the physician who had ordered or prescribed it worked, the date of birth of the patient, their gender and, in the case of prescriptions, the number of defined daily dosages (DDDs) that were prescribed. A more detailed description of the intervention is available in the previously published trial protocol.¹³

Table 4.1 Sets of clinical topics and the number of meetings held for each topic.

Clinical topic	Study arm	Number of meetings on test ordering	Number of meetings on prescribing	Desired direction of change	
				tests	drugs
Anaemia	A	9	9	↓	=
Dyslipidemia	A	2	2	↓	=
Prostate complaints	A	6	6	↓	↓
Rheumatic complaints	A	10	10	↓	=
UTI	A	2	2	↓	↓
Chlamydia trachomatis	B	8	8	↑	↑
Diabetes mellitus II	B	5	4	↓	=
Stomach complaints	B	6	6	↓	↓
Perimenopausal complaints	B	4	4	↓	↓
Thyroid dysfunction	B	6	6	↓	=
Total		58	57		

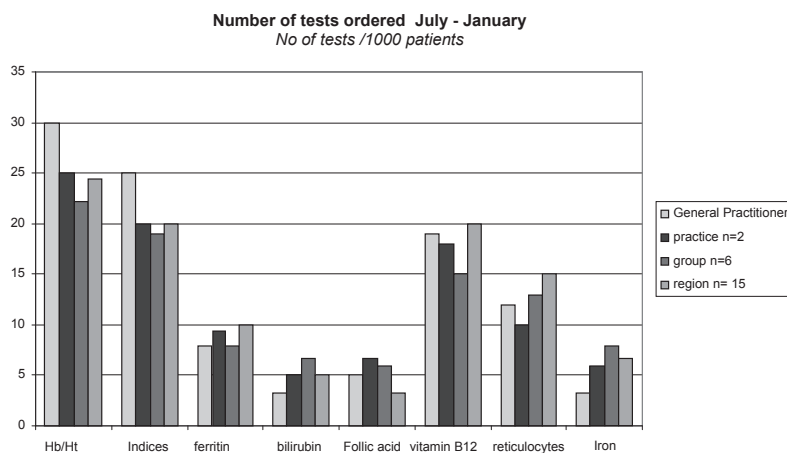


Figure 4.1 Example of the graphical comparative feedback (this image doesn't reflect actual data).

Data collection and main outcome measures

The primary outcome measures were the volumes of tests ordered and drugs prescribed per practice, per 1000 patients, per 6 months. Although data on a large number of diagnostic tests and prescriptions were available (Appendix 4.2), only results on key tests and drugs for each clinical topic are reported in this paper. The identification of these key tests and drugs was based on consensus within the research group and one clinical expert on each topic before the intervention started.

Sample size

We calculated that a total of 44 LQICs would be sufficient to detect a standardised effect size (Cohens d) of 0.5, with a significance level α of 0.05, a power of 0.9, an ICC of 0.1 and a mean group size of seven GPs. Anticipating a dropout rate of 10%, we would need to recruit 50 LQICs.^{10,13}

Randomisation

Prior to the randomisation, the LQICs were stratified on their level of group performance, as assessed by a questionnaire resulting in four levels of group work, from 'poor' to 'good' (Box 4.1). The level of group performance may be a confounder for the ability to establish shared working agreements and for the quality of prescribing behaviour.²⁶⁻²⁹ By stratifying on this, we ensured an equal distribution of these levels over the trial arms (Table 4.2). An independent research assistant produced a computer-generated allocation list and allocated the LQICs to arm A or arm B, while the researcher was blinded to this process. Groups in both trial arms were exposed to the same intervention, but on different clinical topics. Each LQIC in one arm served as an unmatched control for the LQICs in the other arm.^{30,31} Groups were blinded to the clinical topics discussed in the other arm. The researcher was blinded until all data analyses had been completed.

Data analysis

To analyse the overall differences between the two arms we compared performances at the LQIC level for all key tests and drugs during the six months prior to the intervention with performances during the six months after completion of the intervention period. We analysed according to the intention-to-treat principle. In addition, we performed a per-protocol before and after analysis to test for effects in the groups that had actually organised a meeting on a specific topic, with all other groups acting as controls in the analysis.

Table 4.2 Characteristics of the participating GPs and groups.

	Arm A	Arm B	p
Number of groups	10	11	0.40
Number of practices	39	49	0.45
Mean group size (GPs per group)	8.1 (0.64 [‡])	10.55 (1.70 [‡])	0.20
Mean group level	2.70 (0.26 [‡])	2.82 (0.33 [‡])	0.78
No. of men (n total)	53 (81)	78 (116)	0.791
Mean age of GPs	47.9 (0.91 [‡])	47.1 (0.77 [‡])	0.538

[‡] standard deviation

The time intervals for our per-protocol analyses were six months prior to each LQIC meeting, compared with 0-6 months after each LQIC meeting in the case of tests, and 3-9 months after each meeting in the case of prescribing (Figure 4.2). By using this washout period we avoided contamination with long-term prescriptions.

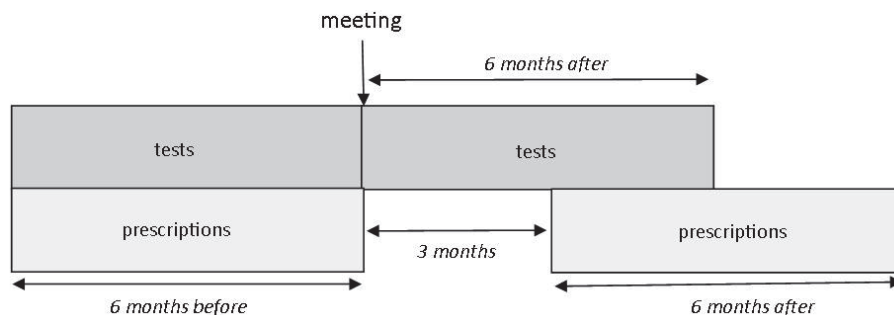


Figure 4.2 Graphical display of the periods defined for the baseline and follow-up measurement of tests ordered and drugs prescribed in the per protocol analysis.

We used a Chi-square test or t-test to check if the stratification of groups had led to an even distribution of the LQIC group performance levels and the characteristics of the participants over the trial arms and topics. The group effect (intervention versus control) on prescribing rate and test ordering rate after the intervention was assessed using a linear mixed model with the LQIC as a random effect to account for the clustering of practices within the LQIC. In addition, group (intervention or control), baseline value of the outcome measure (before the intervention) and interaction between the baseline and group were included as fixed factors. If the interaction term was not statistically significant, it was removed from the model, and only the overall group effect is presented. This method was used for both the crude analysis and the per-protocol analysis.

If the interaction term was statistically significant, the overall group effect (obtained from the model without the interaction term) and the group effects for different baseline values, at the 10th and 90th percentiles of the baseline variable, are presented to assess the effects at both ends of the spectrum. We expected that change would mostly be seen in GPs in the 90th percentile, as this is a clear indication of overuse and marks a need to decrease test ordering or prescription volumes. Although it is not clear what the benchmark is for volumes of tests and prescriptions, we did not expect GPs at the other end of the spectrum – the 10th percentile – to clearly fail in terms of underuse of tests and prescriptions.

All data analyses were performed using IBM SPSS Statistics for Windows, Version 21.0 (Armonk, NY: IBM Corp). P-values ≤ 0.05 were considered statistically significant.

Results

Participants

Out of the 24 primary care diagnostic facilities (laboratories) we approached, 20 consented to participate in the trial. The other facilities reported organisational changes (2 facilities), governmental restrictions (1 facility) or internal conflicts within the organization (1 facility) as a reason not to participate. Eight of the 20 facilities failed to recruit LQICs, for various reasons.²⁵ The 12 remaining facilities managed to recruit 21 LQICs for the trial (Figure 4.3). The 21 groups consisted of 197 GPs working in 88 practices, and 39 community pharmacists. Eight laboratory specialists participated in the groups when test ordering was being discussed (some laboratory specialists covered more than one region). The characteristics of the participating groups and the GP members of these groups are described in Table 4.2.

In arm A, one group had to cancel its last two meetings because feedback data was not available, and there was no time left to schedule a new meeting within the intervention period. In arm B, one group dropped out of the intervention after one meeting, and another did so after two meetings because they doubted the validity of the feedback data. This is described in detail in the qualitative process evaluation of this trial.²⁵

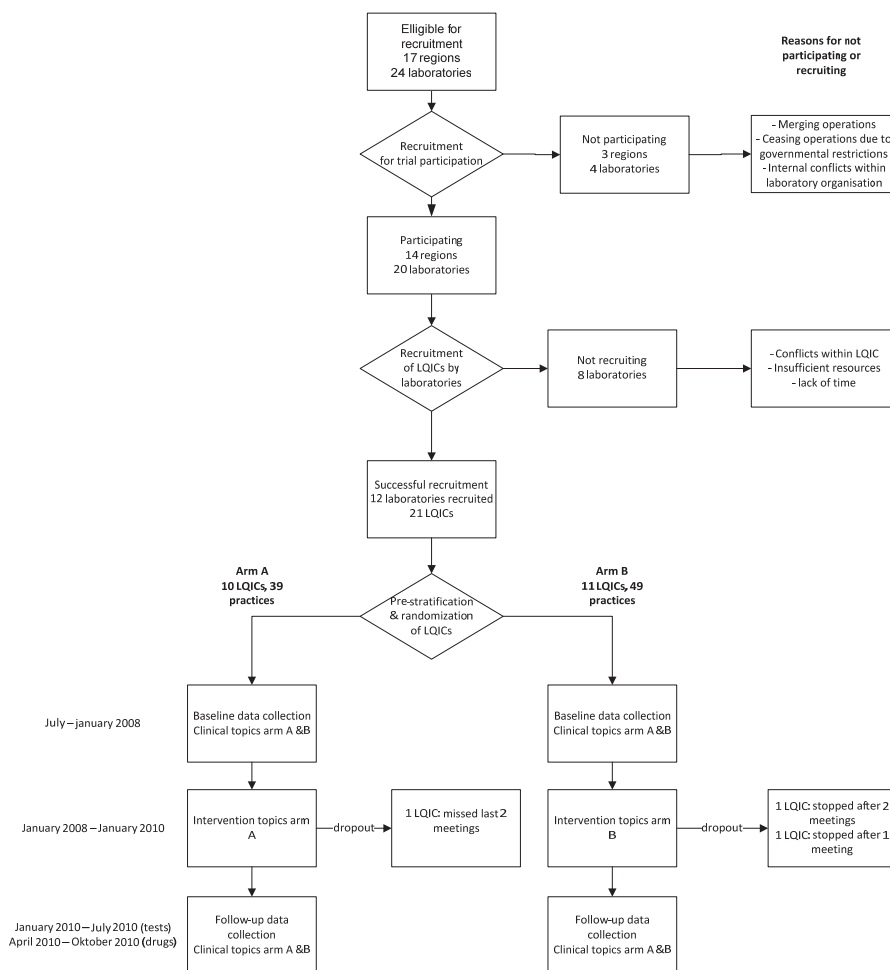


Figure 4.3 Flowchart of recruitment of laboratories, laboratory specialists or regional health officers and their recruitment of LQICs with the number of GPs in brackets.

Exclusion of data before analysis

At the time when the intervention was designed, the recommendations for dyslipidemia and type 2 diabetes mellitus were provided in two separate guidelines. However, at the start of the actual intervention, the guidelines on diabetes management and dyslipidemia treatment were merged into one new, multidisciplinary national guideline on cardiovascular risk management. This was directly followed by a massive government led intervention to transfer care for diabetics and cardiovascular risk patients from specialist care to GPs. Part of this transfer was the institution of a pay

for performance model for these two topics together with the introduction of many outcome indicators being tracked by newly introduced and completely integrated software.³² As part of this campaign much publicity was created in both the professional and public media. This caused substantial contamination of our intervention contrast, resulting in an inability to interpret the results on these clinical topics. Therefore, we chose to exclude the topics of dyslipidemia and type 2 diabetes mellitus from the analyses. The results on test ordering for the clinical topics of urinary tract infections (UTI) and stomach complaints could not be calculated either, due to insufficient data on test ordering from laboratories and diagnostic facilities to report these data (e.g., urine cultures and gastroscopies). Results on prescription rates for Chlamydia are not shown because it proved impossible to link the prescribed antibiotics reliably to this condition. This problem did not occur for UTI, as we confined the data to nitrofurantoin and trimethoprim, which are antibiotics that are only indicated or prescribed for UTI treatment in Dutch primary care.

Results of the intention-to-treat analysis

The intention-to-treat before and after analyses on test ordering did not show any differences between the intervention and control groups, with wide confidence intervals, all including 0, and all with p-values well above 0.05 (Table 4.3). Interaction with the baseline values was present for rheumatic complaints. The adjusted difference between intervention and control groups at the end of the intervention showed a difference of -7.8 tests/1000 patients [95% CI: -22.0, 6.5; p: 0.257] for the lower p10 while the adjusted difference between intervention and control groups at the end of the intervention showed a difference of 16.3 tests/1000 patients [95% CI: 1.7, 30.9; p: 0.031] for the highest p90 groups, all numbers are corrected for baseline differences. The intention-to-treat analysis on drug prescriptions showed a difference in the desired direction for misoprostol of -0.1 DDDs/1000 patients (95%CI: -0.2, -0.0; p: 0.008). Interaction with baseline values was present for misoprostol, the triple therapy for *H. pylori* eradication (PantoPac®), antithyroid preparation drugs and clonidine (Table 4.4), showing changes in the desired direction for misoprostol (P10: -0.1 DDDs/1000 patients [95%CI: -0.2, -0.0; p: 0.039] and P90: -0.1 DDDs/1000 patients [95%CI: -0.2, -0.0; p: 0.039]), for antithyroid preparations (P10: -1.9 DDDs/1000 patients [95%CI: -7.2, 3.4; p: 0.474] and P90: 12.3 DDDs/1000 patients [95%CI: 4.6, 19.9; p: 0.002]) and for clonidine (P10: 0.6 DDDs/1000 patients [95%CI: -0.4, 1.6; p: 0.215] and P90: -2.4 DDDs/1000 patients [95%CI: -3.9, -1.0; p: <0.001]). The analysis with the interaction factor showed no changes for triple therapy at either end of the spectrum (P10: 0.9 DDDs/1000 patients [95%CI: -0.2, 1.9; p: 0.117] and P90: -0.4 DDDs/1000 patients [95%CI: -1.6, 0.7; p: 0.413]). All changes are expressed as adjusted difference between

intervention and control groups at the end of the intervention, corrected for baseline differences.

Results of the per-protocol results

Table 4.5 shows the results of the per-protocol analyses on test ordering volumes for all groups that covered a specific topic (intervention group) compared to all other groups (their controls). We found a difference between both trial arms in the desired direction in test ordering for two clinical topics, thyroid dysfunction (36.6 tests/1000 patients [95%CI 10.5, 62.7; p: 0.007]) and perimenopausal complaints (3.2 tests/1000 patients [95%CI 0.1, 6.4; p: 0.046]).

Testing for interaction with baseline measurements showed a difference in test ordering rates in the desired direction for those GP practices with a baseline test-ordering rate at or above p90 for chlamydia infections (-10.4 tests/1000 patients [95% CI -17.6, -3.3; p: 0.006]), rheumatic complaints (14.8 tests/1000 patients [95%CI 2.1, 27.4; p: 0.025]) and perimenopausal complaints (6.6 tests/1000 patients [95%CI 2.4, 10.7; p:0.002]).

Table 4.6 shows the results of the per-protocol analysis on prescribing performance. A difference in the overall volume of all prescribed drugs between intervention groups and their controls was observed for medication prescribed for prostate complaints (28.5 DDDs/1000 [95%CI 6.5, 50.7; p: 0.016]), stomach complaints (61.1 DDDs/1000 patients [95%CI 15.0, 107.2; p: 0.014]) and thyroid dysfunction (12.6 DDDs/1000 patients [95%CI 0.7, 24.4; p: 0.040]). For each of the clinical topics we also analysed each Anatomical Therapeutic Chemical Classification System (ATC) group in that topic separately, as changes found for specific ATC groups could represent clinically relevant changes. These results are shown in more detail in Table 4.6. Testing for interaction with baseline measurements again showed a statistically significant interaction for several topics and specific ATC groups. All showed larger differences in prescribing rates before and after the intervention for the practices in the p90 range than for those in the p10 range (Table 4.6).

Table 4.3 Crude analysis of changes in test ordering rates. The desired direction of change is given in the column headed 'Direction'.

Clinical topic	Intervention				control				icc	Direction	p
	n	Mean before (sd)	Mean after (sd)	n	Mean before (sd)	Mean after (sd)	B [#] (95% CI)				
Anaemia	39	177.8 (173.1)	208.4 (216.9)	49	208.1 (148.7)	247.3 (176.7)	15.3 (-53.7, 83.3)	0.31	↓	0.630	
Rheumatic complaints	39	12.9 (20.1)	13.1 (23.0)	49	12.9 (16.2)	15.0 (27.5)	2.9 (-9.0, 14.9)	0.31	↓	0.581 *	
Prostate complaints	39	34.4 (28.0)	36.1 (26.2)	49	37.6 (27.9)	40.7 (31.0)	1.8 (-4.9, 8.6)	0.04	↓	0.575	
Chlamydia infections	49	9.9 (15.4)	8.9 (19.5)	39	7.8 (9.6)	6.3 (10.0)	2.2 (-9.5, 13.9)	0.74	↑	0.701	
Thyroid dysfunction	49	178.9 (139.0)	206.1 (155.4)	39	160.6 (158.3)	186.1 (151.7)	3.9 (-26.3, 34.11)	0.03	↓	0.788	
Perimenopausal complaints	49	5.6 (6.2)	5.2 (6.0)	39	6.4 (7.6)	7.4 (8.5)	-1.4 (-5.0, 2.1)	0.29	↓	0.404	

Numbers represent the number of tests prescribed per 6 months per 1000 patients. n = number of practices; [#] Adjusted difference between groups at end of intervention, corrected for baseline; * statistically significant interaction: Rheumatic complaints: P 10: B (95% CI): -7.8 (-22.0, 6.5). p: 0.257; P 90: B (95% CI): 16.3 (1.7, 30.9), p: 0.031.

Table 4.4 Crude analysis of changes in prescribing rates showing sum scores per clinical topic and scores per drug. The desired direction of change is given in the column headed 'Direction'.

Clinical topic	Intervention				Control				icc	Direction	p
	n	Mean before (sd)	Mean after (sd)	n	Mean before (sd)	Mean after (sd)	B [#] (95% CI)				
Anaemia	39	25.0 (20.7)	45.1 (40.0)	49	27.4 (28.4)	49.2 (55.7)	3.9(-13.9, 21.5)	0.21	=	0.644	
Iron preparations, oral	39	43.5 (31.6)	47.3 (36.5)	49	47.5 (38.7)	59.7 (48.7)	11.9 (-3.2, 27.0)	0.21	↓	0.113	
Urinary tract infections	39	59.1 (43.3)	107.1 (69.6)	49	66.5 (58.6)	127.5 (102.2)	20.0 (-20.9, 60.8)	0.38	↓	0.306	
Antibiotics UTI	39	49.6 (35.7)	86.6 (56.2)	49	55.0 (49.5)	103.3 (81.5)	18.5 (-15.3, 52.3)	0.42	↓	0.258	
Prostate complaints	39	9.5 (9.5)	20.4 (18.6)	49	11.5 (11.2)	24.2 (24.7)	1.4 (-7.2, 10.0)	0.01	↓	0.735	
Alpha receptor blockers	39	9.5 (9.5)	20.4 (18.6)	49	11.5 (11.2)	24.2 (24.7)	1.4 (-7.2, 10.0)	0.01	↓	0.735	
5 alpha reductase inhibitors	39	9.5 (9.5)	20.4 (18.6)	49	11.5 (11.2)	24.2 (24.7)	1.4 (-7.2, 10.0)	0.01	↓	0.735	
Stomach complaints	49	329.7 (286.9)	612.0 (477.0)	39	282.6 (210.1)	491.8 (333.4)	106.7 (-70.7, 284.2)	0.63	↓	0.219	
H2 antagonists	49	27.8 (26.4)	32.4 (30.0)	39	22.3 (20.5)	21.3 (16.4)	12.5 (-1.8, 26.8)	0.44	↑	0.081	
Proton pump inhibitors	49	300.8 (262.9)	578.3 (452.3)	39	259.4 (192.3)	469.1 (322.1)	93.7 (-71.9, 259.4)	0.62	↓	0.246	
Misoprostol	49	0.1 (0.4)	0.0 (0.2)	39	0.1 (0.4)	0.2 (0.5)	-0.1 (-0.2, -0.0)	0.00	↓	0.008 *	
Triple therapy (ATC A02BD04)	49	1.0 (1.7)	1.3 (2.0)	39	0.9 (1.3)	1.1 (1.7)	0.2 (-0.7, 1.2)	0.49	↓	0.593 *	
Thyroid dysfunction	49	81.3 (66.3)	137.2 (102.6)	39	71.4 (48.9)	107.1 (71.6)	21.2 (-14.1, 56.6)	0.42	=	0.217	
Thyroid hormones	49	74.8 (60.8)	125.0 (92.6)	39	66.9 (47.4)	100.8 (70.0)	18.2 (-16.0, 52.4)	0.45	=	0.269	
Antithyroid preparations	49	6.5 (8.6)	12.2 (17.8)	39	4.5 (4.4)	6.3 (7.3)	2.6 (-1.7, 7.0)	0.09	↑	0.218 *	
Perimenopausal complaints	49	56.4 (42.7)	56.8 (41.8)	39	46.3 (34.2)	46.0 (29.6)	6.6 (-12.0, 25.2)	0.56	↓	0.459	
Oestrogens, oral	49	14.1 (13.6)	12.7 (12.1)	39	12.3 (11.8)	11.2 (10.4)	0.2 (-3.6, 4.1)	0.01	↓	0.889	
Contraceptives (>50y)	49	3.6 (3.2)	6.4 (4.9)	39	3.2 (3.9)	5.9 (4.2)	1.1 (-1.9, 4.2)	0.34	↓	0.431	
Hormone replacement therapy	49	34.8 (28.6)	34.7 (27.0)	39	28.3 (19.7)	26.3 (17.8)	5.8 (-6.5, 18.0)	0.70	↓	0.331	
Clonidine	49	0.5 (1.8)	0.5 (2.0)	39	0.2 (0.5)	0.4 (2.0)	0.1 (-1.0, 1.2)	0.13	↓	0.908 *	
Tibolone	49	3.4 (4.1)	2.4 (2.7)	39	2.3 (2.8)	2.2 (3.4)	-0.3 (-1.4, 0.8)	0.00	↓	0.615	

Numbers represent the number of DDDs prescribed per 6 months per 1000 patients. n=number of practices; * Adjusted difference between groups at end of intervention, corrected for baseline; * statistically significant interaction. Misoprostol: P 10: B (95% CI): -0.1 (-0.2, -0.0). p: 0.039; P 90: B (95% CI): -0.1 (-0.2, -0.0). p: 0.039. Triple therapy (ATC A02BD04): P10: B (95% CI): 0.9 (-0.2, 1.9). p: 0.117; P90: B (95% CI): -0.4 (-1.6, 0.7). p: 0.413. Antithyroid preparations: P10: B (95% CI): -1.9 (-7.2, 3.4). p: 0.474; P90: B (95% CI): 12.3 (4.6, 19.9). p: 0.002. Clonidine: P10: B (95% CI): 0.6 (-0.4, 1.6). p: 0.215; P90: B (95% CI): -2.4 (-3.9, -1.0). p: <0.001.

Table 4.5 Per-protocol analysis of change in test ordering rates for each topic. The desired direction of change is given in the column headed 'Direction'.

Clinical topic	Intervention				control				B* (95% CI)	icc	Direction	p
	n	Mean before (sd)	Mean after (sd)	n	Mean before (sd)	Mean after (sd)						
Anaemia	35	176.8 (172.2)	209.4 (174.6)	53	199.6 (146.3)	232.5 (177.7)	-9.1 (-96.1, 77.9)	0.60	↓	0.829		
Rheumatic complaints	39	13.3 (22.9)	11.6 (20.7)	49	12.9 (16.2)	14.9 (27.5)	3.2 (-8.1, 14.6)	0.55	↓	0.551 *		
Prostate complaints	26	34.7 (24.6)	30.4 (29.3)	62	39.1 (27.4)	41.8 (29.5)	6.9 (-0.7, 14.6)	0.07	↓	0.072		
Chlamydia infections	30	11.4 (21.5)	14.7 (23.3)	58	7.4 (8.8)	5.8 (9.5)	-5.7 (-12.0, 0.5)	0.40	↑	0.069 *		
Thyroid dysfunction	29	204.8 (159.9)	191.3 (135.4)	59	163.3 (153.2)	191.7 (152.8)	36.6 (10.5, 62.7)	0.00	↓	0.007		
Perimenopausal complaints	24	5.7 (5.5)	3.8 (3.5)	64	6.1 (6.7)	7.3 (8.0)	3.2 (0.1, 6.4)	0.10	↓	0.046 *		

Numbers represent the number of tests prescribed per 6 months per 1000 patients. n=number of practices; * Adjusted difference between groups at end of intervention, corrected for baseline; * statistically significant interaction. Rheumatic complaints: P10: B (95% CI): -3.0 (-13.9, 8.0), p: 0.560; P90: B (95% CI): 14.8 (2.1, 27.4), p: 0.025. Chlamydia infections: P10: B (95% CI): -2.3 (-9.0, 4.4) p: 0.482; P90: B (95% CI): -10.4 (-17.6, -3.3) p: 0.006. Perimenopausal complaints: P10: B (95% CI): 0.7 (-3.0, 4.5), p: 0.696; P90: B (95% CI): 6.6 (2.4, 10.7), p: 0.002.

Table 4.6 Per-protocol analysis of change in prescribing rates, showing sum scores per clinical topic and drug. The desired direction of change is given in the column headed

Clinical topic	Intervention				Control				Direction	p
	n	Mean before (sd)	Mean after (sd)	n	Mean before (sd)	Mean after (sd)	B [#] (95% CI)	icc		
Anaemia	35	29.5 (22.5)	38.7 (25.6)	53	26.6 (27.6)	46.9 (54.2)	18.9 (-3.9, 41.7)	0.43	=	0.098
Iron preparations, oral										
Urinary tract infections	8	40.8 (24.0)	44.5 (29.7)	80	46.5 (36.3)	55.3 (44.8)	6.3 (-21.1, 33.8)	0.33	↓	0.623
Antibiotics UTI	26	68.5 (47.0)	79.5 (63.0)	62	65.5 (53.6)	125.9 (91.9)	28.5 (6.5, 50.7)	0.27	↓	0.016
Prostate complaints	26	57.5 (41.1)	66.5 (54.2)	62	54.0 (44.9)	101.5 (73.2)	44.2 (6.0, 82.3)	0.49	↓	0.027
Alpha receptor blockers	26	11.0 (7.5)	13.0 (11.7)	62	11.5 (11.0)	24.5 (22.9)	11.1 (2.1, 20.2)	0.05	↓	0.020
5 alpha reductase inhibitors	28	455.1 (294.7)	565.2 (303.7)	60	275.7 (253.8)	509.1 (433.0)	61.1 (15.0, 107.2)	0.11	↓	0.014 *
Stomach complaints	28	28.0 (22.4)	32.4 (26.8)	60	23.2 (22.35)	24.2 (22.6)	-2.4 (-21.5, 16.7)	0.68	↑	0.788
H2 antagonists	28	425.9 (276.1)	531.5 (283.8)	60	251.6 (234.0)	483.7 (414.7)	254.8 (60.1, 449.5)	0.66	↓	0.014 *
Proton pump inhibitors	28	0 (0)	0 (0)	60	0.1 (0.4)	0.1 (0.45)	0.1 (-0.1, 0.2)	0.01	↓	0.365
Misoprostol	28	1.2 (2.5)	1.4 (1.8)	60	0.9 (1.2)	1.1 (1.6)	0.1 (-0.7, 0.8)	0.20	↓	0.878 *
Triple therapy (ATC A02BD04)	29	82.2 (45.9)	99.2 (57.5)	59	86.0 (63.1)	128.5 (94.6)	12.6 (0.7, 24.4)	0.08	=	0.040
Thyroid dysfunction	29	76.7 (43.5)	91.1 (52.2)	59	79.5 (58.8)	118.3 (87.8)	25.2 (2.1, 48.3)	0.19	=	0.035
Thyroid hormones	29	5.5 (7.4)	8.2 (13.2)	59	6.5 (7.7)	10.1 (14.0)	0.2 (-4.2, 4.6)	0.11	↑	0.927
Antithyroid preparations	24	59.4 (31.9)	56.5 (25.9)	64	51.2 (40.0)	51.3 (40.0)	1.5 (-2.3, 5.4)	0.21	↓	0.414
Perimenopausal complaints	24	11.5 (8.0)	9.7 (6.2)	64	14.0 (14.0)	12.3 (12.4)	1.0 (-3.6, 5.5)	0.03	↓	0.612
Oestrogens, oral	24	5.1 (2.8)	5.9 (2.7)	64	3.2 (3.6)	6.5 (5.1)	2.5 (-0.9, 5.8)	0.31	↓	0.137
Contraceptives (>50y)	24	38.7 (24.2)	36.3 (21.0)	64	31.1 (24.4)	29.8 (24.4)	2.6 (-9.7, 14.9)	0.60	↓	0.652
Hormone replacement therapy	24	0.6 (1.5)	0.7 (2.7)	64	0.4 (1.5)	0.4 (1.6)	-0.3 (-1.4, 0.7)	0.11	↓	0.508 *
Clonidine	24	3.6 (3.7)	3.8 (4.1)	64	2.6 (3.0)	2.3 (3.3)	-0.7 (-2.0, 0.6)	0.00	↓	0.294 *
Tibolone										

'Direction'. Numbers represent the number of DDDs prescribed per 6 months per 1000 patients, n=number of practices; [#] Adjusted difference between groups at end of intervention, corrected for baseline; * statistically significant interaction: Thyroid dysfunction: P10: B (95% CI): 30.3 (15.1, 45.5), p: <0.001; P90: B (95% CI): 186.7 (88.7, 284.7), p: <0.001. Clonidine: P10: B (95% CI): 0.4 (-0.4, 1.2), p: 0.327; P90: B (95% CI): -1.3 (-2.2, -0.4), p: 0.005. Stomach complaints: P10: B (95% CI): 1.4 (-39.6, 42.6), p: 0.937; P10: B (95% CI): 0.9 (-0.8, 2.6), p: 0.297. Tibolone: P90: B (95% CI): 408.8 (341.1, 476.5), p: <0.001; P90: B (95% CI): -2.3 (-4.0, -0.6), p: 0.008. PPIs: P10: B (95% CI): 91.5 (-120.0, 303.0), p: 0.378; P10: B (95% CI): -0.8 (-1.9, 0.3), p: 0.133. Triple therapy (ATC A02BD04): P90: B (95% CI): 347.4 (159.2, 0.8), p: <0.001; P90: B (95% CI): 1.4 (0.3, 2.5), p: 0.014.

Discussion

Summary of the main findings

Our study found that the beneficial results obtained in earlier, well-controlled studies on audit and feedback with peer review in LQICs in primary care were not confirmed when we introduced this intervention in existing primary care LQICs. The per-protocol analyses showed that GPs from practices with the highest baseline volumes on test-ordering and prescribing showed the largest improvements.

Many factors can be identified as possible explanations for the lack of overall effects of our intervention. These factors can be divided into three different areas:

Participant-related factors

Lack of confidence: the confidence in our intervention was lower than we had expected, resulting in less commitment and a lack of adherence to the intervention. Two of the most important reasons for this lack of confidence were a distrust among participants towards the origin and validity of the feedback, and the perception that the intervention was too complex and ambitious.²⁵

Lack of motivation: cost savings or reduction of testing and prescribing rates could have had low priority for GPs. However, since the LQIC groups were created with the aim of discussing prescribing practices and policies for the future, a lack of motivation to critically review their own performance seems highly unlikely.

Sibley effect: the fact that the groups had much freedom of choice of topics could have influenced the results due to the so-called Sibley effect. This effect occurs when GPs choose those topics for quality improvement for which they already show good performance.³³ Since we used existing groups which were not established on the basis of shared interests but rather on geographical grounds (practices from the same neighbourhood or village) the risk of a Sibley effect seems minimal.

Lack of attendance: as the effects of an intervention will normally be larger for participants who are actually exposed to it, low attendance rates will most likely have a negative impact on the effect.³⁴ However, we assume that attendance to the LQIC meetings was positively influenced by the fact that continuing medical education (CME) credits were awarded for attending meetings, similar to the routines that had been used in the PTAM groups before. We have no signs from our process evaluation that attendance was worse than in the earlier Verstappen trial.

Failed implementation of agreements in practice: it is possible that the LQICs did not use effective implementation strategies to translate working agreements into everyday practice. We received group session reports for 76 of the 115 meetings. Over half of

these reports, however, did not contain specific, 'smart' working agreements that could be evaluated and analysed in a quantitative manner. For example, the reports mentioned working agreements such as 'our group will try to be more compliant with the guideline.'

Context-related factors

A failure to identify important barriers: after the intervention was completed, we analysed the barriers and facilitators to its implementation as perceived by the participants.²⁵ Beforehand, we had simply copied the information on barriers and facilitators from the previous trial by Verstappen 10 years earlier. This proved to be insufficient, as many health care system innovations have been introduced in these 10 years. The Dutch health care system has become more market-regulated and many tasks have been delegated from hospital care to primary care, with Dutch GPs earning a higher income but at the same time feeling threatened in their autonomy and time management. Last but not least, the GPs are more than ever controlled by external parties such as the health inspectorate and healthcare insurers, which may have led to a more defensive attitude among GPs, resulting in higher test ordering rates.

Intervention-specific factors

Failure to identify true opinion leaders: diagnostic facilities for primary care in the Netherlands are often linked to hospitals. The laboratory specialists are often consulted when questions arise on how to interpret test results. Although the role of laboratory specialists or local pharmacists in LQICs is generally undisputed, it is possible that they are not always considered to be opinion leaders by LQIC members.

Source of the feedback: despite complete transparency on the origin of the data we did find that the source of our feedback was not clear to the participants. Despite the training provided to the moderators, many GPs remained suspicious about data stemming from insurance company databases. This critical attitude seems to be reinforced by the GPs' feeling of being controlled by external parties more than ever before.

Insufficient provision of feedback: the criteria for effective audit and feedback as defined by Ivers et al. were only partially met, as we did not exclusively include practices with poor baseline performance, we did not provide repeated feedback according to their criteria and we did not provide predefined goals.³⁵ If we had identified those GPs or practices that showed poor baseline adherence for inclusion in our study, however, this would have forced us to leave the safe and stable environment of the existing LQICs. Also, we would have lost the peer review effect where poorly performing GPs can learn from role models. Setting up new groups would have

introduced the negative effects of an organisational reform.^{8,36,37} In any case, the feedback was provided more than once, as each topic was discussed in two meetings, one on test ordering and one on prescribing on the same topic. Although we did not provide concrete predefined goals, the provision of aggregated results from their own and neighbouring groups, together with the recommendations from clinical guidelines, can be regarded as implicit goal setting.

In their recently published article on how to provide feedback best Brehaut et al. provide 15 suggestions for designing and delivering effective feedback.³⁸ The feedback we provided meets all these suggestions but three. First providing the feedback as soon as possible at optimal intervals was not possible in our trial as we provided feedback on demand of the LQICs. The second suggestion we missed is to provide the feedback in more than one way, this could indeed have been helpful. The last suggestion we (partially) missed is to provide short key messages with optional extra detail being available. This was not possible for us as we chose to use peer review as a means to discuss the feedback. Would we have provided key messages on individual feedback we would have removed the peer review process, an essential part of LQIC work.

Strengths and limitations of the research methods used

Lack of power: our efforts to implement the strategy widely in the southern part of the Netherlands failed to recruit a sufficient number of groups for the trial, leaving us with an underpowered study. This could, in part, have been caused by the pragmatic character of our trial, with local pharmacists and experts on diagnostics leading the recruitment effort and moderating the groups. A major healthcare reform programme was launched shortly after our recruitment started,³⁹ causing frustration among many GPs due to the resulting high administrative burden. This probably reduced their willingness to participate in our trial.

Choice of outcome and lack of quality indicators: we chose to express the volume of prescribed drugs in DDDs. A risk of this is that not all DDDs are compatible with the actual dosage physicians prescribe to a patient. For diclofenac, for instance, the normal dosage is 1.5 to 2 DDDs every day. However, this did not affect the comparability of the two groups, as both were affected by this form of distortion in the same way. If we had been able to provide feedback on quality indicators as well as volume data, a more valid insight into performance might have resulted, but with more interpretation problems for the GPs. Also, using volume data only has been proven to lower volumes especially in areas characterised by overuse.²²

Change of the protocol after its publication: the fact that all groups in the intervention arm, whether or not they had chosen a specific topic, were analysed using the intention-to-treat principle as if they had been exposed to the topic may have diluted

the effects of the intervention. We would rather have analysed the effect of the intervention on changes in the direction of shared working agreements, as stated in the protocol, but as these were hardly established, we decided to use a per-protocol analysis as the second best option, after the protocol had been published.

Minimization of the Hawthorne effect: a strong point of the design we used is that it minimized the Hawthorne effect. On the downside, it might have caused contamination of the effects if we had not exposed all groups in the trial to the intervention at the same time.^{30,40,41}

Comparison with other studies

Evaluating the effect of large-scale implementation of a quality improvement strategy of proven effectiveness using a pragmatic design like ours has not often been performed. Earlier and ongoing work has focused mainly on one particular clinical topic (e.g., prescribing antibiotics for respiratory tract infections or X-rays for low back pain patients), while we applied the peer review strategy to a broad range of topics and focussed on both test ordering and prescribing behaviour.⁴²⁻⁴⁶ By researching whether the results of more fully controlled trials were also found in large-scale implementation, we sought to contribute to the knowledge on ways to improve professional performance.

We are not aware of similar multi-faceted studies using audit and feedback with peer group discussion in this field that would allow direct comparison with our study, although much is known about the individual components we combined in our study.

Much work has been done on evaluating the effects of audit and feedback on both test ordering and prescribing behaviour in well-controlled trials. These interventions show a modest but statistically significant positive effect on changing professional behaviours. However, no solid conclusions can be drawn from these studies, given the difficulties of interpretation due to the heterogeneity of the trials.^{34,35,42,47-52} Although audit and feedback on test ordering behaviour embedded in peer review in small groups has been found to be more effective than audit and feedback alone, it generally remains unclear exactly what factors contribute to this effect.^{10,21,53,54} The use of pragmatic designs in quality improvement research contributes to bridging the gap between academia and clinical practice.^{55,56}

Multifaceted interventions like ours are complex by nature but seem attractive because the individual effects could add up. It remains unclear, however, whether multifaceted interventions or single interventions are more effective. Mostofian concluded in a review of reviews that multifaceted interventions are most effective in changing professional behaviour.⁵⁷ On the other hand, Irwin et al. concluded that there is no

evidence for a larger effect of combined interventions, while Johnson and May find it likely that multifaceted interventions are more effective.^{52,54}

Studies embedding the discussion of clinical topics in LQICs have reported a modest positive effect on prescribing costs and quality.^{15,58-62} Our finding, based on the per-protocol analysis, that groups with the highest volumes at baseline showed the largest improvement, is in line with the results presented by Irwin et al..⁵²

Conclusions

Our intervention, which aimed at changing the test ordering and prescribing behaviour of GPs by means of auditing and feedback, embedded in LQICs, with academia at a distance, shows that the favourable results of earlier work could not be replicated. It appeared that large-scale uptake of evidence-based but complex implementation strategies with a minimum of influence of external researchers, but with the stakeholders in healthcare themselves being responsible for the work that comes with integrating this intervention into their own groups, was not feasible. Although our study suffered from a lack of power, we expect that even if a sufficient number of groups had been included, no clinically relevant changes would have been observed.

Implications for future research

The problems on the fidelity of the feedback and with the uptake of the intervention could best be handled by assuring that a strong leader picks up the group and lead them forward. It may also be helpful to identify GPs with a low quality baseline performance representing an unwarranted deviation from the mean and target those GPs in this type of quality improvement initiatives. Other physicians who are already doing well can concentrate on what they are doing already; delivering high quality care. Further research is needed on whether low baseline performance is consistent behaviour for an individual GP. Also further research on the cut-off point for participants that can benefit from a QI intervention like this is needed to clarify the population to be targeted best. Potential downsides of such an approach such as the loss of peer learning with learning from the best practices need to be addressed as well. Further pragmatic research should be performed to confirm our findings that the results found in earlier well-controlled trials are not easily replicated. We therefore encourage other researchers to perform vigorous large scale evaluations of complex implementation strategies, preferably embedded and owned by the field, as we did.

References

1. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA* 2012; 307(14):1513-1516.
2. Huisarts - Prescripties naar leeftijd [GP: prescriptions per agegroup] [www.nivel.nl/node/3099]
3. Westert G, van den Berg M, Zwakhals S, Heijink R, de Jong J, Verkleij H: Zorgbalans [balanced care], vol. 2014: Bohn Stafleu Van Loghum; 2010.
4. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. *PLoS One* 2013;8(11):e78962.
5. Walraven van C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA* 1998;280(6):550-558.
6. Harteloh PPM. kwaliteit van zorg: van zorginhoudelijke benadering naar bedrijfskundige aanpak [Quality of care: from a care standpoint towards a business management standpoint], 4 edn. Maarssen: Elsevier/ De Tijdstroom; 2001.
7. Braspenning JCC, Schellevis FG, Grol R. Kwaliteit huisartsenzorg belicht [Quality in primary care reviewed]: Nivel; 2004.
8. Burgers JS, Grol RPTM, Zaat JOM, Spies TH, van der Bij AK, Mokkink HGA. Characteristics of effective clinical guidelines for general practice. *Br J Gen Pract* 2003;53(486):15-19.
9. Muijters PE, Grol RP, Sijbrandij J, Janknegt R, Knottnerus JA. Differences in prescribing between GPs. Impact of the cooperation with pharmacists and impact of visits from pharmaceutical industry representatives. *Fam Pract* 2005;22(6):624-630.
10. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T: Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21(4):387-395.
11. Grol R. Implementing guidelines in general practice care. *Qual Health Care* 1992;1(3):184-191.
12. Martens JD, van der Weijden T, Severens JL, de Clercq PA, de Bruijn DP, Kester AD, Winkens RA. The effect of computer reminders on GPs' prescribing behaviour: a cluster-randomised trial. *Int J Med Inform* 2007;76 Suppl 3:S403-416.
13. Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijters P, Winkens R, van Steenkiste B, Grol R, Metsemakers J. A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: Study Protocol. *Implement Sci* 2009; 4:6.
14. Verstappen WHJM, van Merode F, Grimshaw J, Dubois WI, Grol RPTM, van Der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care* 2004;16(5):391-398.
15. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ* 2008;336(7659):1491-1494.
16. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Niessen LW, Grol RP. Short- and long-term effects of a quality improvement collaborative on diabetes management. *Implement Sci* 2010;5:94.
17. Beyer M, Gerlach FM, Flies U, Grol R, Krol Z, Munck A, Olesen F, O'Riordan M, Seuntjens L, Szecsenyi J. The development of quality circles/peer review groups as a method of quality improvement in Europe. Results of a survey in 26 European countries. *Fam Pract* 2003;20(4):443-451.
18. Peck C, McCall M, McLaren B, Rotem T. Continuing medical education and continuing professional development: international comparisons, vol. 320; 2000.
19. Riou F, Piette C, Durand G, Chaperon J. Results of a 12-month quality-circle prescribing improvement programme for GPs. *Br J Gen Pract* 2007;57(540):574-576.
20. Florentinus SR, van Hulten R, Kloth ME, Heerdink ER, Griens AM, Leufkens HG, Groenewegen PP. The effect of pharmacotherapy audit meetings on early new drug prescribing by general practitioners. *Ann Pharmacother* 2007;41(2):319-324.
21. Verstappen WHJM, van der Weijden T, Dubois WI, Smelee I, Hermesen J, Tan FES, Grol RPTM. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med* 2004;2(6):569-575.
22. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smelee I, Hermesen J, Grimshaw J, Grol RPTM. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289(18):2407-2412.

23. Lagerlov P, Loeb M, Andrew M, Hjortdahl P. Improving doctors' prescribing behaviour through reflection on guidelines and prescription feedback: a randomised controlled study. *Qual Saf Health Care* 2000;9(3):159-165.
24. Feldman LS, Shihab HM, Thiemann D, Yeh HC, Ardolino M, Mandell S, Brotman DJ. Impact of providing fee data on laboratory test ordering: a controlled clinical trial. *JAMA Intern Med* 2013;173(10):903-908.
25. Trietsch J, van Steenkiste B, Hobma S, Frericks A, Grol R, Metsemakers J, van der Weijden T. The challenge of transferring an implementation strategy from academia to the field: A process evaluation of local quality improvement collaboratives in Dutch primary care using the normalization process theory. *J Eval Clin Pract* 2014;20(6):1162-1171.
26. Eimers M, de Groot J. FTO peiling 2005, kwaliteit van farmacotherapieoverleg in Nederland in beeld [PTAM audit 2005, quality of pharmacotherapeutic audit meetings in the Netherlands visualised]. In.: DGV Nederlands Instituut voor Verantwoord Medicijngebruik [DGV the Dutch Institute for Rational Use of Medicine]; 2006.
27. Eimers M, van der Aalst A, Pelzer B, Teichert M, de Wit H. Leidt een goed FTO tot beter voorschrijven? [Does an effective PTAM lead to more appropriate prescribing?]. *Huisarts en Wetenschap* 2008; 51(7):340-345.
28. Florentinus SR, van Hulten R, Kloth ME, Heerdink ER, Griens AM, Leufkens HG, Groenewegen PP: fout: The effect of pharmacotherapy audit meetings on early new drug prescribing by general practitioners. *Ann Pharmacother* 2007;41(2):319-324.
29. Teichert M, van der Aalst A, de Wit H, Stroo M, De Smet PA. How useful are prescribing indicators based on the DU90% method to distinguish the quality of prescribing between pharmacotherapy audit meetings with different levels of functioning? *Eur J Clin Pharmacol* 2007;63(12):1171-1177.
30. Verstappen WHJM, van der Weijden T, ter Riet G, Grimshaw J, Winkens R, Grol RPTM. Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. *J Clin Epidemiol* 2004;57(11):1119-1123.
31. Trietsch J, Leffers P, van Steenkiste B, Grol R, van der Weijden T. The balanced incomplete block design is not suitable for the evaluation of complex interventions. *J Clin Epidemiol* 2014;67(12):1295-8
32. van Ginneken E. Perennial Health Care Reform--The Long Dutch Quest for Cost Control and Quality Improvement. *N Engl J Med* 2015;373(10):885-889.
33. Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick V, Fraser W. A randomized trial of continuing medical education. *N Engl J Med* 1982;306(9):511-515.
34. Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien Mary A, Wolf F, Davis D, Odgaard-Jensen J, Oxman Andrew D. Continuing education meetings and workshops: effects on professional practice and health care outcomes. In: Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2009.
35. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O'Brien MA, Johansen M, Grimshaw J, Oxman AD. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012, 6:CD000259.
36. Grol R, Dalhuijsen J, Thomas S, Veld Cit, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998;317(7162):858-861.
37. Grol R, Thomas S, Roberts R. Development and implementation of guidelines for family practice: lessons from The Netherlands. *J Fam Pract* 1995;40(5):435-439.
38. Brehaut JC, Colquhoun HL, Eva KW, Carroll K, Sales A, Michie S, Ivers N, Grimshaw JM. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. *Ann Intern Med* 2016;164(6):435-441.
39. van Weel C, Schers H, Timmermans A. Health care in the Netherlands. *J Am Board Fam Med* 2012;25 Suppl 1:S12-17.
40. Godman B, Wettermark B, Miranda J, Bennie M, Martin A, Malmström RE. Influence of multiple initiatives in Sweden to enhance ARB prescribing efficiency following generic losartan; findings and implications for other countries. *Int J Clin Pract* 2013;67(9):853-862.
41. Konstantinou GN: Pragmatic trials: how to adjust for the 'Hawthorne effect'? *Thorax* 2012, 67(6):562.
42. van der Velden AW, Kuyvenhoven MM, Verheij TJ. Improving antibiotic prescribing quality by an intervention embedded in the primary care practice accreditation: the ARTI4 randomized trial. *J Antimicrob Chemother* 2016;71(1):257-63

43. Prior M, Elouafkaoui P, Elders A, Young L, Duncan EM, Newlands R, Clarkson JE, Ramsay CR. Evaluating an audit and feedback intervention for reducing antibiotic prescribing behaviour in general dental practice (the RAPID trial): a partial factorial cluster randomised trial protocol. *Implement Sci* 2014;9:50.
44. Mold JW, Fox C, Wisniewski A, Lipman PD, Krauss MR, Harris DR, Aspy C, Cohen RA, Elward K, Frame P et al. Implementing asthma guidelines using practice facilitation and local learning collaboratives: a randomized controlled trial. *Ann Fam Med* 2014;12(3):233-240.
45. Gjølstad S, Høy S, Straand J, Brekke M, Dalen I, Lindbaek M. Improving antibiotic prescribing in acute respiratory tract infections: cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ* 2013;347:f4403.
46. Bird GC, Marian K, Bagley B. Effect of a performance improvement CME activity on management of patients with diabetes. *J Contin Educ Health Prof* 2013;33(3):155-163.
47. Ho MJ, Venci J. Improving the success of mailed letter intervention programs to influence prescribing behaviors: a review. *J Manag Care Pharm* 2012;18(8):627-649.
48. Giguere A, Legare F, Grimshaw J, Turcotte S, Fiander M, Grudniewicz A, Makosso-Kallyth S, Wolf FM, Farmer AP, Gagnon MP. Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;10:CD004398.
49. O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in Australia, vol. 318; 1999.
50. Foy R, Eccles MP, Jamtvedt G, Young J, Grimshaw JM, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC Health Serv Res* 2005;5:50.
51. Ivers NM, Tu K, Young J, Francis JJ, Barnsley J, Shah BR, Upshur RE, Moineddin R, Grimshaw JM, Zwarenstein M. Feedback GAP: pragmatic, cluster-randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care. *Implement Sci* 2013;8:142.
52. Irwin R, Stokes T, Marshall T. Practice-level quality improvement interventions in primary care: a review of systematic reviews. *Prim Health Care Res Dev* 2015:1-22.
53. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, Grimshaw JM. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci* 2014;9:14.
54. Johnson MJ, May CR. Promoting professional behaviour change in healthcare: what interventions work, and why? A theory-led overview of systematic reviews. *BMJ Open* 2015;5(9):e008592.
55. Williamson M, Cardona-Morrell M, Elliott JD, Reeve JF, Stocks NP, Emery J, Mackson JM, Gunn JM. Prescribing Data in General Practice Demonstration (PDGPD) project - a cluster randomised controlled trial of a quality improvement intervention to achieve better prescribing for chronic heart failure and hypertension. *BMC Health Serv Res* 2012;12(1):273.
56. Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract* 2000;17(2):192-196.
57. Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing physician behavior: what works? *Am J Manag Care* 2015;21(1):75-84.
58. Wensing M, Broge B, Riens B, Kaufmann-Kolle P, Akkermans R, Grol R, Szecsenyi J. Quality circles to improve prescribing of primary care physicians. Three comparative studies. *Pharmacoepidemiol Drug Saf* 2009;18(9):763-769.
59. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Niessen LW, Grol RP. fout: Short- and long-term effects of a quality improvement collaborative on diabetes management. *Implement Sci* 2010;5:94.
60. Veninga CC, Denig P, Zwaagstra R, Haaijer-Ruskamp FM. Improving drug treatment in general practice. *J Clin Epidemiol* 2000;53(7):762-772.
61. Wensing M, Broge B, Kaufmann-Kolle P, Andres E, Szecsenyi J. Quality circles to improve prescribing patterns in primary medical care: what is their actual impact? *J Eval Clin Pract* 2004;10(3):457-466.
62. Engels Y, Verheijen N, Fleuren M, Mookink H, Grol R. The effect of small peer group continuous quality improvement on the clinical practice of midwives in The Netherlands. *Midwifery* 2003;19(4):250-258.
63. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.

Appendix 4

The complete set of texts sent to GPs together with the feedback.

NHG = Dutch College of General Practitioner.

Module prostate complaints

General

Difficulty in micturition is neither an early symptom nor risk factor for prostate cancer. Difficulty in micturition is caused by obstruction of the enlarged prostate, obstruction by smooth muscle of the bladder neck and/ or bladder dysfunction. Prostatic hyperplasia can occur from the 3rd decade. 90% of the patients with symptoms is 80 years and older. The size of the prostate does not correlate with the presence and severity of symptoms. All of the following opinions and statements relate to men over 50 years of age.

Incidence and prevalence of BPH

	incidence (/1000 men)	Prevalence (/1000 men)
General population of men	2-4	4-6
45-64 years	4-9	8-19
65-74 years	8-24	24-33
75 years and older	10-18	18-36

Because of the different definitions of BPH, there are large differences in incidence and prevalence between studies. In Dutch general practice the prevalence until 75 years is approx. 80/ 1000 patientyears, older than 75 years approx. 165/ 1000 patientyears.

Part A, diagnostics

General

In all patients with LUTS an underlying urinary tract infection should be excluded. Diagnosis and treatment of urinary tract infections are beyond the scope of this module. There is no biochemical marker to diagnose BPH or prostate cancer nor to rule it out. The history and physical examination are the most important criteria to diagnose BPH.

PSA € 8,31

The prevalence of prostate cancer in the general population is 2-5%.

PSA is formed by the prostate epithelium and to a very small extent by other tissues. Increased by: prostate cancer, prostatitis, androgen use, BPH and perhaps after DRE and biopsies.

An abnormal DRE is an indication a referral to secondary care. Men whose prostate cancer is detected by screening have a slightly higher life expectancy than men with prostate cancer diagnosed otherwise.

An abnormal DRE and PSA<4 ng/ml results in 10-20% probability on prostate cancer.

A PSA 2-4 ng/ml gives a 15% chance of having prostate cancer. A PSA>10 ng/ml correlates with a probability of 10% to >50% of prostate cancer.

Free/Total PSA-ratio € 8,31

This test increases the sensitivity of the PSA determination if its result is between 4-10mg / l. In this group the risk is as follows:

V/T PSA >0,25:	8% probability on prostate cancer being present
V/T PSA 0,20-0,25:	16% probability on prostate cancer being present
V/T PSA 0,15-0,20:	20% probability on prostate cancer being present
V/T PSA 0,10-0,15:	28% probability on prostate cancer being present
V/T PSA <0,10:	56% probability on prostate cancer being present

Creatinine € 1,41

In general malaise, with recurrent urinary tract infections or urinary creatinine levels should be determined. This allows the GFR to be calculated.

Several methods are used to estimate the GFR. The Cockcroft-Gault and MDRD methods are the most used. Both are vulnerable to confounding factors (BMI, age, gender) and should be interpreted cautiously.

BPH can cause stowage and eventually renal failure.

Module prostate complaints

Part B, pharmacotherapy

General

Refer for invasive therapy if:

- The patient wishes it because of the severity of symptoms
- In recurrent acute urinary retentions
- In recurrent urinary tract infections
- When renal dysfunction is diagnosed and / or hydronephrosis
- When drug therapy has proven to be ineffective

Combination therapy of α 1-receptor blockers with 5 α -reductase inhibitors has not been proven more effective than treatment with α 1-receptor blockers alone.

Selective α 1-receptor blockers

Group of first choice according to the NHG guideline. Influences the muscle tone in the prostate and the urinary tract. The greatest effect is achieved within 2 weeks after the start of the therapy. The urinary flow improves by 20-30% this medication, the symptom score improves by 20%. These effects are independent of the size of the prostate. First administration may cause orthostatic hypotension. Severe hepatic impairment constitute a contraindication for the use of α 1-receptor blockers

Alfuzosin € 15,89 - € 26,92/month (1 DDD = 7,5 mg)

Drug of first choice according to the NHG guideline

Tamsulosin € 23,72/month (tablet), €12,36/month (caps) (1 DDD = 0,4 mg)

Drug of first choice according to the NHG guideline. Evidence is limited to trials with a short follow-up.

Doxazosin € 10,52 - € 13,15/month (1 DDD = 4 mg)

As effective as Alfuzosin and Tamsulosin, however non-compliance is higher due to side effects

Terazosin € 11,73 - € 12,11/month (1 DDD = 5 mg)

As effective as Alfuzosin and Tamsulosin, however non-compliance is higher due to side effects

Prazosin € 7,53/month (1 DDD = 5 mg)

Not mentioned in the NHG guideline

5 α -reductase blockers

Slowing down the conversion of testosterone to dihydrotestosterone in the prostate. This results in less hyperplasia. In case of relatively large prostates (> 35-40ml), the risk for acute urinary retention and the need for a surgical intervention is reduced.

NHG guideline: not to be administered in primary care. Indicated when the prostate is large enough, in primary care the determination of the prostate size is not sufficiently reliable.

Dutasteride € 30,40/month (1 DDD = 0,5 mg)

The clinical effect is limited. Noticeable effect occurs after 12 months (subjective symptoms), effect on the flow can be detected earlier. The effect is noticeable with a prostate volume >30ml but greater impact when >40ml.

Finasteride € 26,09/month (1 DDD = 5 mg)

Improves symptom scores significantly. The clinical effect is limited in the short and medium term and fully reversible. In a small volume prostate symptom score improved no more than with placebo. Noticeable effect after 2-6 months.

Module chlamydia trachomatis

General

PID: pelvic inflammatory disease, usually based on a salpingitis or adnexitis.

The incidence of Chlamydia trachomatis is 20 / 10,000.

The incidence of PID in women is 10 / 10,000

The incidence of PID in women caused by chlamydia is 5 / 10,000

In 13-40% PID leads to infertility (depending on number of infections) in 9% to EUG

The chance of infection after single or repeated contact is identical, 50-70%

The prevalence is 3-5% in young adults in the general population (<26years) and higher in young Surinamese, Antilleans and Arubans (up to 22%).

Incubation period: 1-3 weeks

After 1 year, 50% of the infections are no longer to traceable

After positive test: proactively perform partner detection up to 6 months prior to detection.

Part A, diagnostics

General

Physical examination with a suspected chlamydia often will yield no abnormalities. In the anamnesis also focus on anal and oropharyngeal complaints as well as complaints including post coital bleeding, spotting and abdominal complaints.

DNA amplification, urine € 44,59-88,79

Suitable for detecting Chlamydia in men. In women only suited for asymptomatic infections. The test result remains positive until 3-4 weeks after completing a treatment successfully. (NHG guideline speaks of 2 weeks).

DNA amplification, swab € 44,59-88,79

Suited for collection in women. A swab of the cervix together with a swab from the urethra is preferred (2 swabs). This way an isolated urethritis is can't be missed. An alternative for a cervical swab is a vaginal swab, this can be performed by the patient herself. The test result remains positive until 3-4 weeks after completing a treatment successfully. (NHG guideline speaks of 2 weeks).

IgG € 8,31

Low sensitivity in asymptomatic infections. High sensitivity in active PID, pneumonia or lyfogranuloma venerum.

Test is specific to the genus; therefore the results can be difficult to interpret.

After the introduction of DNA amplification techniques, this test has become obsolete for detecting active chlamydia infection.

Source: NHG guideline M82, the STD consultation. M50, PID. NHG LESA test ordering rationally 2006. CBO guideline STD and herpes neonatorum 2002. Diagnostic compass 2003. 2006 UK national guideline for the management of genital tract infection with chlamydia trachomatis. Sexually transmitted diseases treatment guidelines, 2006 CDC, Atlanta, VS.

Module chlamydia trachomatis

Part B, pharmacotherapy

General

Treatment with a single dose: 7 days no sexual intercourse

Treatment with a 7 day regime: no sexual intercourse until the treatment has been finished

Re-testing directly after the treatment has been finished has no value as the results remain positive for 2-4 weeks. Re-testing after 3-6 months is advisable. In 10% of the cases a re-infection occurs

Azithromycin € 9,05/ 1500mg (1DDD = 300 mg)

Drug of first choice according to the NHG guideline. A single administration of 1gr cures >95% of the infections. Is preferred over doxycycline because of a higher compliance rate. Can be administered during pregnancy (since august 2007). Therefore also drug of first choice in pregnant women.

Doxycycline €1,86-2,16/ week (1DDD = 100mg)

Drug of second choice according to NHG guideline. Treat for 7 days twice daily with 100mg. Do not prescribe to pregnant women.

Amoxicillin € 4,75/ 1125mg (1DDD = 1000 mg)

According to the NHG guideline, first choice in pregnant women. Treat 7 days 3 times daily with 500mg. 70% chance on cure of an infection.

NB: this advice is no longer valid; azithromycin has become drug of first choice in pregnancy as well.

Erythromycin €4,88-6,90/ 1500mg (1 DDD = 1000 mg)

Drug of second choice in pregnancy according to the NHG guideline, 4 times 500mg. also to be prescribed in cases of allergies. The chances of a successful treatment are 70%. In this case a re-test after 3 weeks is indicated.

Source: NHG guideline M82, the STD consultation. M50, PID. CBO guideline STD and herpes neonatorum 2002. 2006 UK national guideline for the management of genital tract infection with chlamydia trachomatis. Sexually transmitted diseases treatment guidelines, 2006 CDC, Atlanta, VS. pharmacotherapeutic compass 2007.

Module rheumatic complaints

Notions

Prevalence RA open population:	10/ 1000
Prevalence RA GP, men:	2/ 1000
Prevalence RA GP, women:	5/ 1000
Incidence RA:	1-2/ 1000/ year/ GP
Men: women:	1:3 (under 45 years) 5:6 (65+)

ACR criteria for the classification of RA (note: these are not diagnostic criteria!)

Main criteria (only valid when present more than 6 weeks)

1. Morning stiffness, longer than one hour
2. arthritis simultaneously present in three or more joint areas (left or right PIPs, MCPs, wrist, elbow, knee, ankle, MTPs)
3. arthritis in at least one hand joint: wrist, MCP or PIP
4. symmetric arthritis

secondary criteria

5. subcutaneously rheumatic nodules
6. serum rheumatoid factor (either test)
7. X-ray proven changes (X-hand/ wrist or foot)

A patient is said to have RA if 4 out of 7 criteria are satisfied

Part A, diagnostics

General

Rheumatoid arthritis is a clinical diagnosis. For diagnosing rheumatoid arthritis diagnostic testing of little significance. For the differentiation of rheumatism and for instance gout testing might be important or when a patient does not fully meet the diagnostic criteria.

Serum rheumatoid factor

Qualitative or semi-qualitative methods such as the latex fixation test and Waaler-Rose test are being replaced by the quantitative IgM test. The Waaler-Rose test adds nothing anymore.

Waaler-Rose € 11,19 (indication)

Latex fixation test € 7,62 (indicatief)

IgM- rheuma factor € 16,63

Quantitative test, test results of different laboratories are comparable. IgM-RF seems to fit better for criterion 6.

Anti-CCP price depends on laboratory

Antibody titer, the sensitivity matches IgM-RF sensitivity, has a very high specificity. In juvenile RA anti-CCP is often undetectable. NHG literature does not mention this test.

ESR € 1,41

ESR testing can sometimes help to differentiate between RA and a non-inflammatory disease. An elevated ESR makes it more likely that RA is present, a normal ESR does not exclude RA. An elevated ESR is especially useful in assessing disease activity.

CRP € 3,88

Acute phase protein, can be elevated in chronic inflammatory diseases as well. ESR is a better proxy for the severity of the disease. CRP is also suitable for assessing disease activity. Testing ESR and CRP together adds nothing.

X-wrist/ foot/ hand € 40,40 - 47,20

Indicated with diagnostic uncertainty after taking the history and a careful examination and when serology tests return negative. The severity of radiological abnormalities correlate poorly with symptoms and functionality

AST € 7,22

Measuring antistreptolysin levels is pointless when diagnosing RA. Antibodies against group A, C of G streptococci are produced starting in the first week of the disease. Levels start to decrease after 6 weeks and return to normal levels within months. AST levels in healthy individuals are age dependent.

Module rheumatic complaints

Part B, pharmacotherapy

General

The GP will usually prescribe an NSAID when RA is suspected.

Initiating and monitoring DMARD prescriptions will usually be done by the rheumatologist. When a GP prolongs a prescription of DMARDs he or she shares the responsibility for the monitoring.

NSAIDs (cox-1) are mentioned as separate drugs on your own feedback. A part of this group is aggregated to a single group as the prescription volumes are very low.

Note: medication to protect the stomach is beyond the scope of this module

Ibuprofen/ Naproxen € 5,44-15,54/ 4,24-6,20/month (1DDD = 1200mg/ 500mg)

Propionic acid derivatives.

Ibuprofen is the least harmful to the stomach of all classic NSAIDs, with high dosages this advantage diminishes. Drug of first choice together with diclofenac.

Naproxen increases the risk for stomach bleeding more than ibuprofen and diclofenac.

Indometacin/ Diclofenac € 2,96-5,81/ 5,98-12,47/ mth (1DDD = 100mg/ 100mg)

Acetic acid derivatives

Diclofenac is the least harmful for the stomach, except for ibuprofen; both are drugs of first choice. Indomethacin is comparable with naproxen considering the stomach

other NSAIDs (cox-1) (DDDs depend on drug of choice)

None of the other NSAIDs is preferred. Based on past experience, for harm to the stomach and price the preference is for ibuprofen, diclofenac and naproxen.

Meloxicam/ Nabumetone/ Piroxicam € 9,32-14,26/ 10,03-15,56/ 5,77-9,86/ month (1DDD = 15mg/ 1000mg/ 20mg)

Are considered preferential cox-2 inhibitors because of their effects on COX-1 as well. In higher dosages COX-1 and COX-2 are inhibited. Ranking position of these drugs remains unclear.

Piroxicam is of the same group as meloxicam. EMEA limited the use and lowered the maximum dosage of piroxicam, never to be considered a drug of first choice.

coxibs (cox-2)

Celecoxib/ Etoricoxib/ Lumiracoxib/ Parecoxib € 25,40/ 31,63/ 23,51/238,50/ month (1DDD = 200mg/ 60mg/ 100mg/ 40mg)

Theoretically these drugs should cause fewer problems with the stomach and the kidneys. The evidence for this claim is limited.

Het European Medicines Agency (EMA) considers the cardiovascular side effects of coxibs to be applicable to all coxibs. Coxibs are contraindicated in peripheral arterial vascular disease, ischemic heart diseases or with a history of CVA.

Parecoxib is only available for parenteral use. Experience with Lumiracoxib is limited. It seems to be suited for administration in arthrosis.

Diclofenac with Misoprostol € 17,78-23,22 (combination: no DDD given)

Misoprostol 800µg/day has proven to be more effective than protonpump inhibitors or H₂-antagonists in a doubled dosage. 800µg/day leads to much gastrointestinal side effects.

For the prevention of stomach ulcers in NSAID usage this combination is a sound option. The recommended dosage is: 3 times daily 1 tablet diclofenac/ misoprostol (50mg/200µg)

DMARD's (DDDs depend on drug of choice)

Most DMARDs are to be administered only by a specialist with specific knowledge and experience on these drugs. GPs however can be confronted with questions on side effects, interactions or the request for prolongation of these drugs. GPs should at least know and recognize the side effects of DMARDs.

Rheumatologists are to be consulted easily with questions concerning usage of DMARDs

common DMARDs:	Side effects	Interactions/ warnings
Sulfasalazine	Gastrointestinal complaints, headache, rash, disturbance of liver function, depressive feelings, leukocytopenia, agranulocytosis	
Methotrexate	Gastrointestinal complaints, disturbance of liver function, leukocytopenia, thrombocytopenia, pneumonitis, stomatitis, subcutaneous nodules	Trimethoprim containing drugs strongly increase the risk for leukocytopenia. Up to 3 month after ceasing use: ♀ avoid getting pregnant, ♂ avoid causing pregnancy
Leflunomide	Gastrointestinal complaints, disturbance of liver function, leukocytopenia, hypertension	Up to 2 years after ceasing use: ♀ avoid getting pregnant, ♂ avoid causing pregnancy
Etanercept, Infliximab, Adalimumab	No specific side effects. Possible increased risk for malignancies remains unclear.	Increased susceptibility for TBC, risk for reactivation of latent TBC. Mild bacterial infections can suddenly turn into severe disease.. always consult a rheumatologist in these cases.
Gold preparations, d-Penicillamine	Gastrointestinal complaints, leukocytopenia, thrombocytopenia, proteinuria, stomatitis, dermatitis	
(Hydroxy)chloroquine	Retinal disorders in long lasting higher dosages, intolerance for sunlight	
Azathioprine	Gastrointestinal complaints, leukocytopenia, thrombocytopenia, influenza like disease, disturbance of liver function	Up to 3 month after ceasing use: ♀ avoid getting pregnant, ♂ avoid causing pregnancy
Ciclosporin	Disturbances of kidney function, hypertension, hypertrichosis, gingival hyperthrophy	

Source: NHG guideline M41, reumatoïd artritis. Pharmacotherapeutic compass 2007. Pharmacotherapeutic directive pain relief, NHG 2005. updated consensus statement on biological agents for the treatment of rheumatic diseases, 2007. DE Furst, FC Breedveld et.al. Ann Rheum Dis 2007;66;2-22.

Module menopausal symptoms

Notions

Perimenopause: period from the start of irregular menses until 1 year after the last menses.

Post-menopause: means the period commencing one year after the last menses.

At the age of 51 years, has an average of 50% of the women reached menopause. Smoking women enter menopause an average of 1-1½ years earlier, women without a uterus average 1-2 years earlier.

Incidence perimenopausal symptoms: 10/1000 women / year

Prevalence perimenopausal symptoms: 80/1000 women / year (category: 54-64 yrs)

Average duration of symptoms: 4 years (range: a few months - 11 years)

Part A, diagnostics

General advice

Diagnose menopause is set on the basis of the anamnesis, **laboratory diagnostics adds nothing.**

In the diagnosis of menopausal symptoms, it is important to rule out other causes before diagnosing transition. Thyroid dysfunction and pregnancy can provide the same pattern of symptoms as the transition.

FSH € 6,65

Fluctuates during the cycle, rising gradually during life, in the menopause a stronger increase. Unsuitable in order to determine or to exclude the biochemical menopause at the individual level. After determination of a high FSH level normal menstruation may still occur. Even in women with no uterus FSH testing is not indicated.

Testing FSH levels is indicated for the diagnosis of primary or secondary amenorrhea.

LH € 6,65

Fluctuates during the cycle. Longitudinal studies showed the LH not to be a good parameter for the transition biochemically.

Even in women with no uterus LH testing is not indicated.

Estradiol € 8,31

Fluctuates during the cycle. Longitudinal studies showed estradiol not to be a good parameter for the transition biochemically. Even in women with no uterus testing Estradiol levels is not indicated.

When premature ovarian failure (<40 years) is suspected testing estradiol levels could be useful.

Source: NHG guideline M73, perimenopause. NHG LESA test ordering rationally 2006. Compass for diagnostics 2003.

Module menopausal symptoms

Part B, pharmacotherapy

General

The transition-related complaints of vaginal dryness and decrease libido seem related to lower estrogen levels. Explanation of this relation and the advice to take more time for foreplay and possibly the use of lubricantia may be a sensible advice to use.

Local estrogen (eg Synapause) is effective without an increased risk of cardiovascular disease, carcinoma of the breasts or endometrium. The risk of urinary tract infections also decreases with its use.

Estrogen alone (various formulations)

Highly effective against vasomotor symptoms resulting of the transition. Usage lowers the risk of fractures

A significant increase in the risk of CVD and thrombosis is present when started after the menopause (especially in the first year of use).

Unopposed estrogen therapy increases the risk of endometrial carcinoma and possibly reduces the risk of breast cancer.

Recently evidence was published showing that these agents also increase the risk of ovarian cancer.

contraceptives (oral and others) in women > 50 years

Not indicated as a therapy against symptoms consistent with the transition not even when continued from the pre-menopause. When a woman is using contraceptives already it seems sensible to agree on an age when usage will be stopped beforehand.

Combination preparations (various formulations)

Highly effective against vasomotor symptoms occurring as a result of the transition. The risk of fractures and colorectal cancer is reduced when using these drugs. 3 years after discontinuation of the medication, this effect will have disappeared. The risk of breast cancer is increased until 3 years after stopping these drugs; however the mortality appears not to be increased.

The risk of endometrial cancer is not changed (in sequential administration) or is decreased (in continuous administration). The risk of cardiovascular disease and thrombosis is increased until 3 years after discontinuation.

Tibolone € 18,74/ month (1 DDD = 2,5 mg)

Steroid with estrogenic, progestagenic and androgenic activity. Effective against vasomotor symptoms as a result of the transition. Works optimally after a few months. Is indicated for symptoms in the postmenopausal period. Use of tibolone increases the risk of occurrence of endometrial carcinoma.

The risk for breast cancer risk has not been established in well-designed research. It may improve sexual function something.

Clonidine € 7,67/ month (1 DDD = 0,45 mg)

Centrally acting antihypertensive agent. Only limited research is available. Clonidine might be effective against vasomotor symptoms as a result of the transition. Caution should be taken when given in combination with other antihypertensive drugs.

β-blockers can lead to a rebound hypertension after discontinuation of clonidine when prescribed at the same time. Usual dosage: 3 times 0,050mg daily.

Management of menopausal symptoms, benefits and risks.

	Effect on vasomotor symptoms	Risk of fractures	Risk of CVD	Risk of breast cancer	Risk of endometrial cancer	Risk of ovarian cancer	Risk of colon carcinoma	Risk of thrombosis
Estrogen alone	++	↓	↑ ¹	↓?	↑ ²	↑?		↑
Combination preparations	++	↓ ⁸	= ⁵	↑	= ³ / ↓ ⁴		↓ ⁷	= ⁶
Tibolone	+			↑?	↑			
Clonidine	+							

↓: decrease

↑: increase

=: unchanged

?: effects unclear

++: large improvement

+: small improvement

¹ when started after menopause, especially during the first year

² with the uterus in situ

³ sequential administration

⁴ continuous administration

⁵ 3 years after discontinuation, first 3 years increased risk

⁶ 3 years after discontinuation, first 3 years increased risk

⁷ 3 years after discontinuation a return to placebo risk levels

⁸ risk seems to return to baseline levels after 3 years

Clear boxes: no statistically significant effects found or has not been researched yet.

Source: NHG guideline M73, perimenopause. NHG manifest on contraceptive use in the menopause, june 2006. Pharmacotherapeutic compass 2007. Gerardo Heiss et al. *Health Risks and Benefits 3 Years After Stopping Randomized Treatment With Estrogen and Progestin*, JAMA 2008;299(9):1036-1045

Module dyslipidemia

Notions

Estimation of cardiovascular risk in the absence of CVD, and in the absence of DM2 is done using the SCORE table

Estimation of the cardiovascular risk in the presence of DM2 is done on the basis of the UKPDS risk engine; free download from: <http://www.dtu.ox.ac.uk/>

Assessing the mortality risk after a previous cardiovascular event is not covered in the guideline cardiovascular risk.

Part A, diagnostics

General

Laboratory tests for cholesterol and triglycerides are prone to errors. The most reliable results are obtained after 12 hours of fasting and in repeated measurements after 1-8 weeks. For determination of the risk profile of a patient a single measurement may be sufficient.

Total cholesterol € 1,41

Important for calculating the risk. Total cholesterol is reduced 10-20% after surgery, trauma or other acute stress (such as myocardial infarction).

HDL € 2,77

Important for calculating the risk. A relative limit applied by the NHG is 0,8mmol / l. Reference values of women are 0.2-0.3 mmol / l higher than that of men.

Homocysteine € 38,80

Hyperhomocysteinemia increases the risk of CVD. The elevated risk is similar to that of high cholesterol levels and smoking. Recent studies however show no effect on cardiovascular events after treatment of hyperhomocysteinemia. The diagnostic value therefore is still limited to the refinement of the risk profile.

Triglycerides € 1,89

Test results are highly dependent on factors such as exercise, type of food, time since last meal and prescribed medication (estrogens, corticosteroids, thiazides, propranolol, chlorthalidone). It is needed for calculating LDL levels when direct assay of LDL levels is not possible. Not a necessary test for the risk assessment.

LDL € 6,07
Usually indirect assay, is calculated by the laboratory on the basis of the Friedewald formula. If the triglyceride level is higher than 4.5, this formula is not reliable. The direct assessment is relatively insensitive to interference by high triglyceride levels.
This value is used to monitor the treatment, preferably a few weeks after the start of the therapy or no later than after 3 months.

Total cholesterol/ HDL ratio
This test is needed for assessing the mortality risk in accordance with the score table (see below).

SCORE table for calculating the mortality risk for patients without prior CVD and DM2

	Women						Men														
	Non smoking		smoking				Non smoking		smoking												
SBP	Age																				
180	8	10	11	13	14	15	18	20	23	26	65	13	15	17	20	22	23	27	31	35	38
160	8	7	8	9	10	11	13	15	17	19		9	11	13	14	16	17	20	23	26	29
140	4	5	8	7	7	8	9	11	12	14		6	8	9	10	12	12	15	17	19	21
120	3	3	4	5	5	5	7	8	9	10		5	6	7	7	8	9	11	12	14	16
180	4	5	6	7	8	8	10	11	13	14		7	9	10	12	13	14	16	19	21	24
160	3	4	4	5	5	6	7	8	9	10	60	5	6	7	8	9	10	12	14	16	17
140	2	3	3	3	4	4	5	6	7	7		4	5	5	6	7	7	9	10	11	13
120	1	2	2	2	3	3	3	4	5	5		3	3	4	4	5	5	6	7	8	9
180	2	3	3	4	4	4	5	6	7	8		4	5	6	7	8	8	10	11	13	15
160	2	2	2	3	3	3	4	4	5	5		3	4	4	5	6	6	7	8	9	11
140	1	1	2	2	2	2	3	3	3	4	55	2	3	3	4	4	4	5	6	7	8
120	1	1	1	1	1	1	2	2	2	3		2	2	2	3	3	3	4	4	5	5
180	1	1	2	2	2	2	3	3	4	4		3	3	4	4	5	5	6	7	8	9
160	1	1	1	1	2	2	2	2	3	3		2	2	3	3	3	4	4	5	6	6
140	1	1	1	1	1	1	1	2	2	2		1	2	2	2	2	3	3	4	4	5
120	0	1	1	1	1	1	1	1	1	2	50	1	1	1	2	2	2	2	3	3	3
180	1	1	1	1	1	1	1	2	2	2		1	1	1	1	2	2	2	2	3	3
160	0	0	1	1	1	1	1	1	1	1		1	1	1	1	1	1	2	2	2	2
140	0	0	0	0	1	1	1	1	1	1		0	1	1	1	1	1	1	1	1	2
120	0	0	0	0	0	0	0	1	1	1		0	0	0	1	1	1	1	1	1	1
	4	5	6	7	8	4	5	6	7	8		4	5	6	7	8	4	5	6	7	8
Total cholesterol/HDL-cholesterol ratio																					
	0 – 4% mortality risk by CVD					5 – 9% mortality risk by CVD					≥10% mortality risk by CVD										

Source: NHG guideline M84, cardiovascular risk management version 1.3 (2007). NHG LESA test ordering rationally 2006. Diagnostic compass 2003.

Module Dyslipidemia

Part B, pharmacotherapy

General

Initiate treatment when:

- In patients without a history of CVD or DM2 with a 10-year mortality risk over 10% in combination with LDL level > 2,5 mmol/l. When additional risk factors are present a lower mortality risk can justify initiation of treatment.
- In patients with a history CVD and/ or DM2 with LDL level > 2,5 or with a history of DM2 and a total cholesterol level > 4,5 mmol/l. When additional risk factors are present a lower mortality risk can justify initiation of treatment.

additional risk factors : - family history of CVD

- unhealthy diet
- little exercise
- BMI > 30 kg/m²
- waist circumference > 102 cm (♂) or 88 cm (♀)

In young patients with a history of DM2 and a favorable risk profile, higher treatment thresholds are held.

Therapy objectives:

- in patients without CVD or DM2: LDL < 2,5 mmol/l or a decrease of 1,0 mmol/l.
- In patients with CVD or DM2: LDL < 2,5 mmol/l.

Simvastatin € 8,31/ 11,13 (zocor) / month based on 15mg (15mg = 1DDD)

Drug of first choice according to the NHG guideline (along with others). Proven to be effective on hard endpoints. Initial dose 1dd 40mg.

Pravastatin € 14,52 - 19,68 (selektine) / month based on 20mg (20mg = 1DDD)

Drug of first choice according to the NHG guideline (along with others). Proven to be effective on hard endpoints. Initial dose 1dd 40mg.

Atorvastatin € 22,92/ month based on 10mg (10mg = 1DDD)

Mentioned in the NHG guideline. As effective as simvastatin and pravastatin.
Without a history of CVD and/ or DM2 not advised (NHG/ CBO).

Fluvastatin € 17,14/ month based on 40mg (40mg = 1DDD)

Mentioned in the NHG guideline, less effective in preventing new cardiovascular events.
Without a history of CVD and/ or DM2 not advised (NHG/ CBO).

Rosuvastatin € 23,93/ month based on 10mg (10mg = 1DDD)

Not yet proven to be effective on hard clinical endpoints. Only to be used when target cholesterol levels are not reached with other drugs.
Without a history of CVD and/ or DM2 not advised.

Bezafibrate € 10,92/ month based on 600mg (600mg = 1DDD)

Increases HDL levels, not yet proven to be effective on hard clinical endpoints.

Gemfibrozil € 13,93/ month based on 1200mg (1200mg = 1DDD)
Increases HDL levels, not yet proven to be effective on hard clinical endpoints.

Ciprofibrate € 19,66 - 38,06 (Hyperlipen) / month based on 100mg (100mg = 1DDD)
Increases HDL levels, not yet proven to be effective on hard clinical endpoints.

Acipimox € 26,90/ month based on 500mg (500mg = 1DDD)
Nicotic acid analog. Increases HDL levels, not yet proven to be effective on hard clinical endpoints.

Ezetimibe € 37,45/ month based on 10mg (10mg = 1DDD)
Increases HDL levels, not yet proven to be effective on hard clinical endpoints.
Decreases the reception of cholesterol and plant sterols from the small intestine.

Xantinol nicotinate € 8,76/ month based on 900mg (900mg = 1DDD)
Increases HDL levels, not yet proven to be effective on hard clinical endpoints. Is dissuaded in the pharmacotherapeutic compass for pharmacotherapeutic reasons because of insufficient proof of effectiveness.

Nicotic acid € 39,97/ month based on 2000mg (2000mg = 1DDD)
vitamin B7. Increases HDL levels, not yet proven to be effective on hard clinical endpoints.

Colestyramine € 33,51/ month based on 14g (14g = 1DDD)
Decreases the reception of cholesterol through the intestine. Increases HDL levels, not yet proven to be effective on hard clinical endpoints.

Ezetimibe/ Simvastatin € 46,85 – 57,78/ month
Nicotic acid analog combined with a statin. No convincing proof for effectiveness on hard clinical endpoints considering ezetimibe. The pharmacotherapeutic compass has not yet evaluated the effectiveness of this combination. 1 DDD = 1 tablet regardless of dosage.

SCORE table for calculating the mortality risk for patients without prior CVD and DM2

	Women						Men																	
	Non smoking			smoking				Non smoking			smoking													
SBP	Age																							
180	8	10	11	13	14		15	18	20	23	26	65	13	15	17	20	22		23	27	31	35	38	
160	8	7	8	9	10		11	13	15	17	19		65	9	11	13	14	16		17	20	23	26	29
140	4	5	8	7	7		8	9	11	12	14		65	6	8	9	10	12		12	15	17	19	21
120	3	3	4	5	5		5	7	8	9	10		65	5	6	7	7	8		9	11	12	14	16
180	4	5	6	7	8		8	10	11	13	14	60	7	9	10	12	13		14	16	19	21	24	
160	3	4	4	5	5		6	7	8	9	10		60	5	6	7	8	9		10	12	14	16	17
140	2	3	3	3	4		4	5	6	7	7		60	4	5	5	6	7		7	9	10	11	13
120	1	2	2	2	3		3	3	4	5	5		60	3	3	4	4	5		5	6	7	8	9
180	2	3	3	4	4		4	5	6	7	8	55	4	5	6	7	8		8	10	11	13	15	
160	2	2	2	3	3		3	4	4	5	5		55	3	4	4	5	6		6	7	8	9	11
140	1	1	2	2	2		2	3	3	3	4		55	2	3	3	4	4		4	5	6	7	8
120	1	1	1	1	1		1	2	2	2	3		55	2	2	2	3	3		3	4	4	5	5
180	1	1	2	2	2		2	3	3	4	4	50	3	3	4	4	5		5	6	7	8	9	
160	1	1	1	1	2		2	2	2	3	3		50	2	2	3	3	3		4	4	5	6	6
140	1	1	1	1	1		1	1	2	2	2		50	1	2	2	2	2		3	3	4	4	5
120	0	1	1	1	1		1	1	1	1	2		50	1	1	1	2	2		2	2	3	3	3
180	1	1	1	1	1		1	1	2	2	2	40	1	1	1	1	2		2	2	2	3	3	
160	0	0	1	1	1		1	1	1	1	1		40	1	1	1	1	1		1	2	2	2	2
140	0	0	0	0	1		1	1	1	1	1		40	0	1	1	1	1		1	1	1	1	2
120	0	0	0	0	0		0	0	1	1	1		40	0	0	0	1	1		1	1	1	1	1
	4	5	6	7			8	5	6	7	8		4	5	6		7		8		4	5	6	
	4						7		8															
Total cholesterol/HDL-cholesterol ratio																								
	0 – 4% mortality risk by CVD						5 – 9% mortality risk by CVD						≥10% mortality risk by CVD											

Source: NHG guideline M84, cardiovascular risk management version 1.3 (2007). pharmacotherapeutic compass 2007.

Module Diabetes Mellitus type 2

Notions

This module does not have the intention to cover the entire field of cardiovascular risk management. Please refer to the standard cardiovascular risk of the NHG for more information

Reference values for the diagnosis of diabetes mellitus and impaired fasting glucose (mmol/l).			
		capillary blood	venous plasma
normal	glucose fasting	<5,6	<6,1
	glucose random	<7,8	<7,8
impaired	glucose fasting	>5,6 en <6,0	>6,1 en <6,9
	glucose random	>6,0	>6,9
diabetes mellitus	glucose fasting	>6,0	>6,9
	glucose random	>11,0	>11,0

prevalence: 36/1000 men, 39/1000 women (type 1 en 2)
 2-4 times higher in: Moroccan, Turkish and Surinamese descent.
 Hindustani
 descent: prevalence even higher and more micro and macro vascular complications.

Incidence: 4,5/1000/year, increasing with age.

Part A, diagnostics

General

Well calibrated portable meters can have a deviation of 10-15%

Different methods are used to estimate kidney function. The Cockcroft-Gault and MDRD are used most frequently. Both are prone to confounders (eg. BMI, age, sex) and need cautious interpretation.

Note: all prices mentioned are indicative only for laboratory testing

Glucose, random € 1,41

To be determined with complaints. Furthermore determine every three years in each clinic visitor > 45 years:

- when DM2 is present with parents/brother/ sister
- with a history of hypertension
- with a history of CVD
- with a history of dyslipaemia
- from Turkish, Moroccan or Surinamese descent
- from Hindustani descent (>35 yr),
- when BMI >27,
- with a history of diabetes gravidarum.

Laboratory results for random glucose are lower with a higher hematocrit present.

May alternatively be assayed instead of fasting glucose when following-up a patient. These values are not interchangeable.

Fasting glucose € 1,41

Mandatory for diagnosing when random glucose <11,0. First step after diagnosing DM2 is lifestyle advice. Oral medication is only to be adjusted after testing fasting glucose levels. To be tested once every three months after diagnosing DM2 and initiating drug treatment. Laboratory results for fasting glucose are lower with a higher hematocrit present.

HbA1c € 6,65

To be tested after diagnosing DM2. Plays no role according to international agreements for diagnosing DM2. To be tested once a year to follow-up on the treatment. Target HbA1c levels: <7%.

Only valid parameter to monitor whether the initiated therapy is effective. False increased levels can be seen with erythrocyte abnormalities, elevated urea or overuse of aspirin.

Creatinine € 1,41

Determine after diagnosing DM 2. This is used to calculate the kidney function, see note above (under "general").

Albumin concentration, urine € 1,41

To be tested annually. Abnormal values should be confirmed by a second assay a few months later. The target value is <20mg / ml. There is a slight preference determining the albumin / creatinine ratio in the urine.

Total cholesterol € 1,41

Determine after diagnosing DM 2 and annually thereafter. Target level is <4.5 mmol / l in DM 2 if the LDL cannot be determined because of high triglyceride levels.

HDL € 2,77

Determine after diagnosing DM 2 and annually thereafter. Target level is > 1,0 mmol/l

LDL € 6,07

LDL is in many cases no laboratory determination in the strict sense; it is calculated on the basis of total cholesterol, triglycerides and HDL.

Determine after diagnosing DM 2 and annually thereafter. A HbA1c > 8.5 leads to underestimation of LDL levels. Policy should then be determined on the basis of total cholesterol. In the case of a high triglyceride level, it is impossible to determine LDL levels. Target level is <2.5 mmol / l in diabetics.

Triglycerides € 1,89

Determine after diagnosing DM 2 and annually thereafter. Variation on the same person can be up to 30%, both in the long term and the short term (day). Increases were observed up to 12 hours after ingestion of food.

Module Diabetes Mellitus type 2

Part B, pharmacotherapy

General

NHG advises to start treatment with three months lifestyle management. If this effect is insufficient treatment with medication should be initiated. The policy should be adjusted based upon the fasting blood glucose and HbA1c. Determination of HbA1c is most useful to determine whether the proposed targets are achieved or to check if a new step is indicated in the policy.

Metformin € 3,81-4,76/ month (2gr = 1DDD)

Drug of first choice. Inhibits glucose production in the liver and enhances peripheral insulin sensitivity. In high doses metformin has favorable effects on total cholesterol and LDL.

Sulfonylureas (excl. glibenclamide)

Promoting insulin release from β -cells. Reducing incidence of microvascular complications has not been proven, mortality is not reduced. Use of sulfonylureas can result in a weight gain of 2-5 kg. There is no significant difference demonstrated between the different sulfonylureas with the exception of glibenclamide.

Glicazide € 6,15-7,74/ month (0,16gr = 1DDD)

Glimepiride € 4,83-5,11/ month (2mg = 1DDD)

Tolbutamide € 2,44/ month (1,5gr = 1DDD)

Glibenclamide € 4,70-8,16/ month (10mg = 1DDD)

Relatively high risk of hypoglycemia, administration is not recommended. See further text under sulfonylureas.

Thiazolidinediones

Improving peripheral insulin sensitivity. In overweighted patients both drugs lower glucose levels and insulin, triglycerides and fatty acids. Liver function disorders are a relative contraindication as is heart failure. Based on experience cost and proven effectiveness of metformin this remains the drug of first choice.

Pioglitazon € 34,45/ month (30mg = 1DDD)

A few studies are known with positive results on clinical endpoints. Has a slightly favorable effect on the lipid spectrum compared to rosiglitazone. Classified in step 2 of the NHG guideline.

Rosiglitazon € 44,30/ month (6mg = 1DDD)

Little information available on results at hard endpoints. If a TZD is prescribed: preferably choose Pioglitazone. Recently the potentially increased risk of fracture and coronary heart disease when using this drug is discussed. It is unclear how to deal with this in practice.

Repaglinide € 14,47/ month (4mg = 1DDD)

Functioning similar to those of sulfonylureas, however bind to other binding site, and has shorter working time. The efficacy is comparable to that of sulfonylureas and metformin. However no results are known on clinical endpoints. May be prescribed by impaired renal function.

Acarbose € 17,63/ month (0,3gr = 1DDD)

Alpha glucosidase inhibitor. Because of its many side effects and low blood glucose lowering capacity this drug should be considered obsolete.

Combination preparations:

The NHG guideline does not mention combination preparations.

Metformin-glibenclamide € 6,50/ month

Metformin-rosiglitazone € 34,73-58,31/ month

Glimepiride-rosiglitazone € 45,63-75,79/ month

Insulin 1DDD = 40E

Insulins are distinguished by duration of action:

- Short acting
- Medium long acting
- Long acting
- Mix insuline

The choice for a certain type of insulin is dependent of the treatment regimen that is followed (see below).

Step III: medium long acting insulin once daily to be administered next to oral drugs (stop TZD use). Start with 10^E after dinner. Adjust dosage according to the diagram at the next page (figure 2) guided by the fasting glucose levels.

Step IVa: medium long acting or mix insulin twice daily. Start with 80% of the daily dose of step III. Divide according to: 66% prior to the breakfast and 33% prior to dinner. Adjust dosage guided by fasting glucose levels and glucose levels after the meals. Continue the use of Metformin

Step IVb: Start with 80% of the daily dose of step III. Start with 3x 20% short acting insulin prior to the meals and 1x 40% long acting insulin before sleep

Glucagon € 24,81 (per ampoule) (1mg = 1 ampoule = 1DDD)

Blood glucose enhancer. Promotes glycogenolysis in the liver causing the blood sugar level to rise rapidly. Only works if there is stocked glycogen in the liver. Because of the small number of prescriptions and because it is not clear which part of the prescriptions are actually administered, you will get no feedback on this product.

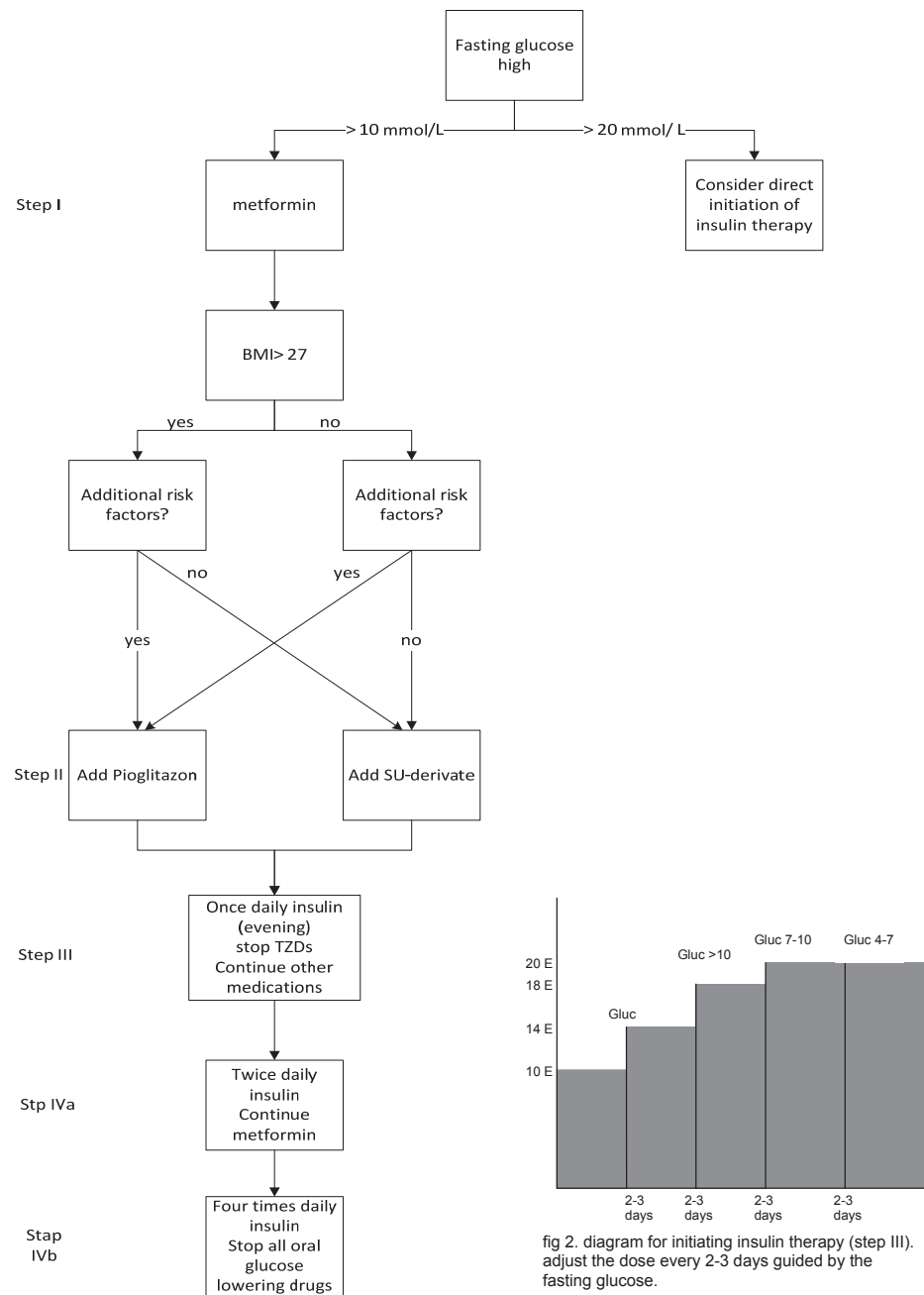


Figure 1 Flowchart diabetes mellitus type 2 treatment. When targets are not reached advance to the next step.

Source: NHG guideline M01, diabetes mellitus type 2. Pharmacotherapeutic compass 2007

Module Urinary tract infections

Notions

Incidence:	acute cystitis	♀ 70/ 1000 patients/ year
		♂ 10/ 1000 patients/ year
	pyelonephritis	♀ 2/ 1000 patients/ year
		♂ <1/ 1000 patients/ year
	prostatitis	2/ 1000 patients/ year

The incidence of acute cystitis in women has a peak between 15 and 24 years and after 60 years. In women 75 years of age, the incidence is 200/1000 patients / year.

In men, the peak incidence peaks after the age of 50. In men of 80 years the incidence is 80/1000 patients / year.

Cystitis is the most common complaint in women consulting a GP.

Part A, diagnostics

General

The diagnosis urinary tract infection can only be determined in a patient with symptoms. Laboratory diagnostics is only valid as a complement; the history is the most important. Physical examination is recommended in patients at risk, with signs of tissue invasion and in women with recurrent urinary tract infections (single study).

Collection of urine: It is no longer preferred to make use of special measures for the reception of urine. Only when a discrepancy between symptoms and findings on examination exists it may be considered to catch the urine after spreading the labia or retract the prepuce.

Use preferably first morning urine. Maximum storage room temp: 2 hours, refrigerator: 24 hours.

Control after an uncomplicated cystitis is **not** necessary, provided the patient is free of symptoms.

Nitrite test urine, POC (text but no feedback) € 0,27 - 1,10 /stick

Based on the presence of reductase containing bacteria, they convert nitrate to nitrite. A positive nitrite test in case of complaints justifies treatment as cystitis. A negative nitrite test with complaints cannot reject the diagnosis of cystitis.

The NHG guideline recommends this test to be carried out first.

Control, in an asymptomatic patient, after an uncomplicated cystitis is not necessary.

Reliability testing has not been studied in this situation.

Leukocyte test urine, POC (text but no feedback) € 0,27 - 1,10 /stick

Has as isolated test little impact on the confirmation or rejection of the diagnosis cystitis. As an addition to the nitrite test this test just seems important if both results are negative a cystitis is highly unlikely. Control in an asymptomatic patient after an uncomplicated cystitis is not necessary.

Reliability testing has not been studied in this situation.

Hemoglobin test urine POC (text but no feedback) € 0,27 - 1,10 /stick

This test provides no added value in addition to the tests above for the confirmation or rejection of the diagnosis cystitis. The NHG guideline does not recommends this test in the context of diagnosing cystitis. Control in an asymptomatic patient after an uncomplicated cystitis is not necessary.

Reliability testing has not been studied in this situation.

Dip slide, POC (text but no feedback) € 1,06 - 1,50 / piece

Can be used in the general practice in order to confirm or exclude cystitis. Cut-off point is 104 bacteria per ml. Older studies indicate that the slide is a reliable test. A 2003 study questions this.

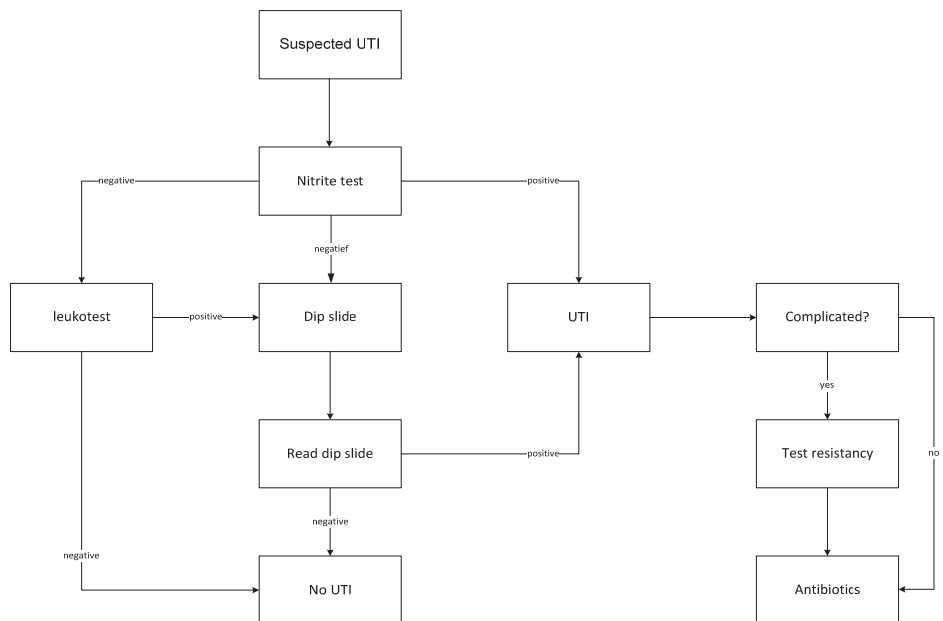
Sediment, POC (text but no feedback) € 1,41

Is to be used in a negative nitrite test to rule out the diagnosis cystitis. Only reliable if done by skilled investigator and the microscope is well maintained.

Urine culture with antibiotic resistance testing (price differs)

First choice after two blind started therapies have failed and in complicated urinary tract infections. The urine must be collected before starting new antibiotic therapy.

If a culture shows a group B streptococcus (GBS) any pregnant woman should be referred to secondary care, regardless of gestational age and susceptibility. Recolonization with GBS occurs very frequently. Neonatal GBS infection has a high morbidity and mortality.



Flowchart UTI

Module urinary tract infections

Part B, pharmacotherapy

General

An uncomplicated urinary tract infection is treated for up to 5 days.

A complicated urinary tract infection is treated for 7-10 days.

Treat as complicated UTI when:

- fever or tissue invasion is present
- in men
- in boys up to 12 years
- in girls up to 4 years
- in presence of anatomic anomalies to kidneys or urinary tract
- with an impaired defense (excl. diabetics)
- with a catheter in situ

Nitrofurantoin € 2,76 /week, **Furabid** € 3,73 /week (1 DDD = 0,2 g)

drug of first choice in an uncomplicated UTI. 2 times daily 100mg for 5 days (as long as Furabid® isn't available: 4dd 50mg). Antibiotic resistance of E. Coli is very limited. A 7 day treatment regimen has the same effects as a 5 day regimen.

Contraindications: renal impairment, G6PD deficiency, just before / during childbirth.

Men, girls between 5 and 12 years, pregnant women and diabetics without tissue invasion: treat for 7 days.

Trimethoprim € 2,35 /week, € 1,64 /week (child, suspension) (1 DDD = 0,4 g)

Drug of second choice according to the NHG guideline. First choice with intolerance for nitrofurantoin. Once daily 300mg for 3 days.

note: local antibiotic resistance patterns can influence the choice of the preferred drug.

Fosfomycin € 5,98 /week (1 sachet) (1 DDD = 3 g)

Drug of third choice according to the NHG guideline. Prescribe 3 grams once. Start when nitrofurantoin and trimethoprim cannot be administered.

Amoxicilline-clavulaanzuur € 11,73 - 27,63 /week, € 2,34 - 8,28 /week (child) (1 DDD = 1 g)

Drug of first choice in complicated UTIs.

Co-trimoxazol € 2,55/week, € 1,17 - 1,30/week (child)

Drug of second choice in complicated UTIs. Prescribe 2 times daily 960mg for 10 days. Not to be prescribed to pregnant women.

Note: local antibiotic resistance patterns can influence the choice of the preferred drug.

Norfloxacin € 5,64 - 5,96/week (1 DDD = 0,8 g)

Reserve antibiotic for urinary tract infections, only prescribe guided by antibiogram. Prescribe twice daily 400mg for 10 days.

Ciprofloxacin € 17,50 - 39,41/week (1 DDD = 1 g)

Reserve antibiotic for urinary tract infections, only prescribe guided by antibiogram. Prescribe twice daily 500mg for 10 days. Do not prescribe to pregnant women or youngsters <16 years.

Module anaemia

Notes

Anaemia:	Low hemoglobin levels considering age and sex (according to normal reference range from regional laboratory)
Mild:	Hb >6,0 mmol/l (women and children) of > 6,5 mmol/l (men)
Moderate:	Hb >5,0 mmol/l but < 6,0 mmol/l (women and children) or < 6,5 mmol/l (men)
Severe:	Hb <5,0 mmol/l
Microcytic:	mean corpuscular volume (MCV) <80 fl
Normocytic:	80 fl < MCV <100 fl
Macrocytic:	MCV >100fl

Part A, diagnostics

Background

Reference values are usually calculated based on a group healthy individuals, the 97,5 and 2.5 percentile are boundaries. Random blood tests performed in healthy thereby results in 5% abnormal results which do not indicate disease. The likelihood of an abnormal result increases as more tests are pending to a person. At 5 tests this chance is increased to 23%. Testing Hb levels because of fatigue does not seem rational. The patient however can consider the "magic" of blood tests as satisfactory.

Risk factors for anaemia:

- bloodloss
- abnormal diet (including alcohol abuse)
- infectious disease within last month (children)
- at risk for thalassemia (family history, immigrant descent)
- gastrointestinal malignancy
- old age
- chronic disease

Incidence: 6 per 1000 patients each year iron deficiency anaemia
0.5 per 1000 patients each year other deficiency anaemias (esp. vit. B₁₂)
0.15 per 1000 patients each year other anaemia

General advice

Since iron deficiency anemia is the most common **simultaneous determination of Hb, MCV and ferritin is the most appropriate.**

Hb/ Ht € 11,08/ 1,41

The common Hb meters have a measurement error of 0.2 mmol / l. The Hb in capillary blood is on average 0.14 mmol / l lower than the value determined from venous blood. The milking of a finger can easily lead to a low value by mixing with tissue fluid. There is no difference in Hb of patients who complain of fatigue and patients who do not. Treatment does not lead to improvement of symptoms. Testing Ht is meaningful in cases of a slightly increased or slightly decreased Hb. This could be due to a dilution or thickening effect.

Indices € 1,41 a piece

Of the indices is just the MCV useful for the diagnosis of anemia in primary care. MCV helps classifying microcytic, normocytic or macrocytic anaemia. An increase in the number of reticulocytes (see there) can lead to an increase in the MCV because reticulocytes are larger than mature erythrocytes.

Ferritin € 8,31

Ferritin is a good indicator for the reserve-iron in the body, 1 µg / l of ferritin = 8µg reserve-iron. Ferritin is an acute phase protein and thus test results can be disrupted by infections. Ferritin <15 µg / l indicates an iron deficiency anemia, further diagnosis of the iron status is not necessary. Values above 100 µg / L exclude an iron deficiency anemia.

Bilirubin € 1,41

Has no place in the diagnosis of anemia in the first line. Not useful when diagnosing hemolysis.

Folic acid € 6,65

Folic acid is responsible for the synthesis of nucleic acid (DNA). Absorption takes place in the small intestine. Folate deficiency anaemia are rare in primary care.

Vitamin B₁₂ € 8,31

A macrocytic anaemia is usually caused by a vitamin B12 deficiency. The most common cause of vitamin B12 deficiency is the presence antibodies against intrinsic factor in the stomach (pernicious anemia).

Reticulocytes € 2,77

Determination only makes sense when further investigation is needed in the event of a macrocytic or normocytic anaemia.

Increases in the value indicative of hemolysis or blood loss and the exertion of the body to supplement the shortage of erythrocytes. A decrease indicates a shortage of nutrients (vitamin B12, folic acid or iron) or a production disorder. 4-5 days after onset of replacement therapy increases the value with a peak at day 7 (reticulocyte crisis).

Serum iron € 5,55

Useless in following up on iron deficiency.

Needs rarely be tested in primary care.

50% of the patients exhibit a circadian difference in serum iron content of 30%. Serum iron therefore should be tested preferably in the morning (fasting). In over or improper supplementation the serum iron will be elevated with normal or reduced transferrin levels.

LDH € 1,41

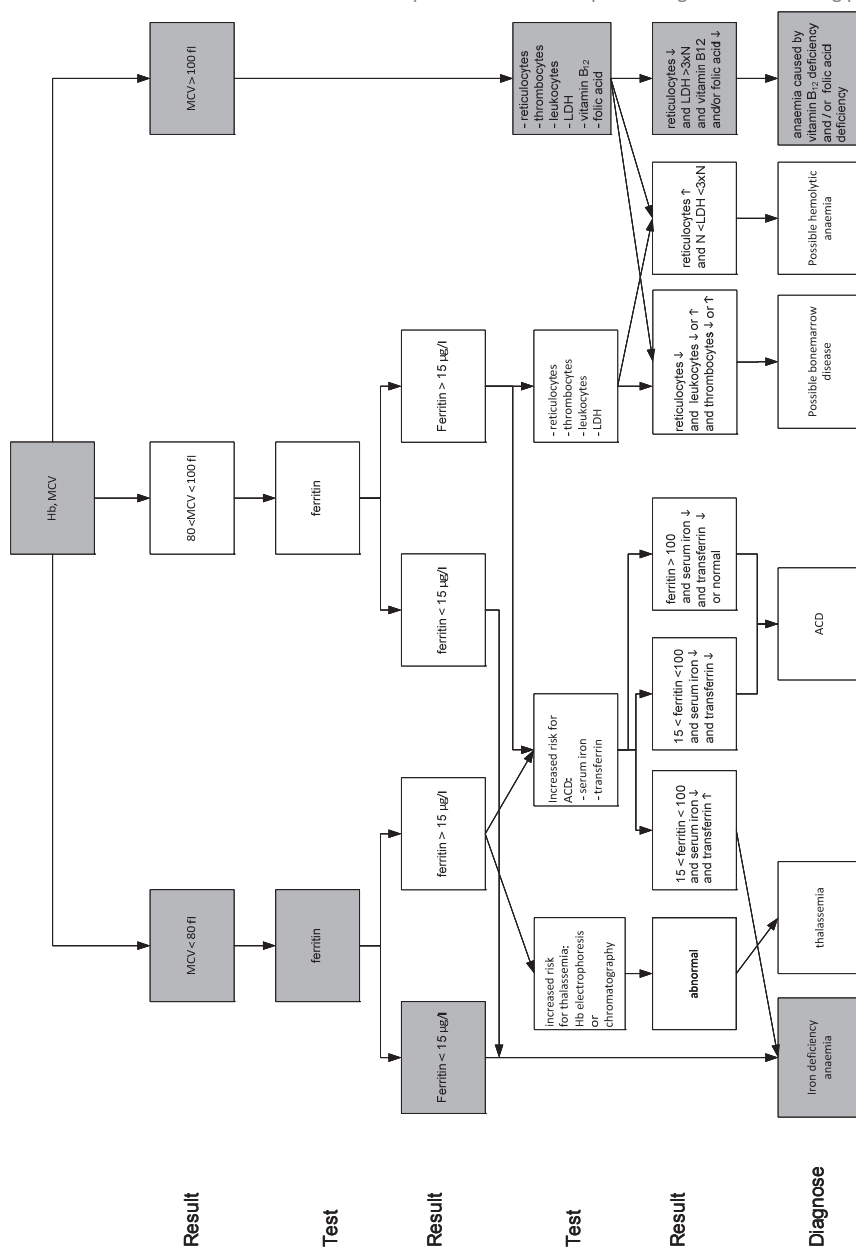
Is rarely needed, vitamin B12 and folic acid levels provide sufficient information. LDH is present in every cell of the body. Damage to the cell leads to release of LDH. In the case of a haemolytic anaemia the LDH is increased up to 3 times the normal value. In the case of an untreated megaloblastic anaemia due to deficiency of folic acid LDH levels will have increased more than 3-fold.

Transferrin/ Total iron binding capacity (TIBC)/ transferrin saturation € 5,55

Transferrin: hardly ever indicated in primary care considering anaemia diagnostics.

TIBC: complete iron binding capacity of transferrin together with a small contribution of albumin. Is considered to be obsolete.

Transferrin saturation: hardly ever indicated in primary. In repeated tests >60%: higher probability of hemochromatosis.



This algorithm provides a pathway to interpret the most common combinations of test results: Less common combinations of results are more difficult to interpret.

Hb: Hemoglobin; N: normal value; ACD: anaemia caused by chronic disease; LDH: lactate dehydrogenase; MCV: mean corpuscular volume; ☐: most frequent diagnosed

Module anaemia

Part B, Pharmacotherapy

General

The following drugs serve as supplementation for an existing shortage of the substance. The cause of a deficiency should be treated when possible. This is beyond the scope of this summary. Doses higher than 100mg ferrous iron per day result in more side effects. Administration with calcium can inhibit the absorption of ferrous iron.

Ferrous sulfate (= slow release) € 4,40/ month (1 DDD = 105mg Fe⁺⁺)

Not recommended.

Because iron is absorbed almost exclusively in the first part of the intestine, slow release preparations lead to a lower absorption of iron and therefore have a lower impact on the hemoglobin content of the blood. The biological availability of divalent iron compounds amounts to 46 to 100%; for slow-release formulations it is at 31-47%.

Ferrous fumarate € 4,38/ month (tabl), € 13,17/ month (susp), € 5,34/ month (child) (1 DDD = 195mg Fe⁺⁺)

Drug of first choice according to the NHG guideline. Divalent iron has a good absorption and bioavailability. 3 times daily 200mg or 2-4mg / kg ferrous as suspension (child) distributed over three doses. The suspension can give tooth discoloration; intake with a straw prevents this.

Ferrous gluconate € 15,28/ month (1 DDD = 154mg Fe⁺⁺), (tablet: 1 DDD = 80mg Fe⁺⁺)

The suspension is drug of second choice according to the NHG guideline. Adults: 3 times daily 6ml; children 2-4mg ferro/kg distributed over 3 doses. The suspension can give tooth discoloration; intake with a straw prevents this.

The effervescent tablet is not included in the standard. It can give tooth discoloration. The claim that there are fewer gastrointestinal side effects in effervescent tablets is not or not sufficiently supported by evidence.

Ferrous chloride € 4,11/ month, € 1,23/ month (child, susp) (1 DDD = 150mg Fe⁺⁺)

Not mentioned in the NHG guideline. The suspension can give tooth discoloration; intake with a straw prevents this.

Folic acid (vit B₁₁) € 0,76 /month (based on 5 mg tablets!) (1 DDD = 0,4 mg)

In proven folic acid deficiency folic acid can be prescribed with 0.5mg once daily. After 6-12 weeks the administration can be discontinued. The cause of the deficiency needs to be treated if possible. Green leafy vegetables are rich in folic acid.

Supplementation without pernicious anaemia ruled out seems unwise because of increased risk of neurological abnormalities. Most times folic acid and vitamin B12 are tested together.

Note: Most folic acid tablets are freely available without prescription; they are not included in the figures as presented.

Vitamine B₁₂ € 0,61/ ampoule 1 ampoule = 50DDD's

In a proven vitamin B12 deficiency the GP starts with 10 injections of 1mg hydroxocobalamin with an interval of 3 days. After this, one injection every two months. In case of a pernicious anaemia: treat lifelong.

Oral therapy does not appear necessary, by the lack of intrinsic factor is the added oral vitamin is not absorbed.

Hydroxocobalamin is the most natural and thus has the greatest bioavailability; hydrocobamine should not be prescribed for this reason.

Module stomach complaints

Backgrounds

Prevalence of *Helicobacter Pylori* infections is much higher in immigrants due to endemic prevalence in the country of origin.

Incidence: 25 per 1000 stomach ache
 7 per 1000 heartburn
 14 per 1000 nausea

Diagnosis: 60-70% functional complaints or cause unknown, prognosis is poor,
 50 to 90% has long term complaints.
 20-25% gastroesophageal reflux
 5% ulcer ventriculi, duodenal ulcer
 <1% malignancies

Part A, diagnostics

General

Alarm Symptoms: haematemesis, melaena, dyspepsia with persistent vomiting, passage disorders, unintentional weight loss and anaemia.

Risk factors for ulcer disease: previous ulcer, male sex, older age, smoking and hunger pains.

In the elderly one should be aware for in serious pathology in an earlier stage.

There are no reliable data on testing for *H. pylori* routinely or in persistent symptoms after eradication therapy. CBO guideline: only if the doctor can order a breath test after eradication, control is recommended at least 30 days after eradication.

¹³C-urea breathtest/ ¹⁴C-urea breathtest € 53,00/ € 121,00

Test of first choice according to the NHG guideline (¹³C- urea breathtest). Sensitivity of the ¹³C method is >98%, specificity is >97%. The ¹⁴C method needs the presence of a nuclear medicine department in the referral hospital. The ¹³C method is used by the clinical chemistry department.

Treatment with PPIs or H₂ receptor antagonists should be stopped 14 days prior to testing.

HP- antibodies/ fecetest € 12,28-18,00/ € 21,85

Some laboratories won't perform *H. Pylori* antibody in serum testing anymore. Antibodies in serum are detectable up to 12 months after eradication.

De fecetest is only validated for use in secondary care. Both are unsuitable for testing after eradication therapy. Sensitivity and specificity of the antibody test are both 80 to >95%. Sensitivity of the fecetest is 80 to 90%, specificity >90%. The patient should not use PPIs 2 weeks prior to the test and no antibiotics 4 weeks prior to the test.

Esophagoscopy/ gastroscopy € 151,20 – 198,45/ € 352,67

Indications for endoscopy:

- If alarm symptoms, the patient is referred for endoscopy or a specialist.
- In elderly people with no history of stomach problems an endoscopy may be considered earlier.
- H. pylori negative patients who do not respond to treatment with acid inhibitors.
- No reduction of symptoms after eradication therapy (if no urea breath test is available).
- Esophagitis grade C and D after 12 weeks of treatment.

Sensitivity and specificity: depending on specialist performing endoscopy and medical history.

In GERD with good response to initial therapy, there is no indication for routine endoscopy.

X-contrast stomach/ esophagus € 114,24

There is hardly any indication for this type of X-ray. Endoscopy has greater diagnostic value and abnormalities seen on a photograph leads to an indication for endoscopy. Small differences are not seen on an X-stomach or esophagus.

In great fear of endoscopy a contrast image can be considered to detect gross abnormalities.

Module stomach complaints

Part B, pharmacotherapy

General

Start with antacids or H₂ receptor antagonists. In many cases this approach is sufficient. After a year 75% of the patients had no symptoms anymore. The placebo effect is substantial.

The discussion of antiemetic and prokinetic drugs is beyond the scope of this module. A distinction is made between a first episode and the last episode of > 1 year ago (without clear cause diagnosed) on the one hand and persistent or recurrent symptoms on the other.

Antacids

First choice at first episode of mild stomach complaints. Alghedrate / magnesiumoxide susp 4dd15ml one hour postprandial and before bedtime. Always subscribe for 2-4 weeks and then evaluate. Try to stop after eight weeks even in mild residual symptoms.

H₂- receptor antagonists € 11,62 - 36,86 / month

First choice at first episode more severe stomach upset and typical reflux symptoms. Always subscribe for 2-4 weeks and then evaluate. Try to stop after eight weeks even in mild residual symptoms. After discontinuation rebound symptoms may occur after 3 weeks. Use in some cases a step-down scheme and sustain with antacids. The degree of rebound is positively correlated with the degree of acid inhibition.

Chronic use: after gastric bleeding or perforation after consultation with the treating specialist. In grade C or D chronic esophagitis continuous use is indicated. In all other cases a complaint-driven treatment is indicated.

Well useable in double dosage for the prevention of NSAID induced ulcers. More expensive than misoprostol.

Famotidine (1DDD=40mg), ranitidine (1DDD=0,3g), Cimetidine (1DDD=0,8g), Nizatidine (1DDD=0,3g)

Protonpump inhibitors € 8,83 – 39,37 / month

Drug of first choice for H. Pylori eradication, highly dosed.

Relative second choice (NHG: to begin with H₂-receptor antagonists) in persistent or recurrent symptoms. Try-out treatment for 2-4 weeks. When results are satisfying: continue up to 8 weeks. Hereinafter administer intermittent (2-4 weeks) or on demand. Evaluate biannually for instance its use and the complaints. 6-14 days after discontinuation might a rebound effect occur. The degree of rebound is positively correlated with the degree of acid inhibition. If a patient H. Pylori is positive then this should be treated as well.

Chronic use: after gastric bleeding or perforation after consultation with the treating specialist. In grade C or D (≈ III-IV) chronic esophagitis continuous use is indicated. In all other cases, a complaint-driven treatment is indicated.

Useful for the prevention of NSAID-induced ulcers. More expensive than misoprostol.

Omeprazole (1DDD=20mg), Pantoprazole (1DDD=40mg), Lansoprazole (1DDD=30mg), Rabeprazole (1DDD=20mg), Esomeprazole (1DDD=30mg)

Pantopac € 275,71 / month

No judgment rendered by NHG. Contains triple therapy in dose as recommended by NHG. Note the NHG guideline mentions no proton pump inhibitor of choice.

Misoprostol € 31,25 / month (1 DDD = 0,8 mg)

prostaglandin E1 analogue with both mucosa protective and gastric acid inhibiting effect. Each dosage can cause diarrhea. The degree of side effects depends on the dose. If tolerated well by the patient it is useful for the prevention of NSAID-induced ulcer disease and in addition it is cheap.

Amoxicilline € 4,75 - 4,77 / week, € 1,23 - 2,28 / week (child) (1 DDD = 1 g)

2 times daily amoxicillin 1000 mg in combination with clarithromycin and a proton pump inhibitor (triple therapy) for 7 days is an effective eradication therapy. Locally based on abnormal resistance patterns different combination of antibiotics or other dosages can be preferred. In penicillin allergy to be replaced by metronidazole.

Clarithromycine € 8,71 - 10,30 / week, € 3,01 - 6,03 / week (kind) (1 DDD = 0,5 g)

Clarithromycin 2 times daily 500 mg in combination with amoxicillin and a proton pump inhibitor (triple therapy) for 7 days is an effective eradication therapy. Locally based on abnormal resistance patterns different combination of antibiotics or other dosages can be preferred.

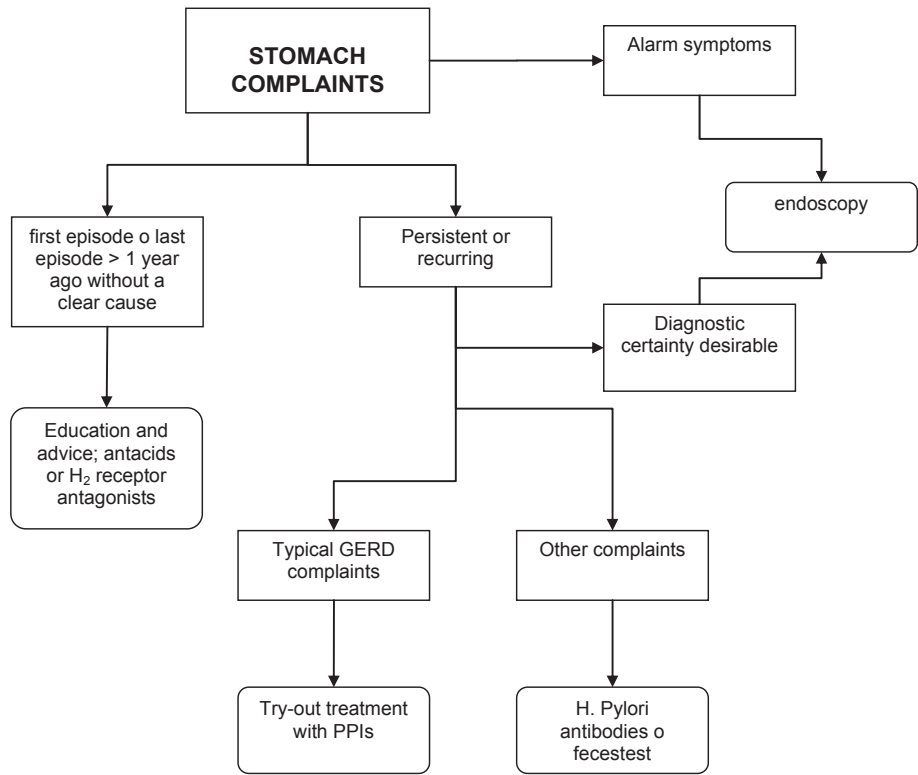
Metronidazol € 3,80/week (1 DDD = 2 g)

Part of quadruple therapy, apply if triple therapy proven to have failed. Dosage 3 times daily 500 mg, 7 days. Combining with proton pump inhibitor, dosed high. The quadruple therapy is the second choice, despite the limited influence of resistance patterns because of the complexity of the treatment and the related lower compliance.

Tetracycline € 1,05 / week (1 DDD = 1 g)

Part of quadruple therapy, apply if triple therapy proven to have failed. Dosage 4 times daily 500 mg, 7 days. Combining with proton pump inhibitor, dosed high. The quadruple therapy is the second choice, despite the limited influence of resistance patterns because of the complexity of the treatment and the related lower compliance.

Flowchart stomach complaints



Source: NHG guideline M36 stomach complaints

Module thyroid dysfunction

Notes

Hypo and hyperthyroidism:	clinical condition characterized by typical complaints proven by testing TSH and FT4 levels.
Hypothyroidism:	TSH ↑ and FT4 ↓
Hyperthyroidism:	TSH ↓ and FT4 ↑
subclinical hypothyroidism:	state in which the TSH deviates and FT4 levels are normal
Myxedema:	pasty swelling of the skin, non-pitting. Occurs in both hypo- and hyperthyroidism. Hyper: especially pretibial, hypo: especially around the eyes and rest of face.
Thyroid nodules:	localized enlargement of the thyroid gland
Struma:	enlargement of the thyroid gland diffuse of several nodules (Multi-nodular).
M. Hashimoto:	chronic autoimmune thyroiditis. Mainly in women. Non-iatrogenic hypothyroidism: 90-95% Hashimoto.
M. Graves	autoimmune thyroiditis TSI (see below) attaches itself to receptors like TSH does.
silent thyroiditis:	autoimmune thyroiditis with destruction and regeneration thyroid tissue recovery in 1-4 months.
Postpartum thyroiditis:	see under silent thyroiditis occurs in 7% of women in the first year postpartum. 25-50% chance of the occurrence of a long-term hypothyroidism.
Incidence primary care:	hypothyroidism: 1.2/ 1000 a year (increased incidence with Down syndrome) Hyperthyroidism: 0.5/ 1000 a year Malignancies: 2.8/ 100.000 a year(♀) en 1,3/ 100.000 a year (♂)
Prevalence primary care:	thyroid nodules: 0.4/ 1000 Struma: 0.7/ 1000

Part A, diagnostics

General advice

Always determine upon suspicion of thyroid dysfunction the presence TSH and FT4 together (if not included in the cascade on the problem-based test ordering form).

Always determine when M. Graves is suspected the TSI (besides TSH and FT4).

Pursue an euthyroid situation in women who are pregnant but have had a (subclinical) hypothyroidism or hyperthyroidism in the past. Thyroid abnormalities impair cerebral development of the child in utero. Consult a qualified specialist internal medicine or gynecologist about referral.

General

The thyroid gland produces 4x more T4 than T3. T4 is metabolized into the biological active T3 in peripheral tissue.

95% of all thyroid dysfunction is primary (reason within the thyroid gland itself). 5% of the thyroid dysfunction is secondary (caused by pituitary gland dysfunctioning or CNS disease).

TSH € 5,55

Thyroid stimulating hormone

Determination of TSH is the best screening test for thyroid problems. A normal TSH virtually excludes thyroid dysfunction.

TSH can be determined if a patient has symptoms consistent with a hypothyroidism or hyperthyroidism. There is no pathognomonic symptom pattern (excl. Ophthalmopathy).

Increased risk of thyroid dysfunction in:

- Treated hyperthyroidism
- history of radiotherapy of head and neck
- Recent parturition (<1 year ago) or history of hypothyroidism postpartum
- Presence of other autoimmune disease (particularly DM type I)
- Down's syndrome
- Positive family history
- Use of iodine-containing drugs (lithium, amiodarone)
- Status after hypovolemic shock; due to necrosis or atrophy of the pituitary gland (trauma, post-partum haemorrhage).

Follow-up: after hyperthyroidism 1 Year after discontinuation of medication. A renewed low TSH means there is a relapsing disease

FT4 € 6,65

Free T4, that part of the total thyroxine in serum that is bound to protein.

Is lower in the third trimester of pregnancy.

Is typically automatically determined by the laboratory at an abnormal TSH level. Is also used to monitor the treatment of hyperthyroidism and adjust the therapy.

T3 € 8,31

Tri-iodo thyronine, biologically active thyroid hormone, in particular found in tissues outside the thyroid gland. Provides a better reflection of the biological effect than FT4 or T4. Should not be requested as a separate test. Indication: Excluding T3 toxicosis.

TSI € 8,31

TSH-receptor stimulating antibodies, stimulating the thyroid gland. Presence indicates Graves' disease.

Determine when Graves' disease is suspected and in pregnant women who have had hypothyroidism in the past.

TPO-Ab € 8,31

Antibodies mainly found in Hashimoto's thyroiditis, the antibodies destroy the thyroid. In other thyroid disorders slightly increased values can be detected.

Ultrasound thyroid gland € 56,52

Does not contribute to the diagnosis of thyroid dysfunction. Meaningful in the context of diagnosis of palpable abnormalities of the thyroid gland which are not accompanied by thyroid dysfunction.

Determination of ESR, leukocytes and differentiation for diagnosing thyroid dysfunction are so unspecific that no feedback can be provided.

Requesting a nuclear scan of the thyroid gland is for most GPs not possible. Therefore you do not receive feedback on this item

Module thyroid dysfunction

Part B, Pharmacotherapy

General

In the case of treatment of hypothyroidism following applies: "start low, go slow". Ensure that fertile women in your population with a known hypothyroidism contact you if they want to be pregnant or may be pregnant. The need for thyroxine rises in pregnancy. The need is lower in the elderly.

Check treatment of hypothyroidism every 6 weeks (TSH) until levels are appropriate, earlier does not make sense.

Pregnancy: Check every 4 weeks. Optional: start "blind" with twice every week an additional daily dose from the start of the pregnancy.

Levothyroxine sodium € 1,25 - 1,79 / month (1 DDD = 0,15 mg)

Substitution therapy for hypothyroidism, drug of first choice. Average dose in total failure of the thyroid gland is 1.6 micrograms / kg / day. Adjust the dose guided by the TSH and complaints.

Liothyronine sodium € 40,36 / month (1 DDD = 0,60 mg)

Combined with levothyroxine as combination therapy offers no advantage over monotherapy with levothyroxine. Same mechanism of action as levothyroxine, it takes effect a few days earlier and holds a few days less. Is not included in the NHG guideline as monotherapy.

Thiamazole € 1,36 / month (1 DDD = 10 mg)

Titration method: increase the dose guided by the FT4 (eg. in pregnant women.)

Combination method: once daily 30mg, it stops the thyroid gland. Add-back therapy with levothyroxine guided on the FT4. Determination of TSH is not necessary. When M Graves: stop all medication after 1 year and follow thyroid function. When (multi) struma: lifelong treatment.

Note: therapy requires special interest and knowledge.

Propylthiouracil (no feedback due to low volumes)

Antithyroid drug not included in NHG guideline.

Carbimazole (no feedback due to low volumes)

Antithyroid drug not included in NHG guideline.

Source: NHG guideline M31, Thyroid diseases. Pharmacotherapeutic compass 2007.

Appendix 4.2

The complete set of drugs and tests included in the databases for this intervention

Module	ATC	Drug	label on feedback form
Diabetes mellitus 2	A10BA02	metformin	metformin
	A10BB09	glicazide	glicazide
	A10BB12	glimepiride	glimepiride
	A10BB03	tolbutamide	tolbutamide
	A10BB01	glibenclamide	glibenclamide
	A10BG02	rosiglitazone	rosiglitazone
	A10BG03	pioglitazone	pioglitazone
	A10BX02	repaglinide	other oral antidiabetics
	A10BF01	acarbose	other oral antidiabetics
	A10BH01	sitagliptin	DPP4 inhibitors
	A10BH02	vildagliptin	DPP4 inhibitors
	A10BH03	saxagliptin	DPP4 inhibitors
	A10BX04	exenatide	incretines
	A10BX07	liraglutide	incretines
	A10BD04	glimepiride/ rosiglitazone	combination preparations
	A10BD02	metformin/ glibenclamide	combination preparations
	A10BD05	pioglitazone/ metformin	combination preparations
	A10BD03	metformin/ rosiglitazone	combination preparations
	A10BD07	sitagliptin/ metformin	combination preparations
	A10BD08	vildagliptin/ metformin	combination preparations
	A10AB	insuline short acting	insuline short
	A10AC	insuline medium long acting	insuline medium long
	A10AD	insuline mix	insuline mix
	A10AE	insuline long acting	insuline long
stomach complaints	A02AA	Magnesium compounds	antacids
	A02AB	Aluminium compounds	antacids
	A02AD	combinations and complexes of al- ca- en mg-compounds	antacids
	A02AH	Antacids with sodium bicarbonate	antacids
	A02BA03	famotidine	other H2-antagonists
	A02BA02	ranitidine	ranitidine
	A02BA01	Cimetidine	other H2-antagonists
	A02BA04	Nizatidine	other H2-antagonists
	A02BC01	Omeprazole	omeprazole
	A02BC02	Pantoprazole	pantoprazole
	A02BC03	Lansoprazole	lansoprazole
	A02BC04	Rabeprazole	rabeprazole
	A02BC05	Esomeprazole	esomeprazole
	A02BB01	Misoprostol	misoprostol
	A02BD04	pantoprazole, amoxicillin and clarithromycin, fixed combination	pantopac

(Continued)

Module	ATC	Drug	label on feedback form
Anaemia	B03AA07	ferrous sulfate	ferrous sulfate
	B03AA02	ferrous fumarate	ferrous fumarate
	B03AA03	ferrous gluconate	ferrous gluconate
	B03AA05	ferrous chloride	ferrous chloride
	B03BA01	vitamin B12	vitamin B12
	B03BA03	vitamin B12	vitamin B12
	B03BB01	folic acid	folic acid
Chlamydia infections	J01AA02	Doxycycline	doxycycline
	J01FA10	Azithromycin	azithromycin
	J01CA04	Amoxicillin	amoxicillin
	J01FA01	Erythromycin	erythromycin
Prostate complaints	G04CA01	Alfuzosin	alfuzosin
	G04CA02	Tamsulosin	tamsulosin
	C02CA04	Doxazosin	other $\alpha 1$ blockers
	G04CA03	Terazosin	other $\alpha 1$ blockers
	C02CA01	Prazosin	other $\alpha 1$ blockers
	G04CB01	Finasteride	finasteride
	G04CB02	Dutasteride	dutasteride
Rheumatic complaints	M01AA01	phenylbutazone	other NSAIDs
	M01AB01	Indometacin	indometacin
	M01AB05	Diclofenac	diclofenac
	M01AB02	Sulindac	other NSAIDs
	M01AB16	Aceclofenac	other NSAIDs
	M01AC06	meloxicam	preferential cox-2 inhibitors
	M01AC01	piroxicam	preferential cox-2 inhibitors
	M01AC02	tenoxicam	other NSAIDs
	M01AE01	ibuprofen	ibuprofen
	M01AE02	naproxen	naproxen
	M01AE03	ketoprofen	other NSAIDs
	M01AE09	flurbiprofen	other NSAIDs
	M01AE11	tiaprofenic acid	other NSAIDs
	M01AE14	dexibuprofen	other NSAIDs
	M01AE17	dexketoprofen	other NSAIDs
	M01AG02	tolfenamic acid	other NSAIDs
	M01AX04	azapropazone	other NSAIDs
	N02BA11	diflusal	other NSAIDs
	N02BB02	metamizole sodium	other NSAIDs
	M01AH01	celecoxib	coxibs
	M01AH05	etoricoxib	coxibs
	M01AX01	nabumetone	preferential cox-2 inhibitors
	M01AB55	diclofenac+ misoprostol	diclofenac+ misoprostol
	A07EC01	Sulfasalazine (tekst but no feedback)	DMARD's
	L01BA01	Methotrexate (tekst but no feedback)	DMARD's
	L04AA13	Leflunomide (tekst but no feedback)	DMARD's
	L04AA11	Etanercept (tekst but no feedback)	DMARD's
	L04AA12	Infliximab (tekst but no feedback)	DMARD's
	M01CB	gold preparations (tekst but no feedback)	DMARD's

(Continued)

Module	ATC	Drug	label on feedback form
	M01CC01	d-penicillamine (tekst but no feedback)	DMARD's
	P01BA01	chloroquine (tekst but no feedback)	DMARD's
	L04AX01	azathioprine (tekst but no feedback)	DMARD's
	L04AD01	ciclosporin (tekst but no feedback)	DMARD's
Thyroid dysfunction	H03AA01	Levothyroxine sodium	Levothyroxine
	H03AA02	Liothyronine sodium	Liothyronine
	H03BB02	Thiamazole	Thiamazole
Urinary tract infections	J01XE01	Nitrofurantoin	Nitrofurantoin
	J01EA01	Trimethoprim	Trimethoprim
	J01XX01	Fosfomycin	Fosfomycin
	J01CR02	Amoxicillin with enzyme inhibitor	Amoxicilline-clavulaanzuur
	J01EE01	sulfamethoxazole with trimethoprim	Co-trimoxazol
	J01MA06	Norfloxacin	Norfloxacin
	J01MA02	ciprofloxacin	ciprofloxacin
Dyslipaemia	C10AA01	Simvastatin	simvastatin
	C10AA03	Pravastatin	pravastatin
	C10AA05	Atorvastatin	atorvastatin
	C10AA04	Fluvastatin	fluvastatin
	C10AA07	Rosuvastatin	rosuvastatin
	C10AB02	Bezafibrate	fibrates
	C10AB04	Gemfibrozil	fibrates
	C10AB08	Ciprofibrate	fibrates
	C10AD06	Acipimox	nicotic acid (analogous)
	C10AX09	ezetimibe	ezetimibe
	C04AD02	xantinol nicotinate	nicotic acid (analogous)
	C10AD02	nicotic acid	nicotic acid (analogous)
	C10AC01	Colestyramine	colestyramine
	C10AX	ezetimibe/ simvastatin	combination preparete
perimenopausal complaints	G03CA03	estradiol	estrogen monotherapy
	G03CA04	estriol	estrogen monotherapy
	G03CA01	ethinylestradiol	estrogen monotherapy
	G03CA57	geconjugated estrogens	estrogen monotherapy
	G03AA	progestogens and estrogens, fixed combinations	contraceptives, >50 jaar
	G03AB	progestogens and estrogens, fixed combinations	contraceptives, >50 jaar
	G03AC	progestogens and estrogens, fixed combinations	contraceptives, >50 jaar
	G03BB01	progestogens and estrogens, fixed combinations	contraceptives, >50 jaar
	G03HB01	estradiol/cyproterone (Climene '28')	combination preparates
	G03FA17	estradiol/drospirenone (Angeliq)	combination preparates
	G03FB08	estradiol/dydrogesterone (Femoston)	combination preparates
	G03FA14	estradiol/dydrogesterone continuous (Femoston continuu)	combination preparates
	G03FA01	estradiol/norethisterone continuous (Activelle, Kliogest)	combination preparates
	G03FB05	estradiol/norethisterone (Estracomb TTS, Trisequens)	combination preparates
	C02AC01	clonidine	clonidine
	G03DC05	tibolone	tibolone

Module	Item	codeNr	description labcode table	Mnemonic	Mat	Group	label on feedback form
Diabetes mellitus 2	Glucose random	370	Glucose time	GLUC	B	KC	glucose
	Glucose not sober	371	Glucose not sober	GLUC	B	KC	glucose
	Glucose sober	372	Glucose sober	GLUC	B	KC	glucose
	HbA1c	368	glycohemoglobin (HbA1c)	GLHB	B	KC	HbA1c
	Total cholesterol	192	cholesterol total	CHOL	B	KC	T-cholesterol
	HDL	446	HDL-cholesterol	HDL	B	KC	HDL
	LDL	542	LDL-cholesterol	LDL	B	KC	LDL
	triglycerides	1377	triglycerides	TRIG	B	KC	triglyc
	creatinine	523	creatinine	KREA	B	KC	creat
	Albumin concentration, urine	38	albumin (micro-) urine portion	ALB	U	KC	albumin urine
stomach complaints	¹³ C-urea breathtest	2130	13C urea breathtest	UREU	RU	FO	breathtest
	¹⁴ C-urea breathtest		14C urea breathtest				breathtest
	Fecestest HP	2184	Helicobacter pylori ag. feces	HPYA	DF	IM	fecetest HP
	Fecestest HP	2107	Helicobacter pylori ag. feces (PCR)	HPYP	DF	IM	fecetest HP
	antibodies helicobacter pylori, IgG	478	Helicobacter pylori ab. IgG	HPYG	B	IM	antibodies HP
	antibodies helicobacter pylori, general	479	Helicobacter pylori ab.	HPYI	B	IM	antibodies HP
	antibodies helicobacter pylori, IgM	480	Helicobacter pylori ab. IgM	HPYM	B	IM	antibodies HP
	gastroscopy	2359	gastroscopy	SCOP	DU	BO	gastroscopy
	X-contrast stomach	2333	X-stomach	X	DU	BO	X-stomach

(continued)

Module	Item	codeNr	description labcode table	Mnemonic	Mat	Group	label on feedback form
Anaemia	Hb	412	hemoglobin (Hb)	HB	B	HE	Hb/ Ht
	Ht	484	hematocrit (Ht)	HT	B	HE	Hb/ Ht
	anaemia diagnostics (cascade)	2190	anaemia diagnostics (Hb+MCV+flow chart)	ANEM	B	KC	anaemia cascade
	Ferritin	328	ferritin	FERR	B	KC	ferritin
	Bilirubin	111	bilirubin total	BILI	B	KC	bilirubin
	folic acid	336	folic acid	FOLI	B	KC	folic acid
	Vitamin B ₁₂	1412	vitamin B ₁₂	VB12	B	KC	vit B ₁₂
	Reticulocytes	1276	Reticulocytes	RETI	B	HE	Reticulocytes
	Serum iron	323	iron	FE	B	KC	Serum iron
	LDH	534	lactaat dehydrogenase (LDH)	LDH	B	KC	LDH
	total iron binding capacity	1394	iron binding capacity total	TYBC	B	KC	transerrin/ TUBC
	Transferrin	1376	transferrin	TRFE	B	KC	transerrin/ TUBC
	indices MCV	582	MCV	MCV	B	HE	MCV
Chlamydia infections	chlamydia DNAtest, urine (PCR)	176	Chlamydia DNA urine	CHDN	U	IM	chlamydia urine
	chlamydia DNAtest, endocervical (PCR)	177	Chlamydia DNA urethra	CHDN	UU	IM	chlamydia endocervical/ vaginal
	chlamydia DNAtest, vaginal swab (PCR)	178	Chlamydia DNA cervix	CHDN	XC	IM	chlamydia endocervical/ vaginal
	antibodies chlamydia trachomatis, IgA	182	Chlamydia ab. IgA	CHIA	B	IM	antibodies CT
	antibodies chlamydia trachomatis, IgG	185	Chlamydia ab. IgG	CHLG	B	IM	antibodies CT
	antibodies chlamydia trachomatis, general	186	Chlamydia ab.	CHLI	B	IM	antibodies CT
	antibodies chlamydia trachomatis, IgM	191	Chlamydia ab. IgM	CHLM	B	IM	antibodies CT

(continued)

Module	Item	codeNr	description labcode table	Mnemonic	Mat	Group	label on feedback form
Prostate complaints	creatinine	523	creatinine	KREA	B	KC	creatinine
	PSA	896	Prostate spec. ag. (PSA)	PSA	B	KC	PSA
	PSA complex	1921	Prostate spec. ag. complex (PSA complex)	PSAC	B	KC	PSA
	Free/Total PSA-ratio	2124	Free/Total PSA-ratio	PSAR	B	KC	V/T-PSA
Rheumatic complaints	Waaler-Rose	1279	Rheumatoid factors, Waaler-Rose	RFWR	B	IM	(semi-)qualitative rheumatoid factors
	Latex fixation test	2154	Rheumatoid factor, Latex fixation	RFLF	B	IM	(semi-)qualitative rheumatoid factors
	IgM-rheumafactor	1278	Rheumatoid factor quantitative	RF	B	IM	IgM rheumatoid factor
	anti-cyclic citrullin peptide antibodies	2153	Rheumatoid factor (IgM ELISA)	RFIM	B	IM	anti-CCP
	ESR	134	sedimentation (ESR)	BSE	B	HE	ESR
	CRP	227	C-reactive protein (CRP)	CRP	B	IM	CRP
	X-wrist left	2239	xwrist le	X	LO	BO	X-hand/ wrist
	X-wrist right	2240	xwrist ri	X	LO	BO	X-hand/ wrist
	X-hand left	2243	xhand le	X	LM	BO	X-hand/ wrist
	X-hand right	2244	xhand ri	X	LM	BO	X-hand/ wrist
	antistreptolysin titer	81	antistreptolysin titer (AST)	AST	B	IM	AST

(continued)

Module	Item	codeNr	description labcode table	Mnemonic	Mat	Group	label on feedback form
thyroid dysfunction	TSH	1385	thyroid stimulating hormone (TSH)	TSH	B	KC	TSH
	thyroid diagnostics (cascade)						
	FT4	2059	thyroiddiag. (TSH if needed FT4)	SCHK	B	KC	thyroid cascade
	T3	348	T4 Free thyroxine	FT4	B	KC	FT4
		347	T3 Free thyroxine	FT3	B	KC	T3
	TSH-receptor stimulating antibodies	1390	TSH receptor ab.	TSRI	B	KC	TSI
	thyroid peroxidase (TPO) antibodies	1365	thyroidtissue ab.	THYI	B	IM	TPO-Ab
	ESR	134	sedimentation (ESR)	BSE	B	HE	ESR
	leucocytes	547	leucocytes	LEUK	B	HE	Leuco diff
	leucocytes differentiation	241	leucocytes, differentiation	DIFF	B	HE	Leuco diff
Urinary tract infections	ultrasound thyroidgland	2415	ultrasound thyroidgland	ECHO	TA	BO	US thyroid
	Urine culture with antibiotic resistance testing	1338	urine culture	STAK	U	BA	urine culture
Dyslipaemia	Total cholesterol	192	Total cholesterol	CHOL	B	KC	T-cholesterol
	HDL	446	HDL-cholesterol	HDL	B	KC	HDL
	LDL	542	LDL-cholesterol	LDL	B	KC	LDL
	triglycerides	1377	triglyceriden	TRIG	B	KC	triglycerides
	total cholesterol/ HDL ratio	181	cholesterol/HDL-cholesterol ratio	CHHD	B	KC	ratio T-cho/HDL
	Homocysteine	470	Homocysteine	HOMC	B	KC	Homocysteine
perimenopausal complaints	follikel stimulating hormone (FSH)	346	follikelstimulating hormone	FSH	B	KC	FSH
	luteinising hormone (LH)	553	luteinising hormone (LH)	LH	B	KC	LH
	estradiol	768	estradiol	OEST	B	KC	estradiol



Chapter 5

**The balanced incomplete block design is
not suitable for the evaluation of complex
interventions**

Jasper Trietsch, P Leffers, Ben van Steenkiste, Richard Grol, Trudy van der Weijden

J Clin Epidemiol 2014;67:1295-1298

Abstract

Objective

In quality of care research the Balanced Incomplete Block (BIB) design is regularly claimed to have been used when evaluating complex interventions. In this paper we reflect on the appropriateness of using this design for evaluating complex interventions.

Study design and setting

Literature study using pubmed and handbooks.

Results

After studying various papers on health services research that claim to have applied the Balanced Incomplete Block and the original methodological literature on this design it became clear that the applied method is in fact not a Balanced Incomplete Block design.

Conclusion

We conclude that the use of this design is not suited for evaluating complex interventions. We stress that, in order to prevent improper use of terms, more attention should be paid to proper referencing of the original methodological literature.

Introduction

Incorrect citing in scientific literature can lead to a chain of erroneous interpretations or the use of incorrect methods or terms. In this short report we describe the case of the balanced incomplete block (BIB) design. In quality of care research the BIB design is regularly mentioned in the sections reporting on the applied study design. Study of the literature on the BIB design led us to original work by Cochran and Cox from the 1950's describing a design that appeared to be very different from the designs that were actually used in the quality of care research papers claiming to have applied the BIB. This raised questions about possible consequences of improper reporting of the BIB design for our own research with respect to the interpretation of study results.

In this paper we describe the features of the original BIB design, and we compare it with the design that we used in our earlier research projects. We illustrate where we failed in citing of the literature on the BIB design, and we describe the implications of this error for the validity of conclusions of our research.

Features of the original BIB design

Comparing the outcomes of multiple interventions under various conditions is a well-known challenge in experimental research. It is often impossible to carry out such studies because of limitations on the number of available research subjects and because of limited resources. In agriculture a solution for this problem was developed for crop-optimization studies by statisticians involved in combinatorial problems research. These methods were thoroughly described by Cochran and Cox.¹ For instance, when testing several new genetic varieties of corn under different growth conditions, vast areas of land would be needed. The BIB design enables researchers to compare harvest returns of the varieties using plots of land (blocks) which each have different conditions between, but uniform conditions within these plots. Not all varieties of corn will be grown on each of these plots: incomplete testing. By balancing the allocation of the varieties over the different plots, a comparison of outcome (e.g. returns in harvest) can be made between varieties that were never really compared under the same conditions. The simplest example is that of difference in harvest between varieties A and C under condition I can be estimated, although they were not directly tested against each other under condition I. Under the assumption of absence of effect modification by condition the harvest difference between A and C can be calculated by comparing varieties A and B under condition I, and varieties B and C under condition II. Statistical testing is done using analysis of variance.^{1,2}

To speak of a BIB design several conditions have to be met. A theorem is available to test whether a design meets the requirements of a BIB design. The notation of the parameters of a BIB design is $\{b, r, k, v, \lambda\}$ (Box 5.1). Figure 5.1 shows a $\{3, 2, 2, 3, 1\}$ -design: a design of three blocks with a variation replication rate of 2, two varieties per block, three different varieties, and a pair replication rate of 1, as described previously.

Box 5.1 General theorem for BIB designs.

$bk = vr$ and $r(k - 1) = \lambda(v - 1)$ When: $v > k > 0$ $r > 0$ $\lambda > 0$	v = the number of compared (genetic) varieties b = the number of plots of land (blocks) r = the number of plots in which each variety is present k = the number of varieties per plot λ = the number of plots in which each pair of varieties is present (pair replication)
--	---

Features of the BIB design as applied in quality improvement research

According to Cochran and Cox the BIB design is suitable for situations in which repeated testing of varieties will lead to the same result, as can be expected when conditions can be well controlled such as in agricultural or laboratory sciences. Unfortunately in most types of clinical research patients will be permanently influenced by the intervention that is being evaluated and therefore repeated testing cannot be expected to lead to the same result. As a consequence the BIB design cannot be used for patient centered research. However several publications within quality of care research on complex healthcare interventions report on the use of the BIB³⁻⁶ and it was also advocated as appropriate for complex guideline implementation trials.⁷

Testing all components of the complex interventions separately is generally not possible because of limited resources or limitations on the number of available research subjects, let alone that all components can be tested under the various conditions. The applied design is claimed to overcome these limitations while it is also considered to be attractive because it controls for the Hawthorne effect.⁸ However,

from the published reports it can be concluded that the BIB design was not applied at all.

Our research group has used the same design in studies on complex quality improvement interventions.⁹⁻¹¹ An example is the work of Verstappen et al.. They performed a cluster RCT and claimed to have applied the BIB design. The strategy under study was the 'small group peer review', and the aim was reducing inappropriate test ordering by general practitioners. Thirteen general practitioner (GP) groups participated as units of research. They were randomized to receive feedback on one of two sets of clinical topics relating to the treatment of their patients (incomplete testing). Both arms acted as 'placebo comparator' for the other arm with the treatment and the placebo aspect of each treatment completely linked. Therefore, this was a cluster randomized trial with only two conditions. Figure 5.2 shows the typical design used in such trials. Put in terms of the BIB design: There are two varieties (GP groups with their allocated treatments; $v=2$) under evaluation. There are only two blocks (two variations of the intervention; $b=2$), each variety appears only once ($r=1$), the number of varieties per block is 1 ($k=1$), and the number of blocks in which each pair of varieties appear is 2 ($\lambda=2$). From checking these numbers with the conditions in Box 5.1, it is immediately clear that these studies do not fulfil the criteria for a BIB design.

Recently we completed the evaluation of a complex implementation strategy based on audit, provision of feedback, and educational materials with peer group discussion guided by clinical guidelines, claiming to use the BIB design.¹² This strategy was implemented in groups of GPs from the same region, and was moderated by community pharmacists and test ordering experts (local opinion leaders). We integrated the small group peer review strategy in routine health care, which resulted in an extremely pragmatic trial.¹³ With hindsight we now conclude that our strategy also did not meet the criteria of a proper BIB design.

Figures 5.1 and 5.2 show schematic representations of the simplest possible BIB design and of a typical trial in quality improvement research. It is immediately clear that a 2 armed trial cannot be a BIB design.

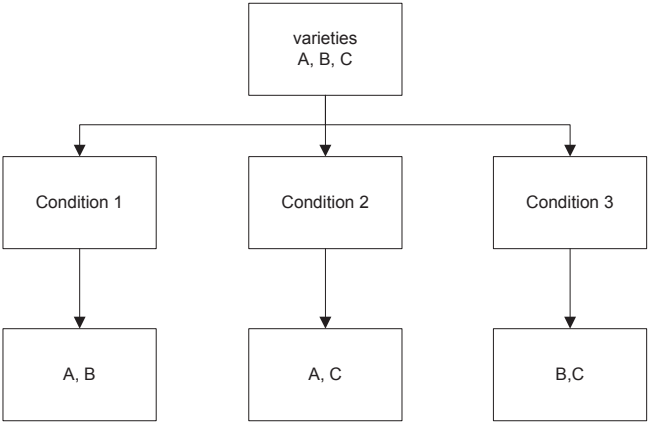


Figure 5.1 Example of BIBD with {3,2,2,3,1}-design.

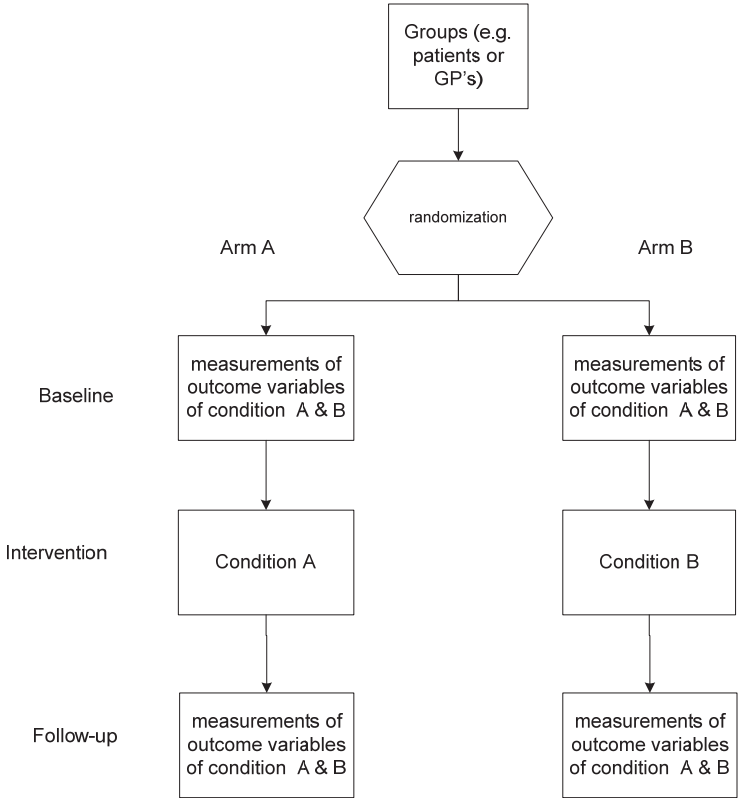


Figure 5.2 Typical design used in quality improvement research.

What went wrong and what is the impact?

Eccles and Grimshaw described using the BIB design in a similar setting in the area of quality improvement research as we did.³ When checking the references of their published studies, we were struck by the fact that authors regularly refer to earlier work from colleagues in the field and not to the original publications on the BIB design. The BIB design can be traced back in time to the work of Cochran and Cox and further back to the work of Bose and that of Yates and even to the famous ‘Kirkman’s schoolgirl problem’.^{1,14,15} It became clear that the used design does not match with the description of the BIB design by Cochran and Cox at all. The fact that that description is difficult to understand for non-statisticians makes it understandable that authors tend not to refer to it or refer to it without having read and understood the contents. This way, an interpretation error made by one group of authors has led to systematic inappropriate use of a methodological term. We think that the smaller the number of researchers in the field the less probable it seems that others will notice such an error.

The question remains what the implications of the improper use of the BIB design are. Fortunately we found that the authors performed appropriate statistical analyses for the design that they actually used. The validity of the conclusions of the publications should be judged on the assumptions that underlie the comparisons in these studies. The main assumption is that the feedback strategies can act as placebo strategies for each other. The improper use of the term ‘BIB design’ has no bearing on this.

Conclusions

The consequences of the improper citing to the BIB design have been very limited. However, authors, reviewers and editors should be alert when indirect citations are used maybe especially when complex methodological or statistical methods are involved that are not commonly used. In the present case, indirect referencing has only led to improper use of a beautiful and complicated name for a very standard study design. As admitting not being acquainted with a design may feel uncomfortable for many, we fear that the unjustified use of a complicated term will impede critical debate about the validity of methods of studies people. It certainly leads to failure to honour developers of research methodology for their contributions to science.

References

1. Cochran WG, Cox GM. Experimental designs. New York,: Wiley; 1950.
2. Wakeling IN, Buck D: Balanced incomplete block designs useful for consumer experimentation. *Food Quality and Preference* 2001;12:265-268.
3. Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D, Purves I. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ* 2002;325:941.
4. Kasje WN, Denig P, Stewart RE, de Graeff PA, Haaijer-Ruskamp FM. An educational programme for peer review groups to improve treatment of chronic heart failure and diabetes mellitus type 2 in general practice. *J Eval Clin Pract* 2006;12:613-621.
5. Campbell BF, Sengupta S, Santos C, Lorig KR. Balanced incomplete block design: description, case study, and implications for practice. *Health Educ Q* 1995;22:201-210.
6. Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D. An Evaluation of Computerised Guidelines for the management of Two Chronic Conditions. In: Newcastle: Centre for Health Services Research; 2002.
7. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract* 2000;17 Suppl 1:S11-16.
8. Verstappen WHJM, van der Weijden T, ter Riet G, Grimshaw J, Winkens R, Grol RPTM. Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. *J Clin Epidemiol* 2004;57:1119-1123.
9. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermesen J, Grimshaw J, Grol RPTM. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289:2407-2412.
10. Verstappen WHJM, van der Weijden T, Dubois WI, Smeele I, Hermesen J, Tan FES, Grol RPTM. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med* 2004;2:569-575.
11. Bindels R, Hasman A, Kester AD, Talmon JL, de Clercq PA, Winkens RAG. The efficacy of an automated feedback system for general practitioners. *Int J Med Informatics* 2003;11:69-74.
12. Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijters P, Winkens R, van Steenkiste B, Grol R, Metsemakers J. A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: Study Protocol. *Implement Sci* 2009;4:6.
13. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464-475.
14. Bose RC. On the construction of balanced incomplete block designs. *Annals of Human Genetics* 1939; 9:353-399.
15. Yates F. Incomplete randomized blocks. *Annals of Human Genetics* 1936;7:121-140.



Chapter 6

General discussion

Discussion

Outline of the Study

This chapter starts with presenting the main findings and the lessons learnt in our study. First the barriers and facilitators will be discussed, followed by the discussion on the effects of the intervention under study in the broader perspective of implementation science and quality improvement interventions. Hereafter, the methodological challenges that result from this type of research, will be considered. Finally, the implications for quality improvement interventions aimed at professional care providers, the implications for the Dutch situation and further research are discussed.

Every day general practitioners (GPs) order tests and prescribe medication to treat patients for a huge variety of medical conditions. Which tests are ordered and which drugs are prescribed differs between GPs as are the numbers of test ordering and prescribing. A gap remains between the care that is provided to patients and the recommended care.^{1,2} In the Netherlands local groups of GPs meet on a regular basis to discuss clinical topics and ideally also to set working agreements, in so-called local quality improvement collaboratives (LQICs). It seems however that the effect of this approach has reached its maximum given the remaining stable level of unwarranted practice variation.³⁻⁵ Many approaches have been studied to reduce this unwarranted practice variation. However, only few have been implemented and tested after implementation. We investigated whether the positive results of earlier research on the effect of peer review within LQICs would hold after large-scale implementation in normal educational sessions. We hope to contribute to the worldwide search for successful interventions to change professionals' behaviours and the identification of the effective elements.⁶⁻¹⁰

We chose to use the existing backbone of LQICs in The Netherlands to implement our strategy. The majority of LQICs are used to discuss prescribing behavior for certain conditions. We broadened the content to discussing test ordering performance of GPs. The main components of the intervention consisted of adding feedback to the GPs on prescribing and test ordering behavior together with peer group discussion and setting working agreements in a structured format. We chose not to interfere with the meetings, not even by being present at the meeting as observer or moderator, to mimic as much as possible the natural situation that we expected to occur after dissemination of the intervention.

The research questions for this work were:

- Was the strategy implemented as planned?
- What is the effect of the strategy on test ordering and prescribing by GPs?
- What were the barriers to and facilitators of the implementation of the strategy?
- Is the balanced incomplete block design suitable for use in implementation science?

Main findings

- The pre-existing evidence on effectiveness, the national infrastructure for these collaboratives and a general positive attitude towards the peer review intervention were not sufficient for normalization; implementation of this intervention failed.
- The GP in the LQICs tended to stay in a passive role of wait and see due to a perceived lack of support.
- When the source of the feedback data is disputed, this has a negative impact on its effect in changing practice behavior.
- No effects were found for the intervention, being audit and feedback embedded in small peer groups, after implementation.
- The balanced incomplete block design is hardly ever suitable for research with humans.

Barriers and facilitators

During the execution of this study we encountered more difficulties than anticipated during the planning phase. In-depth interviews with stakeholders showed problems with the implementation strategy at all levels despite the stakeholders' willingness to cooperate in this trial and the tailoring of the intervention to the existing Dutch infrastructure of LQICs.¹¹ The stakeholders can be distinguished by their role in the process and subsequently be divided into three levels, central, intermediate and the local level. The first level is the research group with facilitating bodies such as the IT-department. They only facilitated the process within the LQIC. The intermediate level is the level of the health officers and laboratory specialists as experts on their field and they acted as moderators of the peer review meetings. The local level is the level of the groups with the general practitioners, they were the main target in this study as their prescription behavior and test ordering behavior was at stake.

Central level

First of all we saw problems in the feedback procedures. Feedback is a central element in the intervention. Although we executed a small pilot prior to this study to test the

procedures of generating feedback and delivering it to the LQICs, it appeared to be more complex than expected to generate feedback data and to produce feedback reports in time. Gathering data for feedback took a lot of time and was not always successful, although many of the problems were solved by redesigning parts of the data flow. For example, we anticipated on the low priority given by the laboratories to the data delivery, by requesting data more in advance. Generating feedback under responsibility of laboratory specialists at their own site, hereby skipping the bothersome task of processing data at a central level, may have facilitated the implementation process. However, transferring the feedback-generating technology to these sites and decentralizing prescribing data to these laboratories would have been a difficult operation with a considerable risk of failure. In that case we also would have had to unravel the prescription data that was already available in a much more aggregated way. Also the working area of local laboratories is not always congruent with the geographical area that members of a single LQIC group originate from. It also occurred that some group members choose to send their patients to another laboratory for diagnostic testing, outside their own region. By decentralizing the generation of feedback to local laboratories this mismatch reduces the comparability and the quality of the feedback as the feedback is based only on tests ordered from one diagnostic facility. When a group member is located on the geographical border of that specific facility many of his patients might choose to go to an adjacent diagnostic facility, these orders would in this case not be presented in the feedback.

A second important element is a good moderation of the local groups. We anticipated on routines and tendencies in LQIC groups by training the moderators in the intervention to specifically be alert to deal with strong local routines. It has to be concluded that this approach was insufficient. The moderators that were provided by the research project group in the successful prior study apparently had been talented early innovators who were able to build on extensive prior experience for this task.² In this trial a similar intervention like ours, with audit and feedback in LQICs, was tested but in a more controlled setting and only aimed at test ordering performance. We did not do a new search for barriers and facilitators perceived by the participants during the pilot phase as we trusted on our experience with this former study, assuming that the important lessons learned would still be valid. With hindsight it would have been wiser to identify the barriers prospectively as this might have helped tailoring the intervention to the needs and beliefs of the current participants and thus could have increased acceptance and implementation.¹²⁻¹⁴ However, whether this would have really helped tailoring the intervention remains unclear as it is still not known what methods for tailoring could best be used.¹⁵

Intermediate and local level

At the intermediate and local level, stakeholders reported that it was unclear to them who were responsible for the work (generating test-ordering data, organizing meetings and moderating groups). It seems that group members tended to adopt a passive role in taking up this implementation. In our extremely pragmatic trial, only an outline of the strategy was given, together with the minimum requirements that had to be met. In the classical way these groups usually work with one GP being responsible together with a local pharmacist to prepare a meeting from beginning to end. They usually start with presenting one or more case reports that act as a starting point for a presentation on the topic based on the guidelines. The main target is often sharing knowledge without reaching agreement on future policy or setting working agreement. We did however provide them with the feedback on their performance and an excerpt from the available guidelines. With this information they were challenged to discuss their performance in comparison with the performance of their peers in the group and the guidelines provided in a session moderated by experts on test ordering or pharmacotherapy. Many groups were not used to such structured discussions about performance on the basis of feedback. Most groups still seem to held on to outdated CME models known to be hardly effective.^{16,17} Forces within groups such as free-riding, competition and even anti-cooperative behavior were found in other studies.¹⁸ As mentioned before, the training we provided was designed to deal with this shift in routines by instructing the moderators how to achieve this shift. From the interviews the suggestion emerged that many GPs felt that the existing structure with informal meetings did not require much preparation and were an easy way to earn CME credits, while the new way of working was seen as much more elaborate. The GPs were especially critical about the prescribing data. This might partly be caused by distrust towards insurance companies that provided us with these data, but possibly also in reaction to the open way the feedback was provided. These defensive mechanisms towards feedback data have also been found in earlier research.^{19,20} Participants in our trial were pleased with the central registration of CME points, which was an unexpected facilitator of the process. However, whereas the aim of the study was to transfer responsibilities to the field, this aspect of certification ironically shifted from the local to the central (academic) level.

The framework used for evaluating the implementation

We used the normalization process model for gaining insight in the barriers and facilitators of this study. This model was derived from the normalization process theory developed by May et al..²¹ The model describes all different levels of organizations that are involved in implementing change in healthcare and within these levels the

resources that are needed in order to achieve the desired change. The resources are not only financial resources but also the work that has to be done or the shift in mind set that is needed as well as practical issues such as where are we going to meet. It helps mapping all relevant factors that contribute or hamper the implementation. As the model originates from social sciences the description from the dimensions has a sociological background. We found the model with its sociological perspective and the focus on the work that comes with implementing helpful in mapping the barriers and facilitators. Nevertheless we feel that the descriptions of the dimensions are difficult to grasp. They are often elaborately described in a language that is difficult to understand. Also the different ecological levels that are involved in implementation such as the research group, opinion leaders, and the participants cannot always clearly be categorized when using the NPM. Improving the language to more plain language and providing guidance in how to deal with the complexity of the different ecological levels would help when using the NPM.

Effects of the intervention

Our study shows that the impact of our intervention in existing LQICs in general practice was less than expected. This expectation was based on the results from earlier well-controlled studies on LQICs in primary care. The impact of audit and feedback with peer review in LQICs in our study was absent or small, particularly in the crude before and after analysis. The per-protocol analyses showed effect on several clinical topics but these results are not as convincing as found in the earlier trials. However, we observed that GPs from practices known for their high volume of tests ordered and drugs prescribed showed the largest improvement. Although the effect sizes found were often significant, clinical relevance was sometimes lacking due to small baseline volumes. Irwin et al. also reported in a review of reviews on quality improvement at the practice level in primary care that physicians with the lowest baseline performance showed most improvement.²² Whether their understanding of low baseline performance includes the high volume practices we defined remains unclear but seems logical.

Although audit and feedback on test ordering behaviour embedded in peer review in small groups is more effective than audit and feedback alone, it generally remains unclear exactly what factors contribute to this effect.^{14,15,23-26} Multifaceted interventions like ours are complex by nature but seem attractive assuming that the effects of separate components add up. Whether this assumption stands is still unclear as is shown in different reviews where effectiveness of single versus multifaceted interventions is compared. Mostofian concludes in a review of reviews that multifaceted interventions are most effective in changing professional behaviour.²⁷ This

view is shared by Johnson and May as well as other researchers; they find it likely that multifaceted interventions are more effective than single interventions.^{2,14,22} On the other hand Irwin et al. conclude that there is no evidence for a larger effect for combined interventions while single interventions aimed at professionals such as educational outreach, academic detailing and workshops have been proven to be effective in changing the targeted behaviour also.^{17,27-30} The reports that underpin these results however stem mainly from research conducted in settings outside general practice. In our opinion the hypothesis that multifaceted interventions have a larger impact than single interventions^{22,31} is very likely to be true in the general practice setting, due to the wide variety of influencing factors.

Embedding the discussion of clinical topics in LQICs showed a modest positive effect on prescribing costs and quality.^{9,32-37} Much remains unclear about what factors are decisive for the effect of LQICs.¹⁰ In trials using LQICs for research on effective change strategies often new groups had to be formed. Using the existing infrastructure of groups is likely to change practice more effectively than when new groups had to be formed. This however is to our knowledge not tested against each other.

The use of a pragmatic design to mimic real life

Much work has been done on evaluating the effects of audit and feedback on both test ordering and prescribing behaviour in well-controlled trials.¹⁵ As stated in the introduction of this thesis evaluating the effect of large-scale implementation using an evidence-based effective quality improvement (QI) strategy in a pragmatic design is not often performed. We found that most research is focused on one particular clinical topic (e.g., antibiotic prescribing in urinary tract infections, X-ray in low back pain patients), while we applied the peer review strategy in a broad range of test ordering and prescribing behaviours. The interventions in earlier studies show on average a modest effect on changing professional behaviour. However, no strong conclusions can be drawn from these studies, given the difficulties in interpretation due to heterogeneity of the clinical topics, designs and settings of the trials performed.^{17,22,26,31,38-43} They also were more situated on the explanatory side of the explanatory-pragmatic continuum because most researchers acted as moderators for the study groups themselves or new groups were formed to discuss test ordering or prescribing. After finishing this thesis this is still one of the first reports on an intervention with a pragmatic design in QI research. The use of such a design in QI research helps to bridge the still present gap between academia and the field.^{2,44} Whether our intervention was really a pragmatic intervention can be argued. As we did plan to alter the routines of the existing groups and trained the local opinion leader to moderate the group. Still, we feel that we rightfully claim this intervention to be

pragmatic because we transferred the knowledge on how to improve the effectiveness of LQICs, working with feedback and goal setting to the moderators. Hereafter we only provided the groups with feedback by post or email, without any contact otherwise. We believe this is in line with the definition of pragmatic used in other reports on pragmatic trials.^{42,45} By choosing this approach we did have little influence on the diagnostic experts recruiting the groups even if it failed. We could have changed this strategy halfway the recruitment phase. This would have meant however that we would have had to gain insight in the organisation of groups in the south of the Netherlands in detail, thus creating an unrealistic large workload for our small team and decreasing the pragmatic nature of our trial to great extent. As the diagnostic experts did not have a close relationship to the general practitioners in their area before, they were supported by the regional primary care association in identifying and approaching the right contacts of groups. We conclude that for effective implementation in normal practice there has to be present a well-functioning infrastructure but also an organisation that takes the lead. This last condition was not sufficiently met in our intervention, possibly because of the pragmatic nature of our intervention.

Working agreements and cost-benefits

Unfortunately only a minority of the working agreements of each group could be analysed for the specific goals that were set. The groups reported 76 working agreements made in 115 meetings in total. Over half of these working agreements, however, did not contain specific, measurable, attainable, relevant and timely (SMART) agreements that could be measured and analysed in a quantitative manner. For example, we received working agreements such as “our group will try to be more compliant with the guideline”, which doesn’t provide a SMART goal that allowed us to analyse if their own working agreement was implemented into practice.

Because of the small effects found in this intervention we did not execute a cost-benefit analysis as was planned in the protocol. With such small effect sizes we considered it to be highly unlikely that the savings in tests ordered and drugs prescribed would outweigh the costs. Also because of the small effect sizes we did not analyze the predictive value of the gain in level of group performance achieved on the volume of tests and drugs.

Strengths and limitations of the methods used

We repeated an earlier trial from our research group, and applied it within existing structures for QI in primary care.⁴⁶ We added the component of discussing prescribing behaviour. This component was also previously tested in another well-controlled

positive trial on prescribing behaviour for one clinical topic.⁴⁷ Testing the effects of a QI-intervention using a pragmatic cluster-RCT is a solid method for testing whether the effects shown in previous trials hold after implementation. In our current pragmatic trial, it was difficult to control for all influencing factors, since we aimed to mimic normal QI-practice as much as possible. This might have compromised the internal validity of the study. However, by using cluster randomisation, blinded analyses of the results and gathering baseline data prior to randomisation, we expected to guarantee sufficient level of internal validity without decreasing the external validity.⁴⁸ Less groups than calculated before entered the trial leading to an underpowered study. As is discussed later a possible source for this is the healthcare reform that was initiated by the government after initiation of the trial. Would we have been able to include more groups the effects probably would have been clearer, not necessarily larger.

We designed our study according to a so-called balanced incomplete block design. This is a special design from the group of Latin square designs. The balanced incomplete block design is a classical block design where not all combinations of species and circumstances are tested (incomplete) against each other but some are left out according to several rules and assumptions (balanced). This model seems to work perfectly for agriculture; a specific genus of corn can be planted at the same time in field A and in field B under different conditions and alongside different genera without it even knowing that an exact copy of itself is growing in another field. In research using human subjects these assumptions are only rarely logical or valid. The only way to meet the criteria is when a person is tested first under a certain condition against other people and hereafter tested again under different conditions or against other persons. Would such a design be executed the second test result would be influenced as the subject has learned from the first test. Also the analysis of the results from a BIBD requires specific statistical tests and relies on several assumptions that can hardly be considered a reflection of the reality in a pragmatic trial in humans.

The design that is commonly used in implementation research should therefore better be named a 2-armed trial or a 2 by 2 Latin square design. The statistical tests needed then are also less complicated and no assumptions are required for the analysis. The BIB design is however attractive because of the elimination of non-specific effects; the so called Hawthorne effect.⁴⁹⁻⁵² One of these non-specific effects is the effect that people will act differently, presumably better, when they know they are being observed.^{53,54} The design we chose, erroneously referred to as a BIB design in our protocol paper, helped to minimize these kinds of non-specific effects as both trial arms were exposed to the same type of intervention. But as the groups were free to choose their clinical topics as well as the timing of sessions and the order of the topics a learning effect could have occurred and influenced the results. Some groups discussed

a specific topic at the start of the intervention period, while in other groups this topic was discussed at the end of this period. The first group then would have been naive towards how the intervention works while the second group already has experience on working with the feedback at the time they discuss the same topic. In future research this learning curve should be considered as a variable.

A dilution of the effects of the intervention could have occurred due to the intention to treat analysis. All groups in a single arm were analysed as if they had worked on all the clinical topics available in their arm but only three out of five clinical topics were chosen. Before the effects of the intervention were calculated we therefore added a per protocol analysis, additional to our plans for analysis as described in the protocol,⁵⁰ that only takes into account those topics that were actually discussed for each group.

As low baseline adherence is a known factor for effective implementation of audit and feedback strategies, it might have been attractive to identify practices with the highest test ordering and prescribing volumes beforehand and target them in our trial.^{22,31} However this would have forced us to leave the attractive environment of existing LQICs and form new groups. We deliberately choose this existing structure as target for our study as we expected several benefits from this. Firstly, discussing personal feedback data is easier because many physicians belong to the same LQIC group for years assuring a stable and safe environment. Secondly, implementing our strategy only demanded adding of experts on test ordering to the LQICs, leaving the basic structure untouched. As these groups were existent for many years with the goal to discuss prescribing behaviour, a lack of motivation to critically assess their own performance seems highly unlikely. The negative effects of an organisational reform, which are often encountered in complex interventions, were prevented by choosing the existing groups for our intervention.⁵⁵⁻⁵⁷

Format and source of the feedback

One of the criteria for effective audit and feedback as defined by Ivers et al. in their Cochrane review is on the format of the feedback provided. They state that feedback should be provided more than once preferably in a written and verbal format, include clear goals and should best be delivered by a senior or colleague.³¹ Although we did provide the feedback through the moderators and provided the feedback more than once as it was discussed in two meetings, we did not provide predefined benchmarks or goals. We did however provide some benchmark for the groups in the form of the aggregated results from their own and neighbouring groups, presented together with the relevant recommendations from clinical guidelines. One of the main issues of concern learned during the study was that the source of our feedback was not clear to the participants. Training of the moderators on how to facilitate setting specific targets

for improvement on group (and individual) level and also on answering questions on the source of the data was insufficient to deal with this last point.

We chose to express the volume of prescribed drugs in daily defined dosages (DDDs). A risk of using this method is that not all DDDs are compatible with the actual dosage physicians prescribe to a patient. For diclofenac, for instance, the normal dosage is 1.5 to 2 DDDs every day. This does however not affect comparability between groups, as both groups are affected by this form of distortion in the same way. Other available scales, such as the drug utilisation 90% (DU90%), have different problems inherited, resulting in over- or underestimating volumes and reliability.^{58,59} This indicator focuses on 90% of the delivered drug volume in a specific group of drugs, thus allowing some variation based on individual characteristics. The DU90% reflects how many different specific drug types are present within the first 90%. A high prescription rate of non-preferred drugs could thus still lead to a high score on the DU90%.

Volume as indicator is easy usable for GPs where the use of quality indicators, expressed as proportions of patients with indications for certain interventions, may be less comprehensible for GPs. To expand and sustain the peer review strategy after ending this intervention, an easy transferrable method such as simple volume data had to be available for use on a decentral level. All other quality indicators demand that they are edited by more specialised persons and are less easy to explain to physicians. Feedback on quality indicators next to the volume data might have generated stronger and more valid insights in performance. But volume data provide meaningful insights, especially with regard to practice variation. Volume data proved to lower volumes in the previous trials especially in areas with overuse.⁴⁶

Prescription behaviour is difficult to change, or at least to measure change. Due to the washout of recent prescriptions and the challenge of informing patients about changes in prescribing routines, it takes time to change, and change will only be gradual.⁶⁰ Also we were only able to provide feedback on prescribed medications. Therefore, some drugs that can be prescribed by physicians but are also available as over the counter drugs (OTC) are not included in this trial. Antacids for instance are not included in the feedback, although according to guidelines an increase of prescriptions of these drugs would be preferred combined with a concurrent decrease in short-term protonpump inhibitors prescribing rates.

Changing professional behavior in an era of transition

In 2006 a large healthcare reform was set in motion in the Netherlands. The goal of this reform was cost containment by shifting tasks from specialist care to general practice. Insurance companies were put in the lead of controlling costs and improving quality of care. At the same time a shift towards a mixed pay for performance model with

capitation fees was effectuated.⁶¹ This caused a disruption of usual practice with many uncertainties for GPs. As the Dutch health care system became more and more market regulated and many tasks were transferred to general practice, GPs started feeling threatened in autonomy and time management. Also the GPs were more than ever controlled by external parties such as healthcare insurers and, to a lesser degree, the inspectorate of health. Many GPs felt overburdened after this reformation was initiated, this resulted in resistance to new interventions and to insurance companies in general. As our intervention was partly sponsored by a large insurance company this could have had a negative influence on recruitment. We anticipated on the feeling of being overburdened by designing our strategy so that the burden of administration that could come with this intervention was limited. However one of the difficulties reported by groups during the implementation process was that the confidence in our strategy was low, even lower than we had expected, resulting in less engagement and a lack of participation with the intervention.

The topics we chose for this intervention were balanced over the two arms concerning the prevalence of the medical conditions, the domain covered (i.e. urological problems) and the relevancy to GPs. All topics are well described in national and international guidelines. Regretfully, after enrollment of LQICs into the trial, an incentive driven attention for high prevalence diseases such as diabetes type 2 and cardiovascular risk management was initiated as part of the healthcare reformation. This forced us to neglect all data from these major topics. Would we have been able to include these topics in our results, probably more robust results would have shown as was found elsewhere.⁶²⁻⁶⁵

Policymakers at the government and insurance companies in the Netherlands are looking for ways to improve the quality of care. It seems hard for them to get an understanding of what high quality of care precisely is about. Since the start of the reformation they search for ways to influence professionals with the goal to improve quality. Many new regulations, forms to be filled out and sets of indicators to be ticked where introduced in healthcare. Individual GPs seem in general not fully aware of the options available to policymakers to enforce the change in quality of care they perceive as needed using financial tools. In an attempt to prove the added value of their members and to prevent stricter regulation, professional associations also started to implement new sets of indicators, audits and peer audits. The healthcare professionals are on the receiving end of all these initiatives and feel overburdened and overmanaged. Therefore researchers and policymakers should operate cautiously when implementing new sets of indicators and interventions.

The final stage of implementation

Much still remains unknown about the effects of interventions after implementation especially concerning changing professional behaviour. Policy makers tend to adopt the effects of relative small and well controlled trials and apply these to the field as a whole. If we would accept this in case of implementing new drugs, we would allow new drugs to enter the market after phase II or III research was completed. The last phases then would almost never be executed. So, it is not logical to implement new strategies to improve quality of care or to change professional behaviour without complete evidence. Policymakers should be encouraged not to initiate implementation of strategies before sound phase III evidence is available. Furthermore together with implementing a new strategy it should be obliged to plan, fund and initiate phase IV post marketing surveillance of the side-effect and costs of implementation strategies. It would be a large improvement when interventions are designed together with practitioners instead of top-down and when they do not have a regulatory or financial background. This approach has been proven highly effective in changing practices in nuclear facility safety.^{66,67} In general practice the LQICs could take up part of that role if properly trained and moderated. A new program would be needed comparable to the program initiated in the 1980's that was used to promote LQICs in general practice in the Netherlands.⁶⁸ Researchers worldwide also should better finish their phase II and III research with the notion that the results found are only valid under the circumstances tested and are not necessary predictive of the effects after large-scale implementation.

Conclusions and recommendations

The impact of our intervention on GP performance was very small. The burden of running the intervention and the administration that came with it proved to be high. Would this intervention be implemented in all LQIC groups nationwide a large amount of intensified support would have to be made available. Despite that we combined interventions of proven effectiveness and implemented the intervention in an existing network of LQICs the strategy was not adopted widely. As many groups tended to take up a passive role and a wait and see attitude the uptake of the strategy might have been larger would we have made a trained facilitator available to the groups, being not only a moderator. Such a facilitator should actively search for perceived barriers within groups, address them and search for solutions. The problems with the source of the feedback could then also be handled more effectively. In large scale implementation research in a decentralized setting like ours this would mean a serious investment to make such support available. Our expectation that the moderators could act as such

facilitators after being exposed to our short training proved to be incorrect. An intervention with a trained facilitator guiding a group through the process of audit and feedback could be effective but possible overburdening of GPs should be avoided.

Identify GPs with a below quality baseline performance representing an unwarranted deviation from the mean and target those GPs in this type of quality improvement initiatives might be interesting. It could lead researchers and policy makers to shift focus. Other physicians who are already doing well can concentrate on what they are doing already; delivering high quality care. This shift of attention to GPs that show a low baseline performance is only valid when low baseline performance or guideline adherence proves to be consistent behaviour at an individual level. If not, each clinical topic requires a new pre assessment to identify those GPs that are to be targeted for that specific topic. Further research is needed on whether low baseline performance is consistent behaviour for an individual GP. Also further research on the cut-off point for participants that can benefit from a QI intervention like this is needed to identify the population to be targeted best. The meaning of low baseline performance would then have to be clarified and defined. In an era with much emphasis on cost containment in healthcare, an intervention targeted at only the GPs with a low baseline performance could provide policy makers with a tool to improve the quality of care and reduce costs at the same time without the risk of overburdening the whole group. By targeting only the high volume segment of physicians, all other physicians can focus on their regular work and the limited resources could be allocated to improving professional behaviour among those that really need it. This hypothesis needs further research with evaluation of potential downsides of such an approach such as the loss of peer learning with lower performing GPs learning from the best practices of their colleagues in LQICs.

Further research should be performed to confirm our findings that, after large-scale implementation of a multifaceted intervention based on audit and feedback with peer group discussions, the results found in earlier well-controlled trials are not easily replicated. This study raises methodological questions about how to deal with a highly pragmatic design that includes much room for local adaptation. We encourage other researchers to perform vigorous evaluations of large-scale transfers of complex implementation strategies, preferably embedded and owned by the field, as we did.

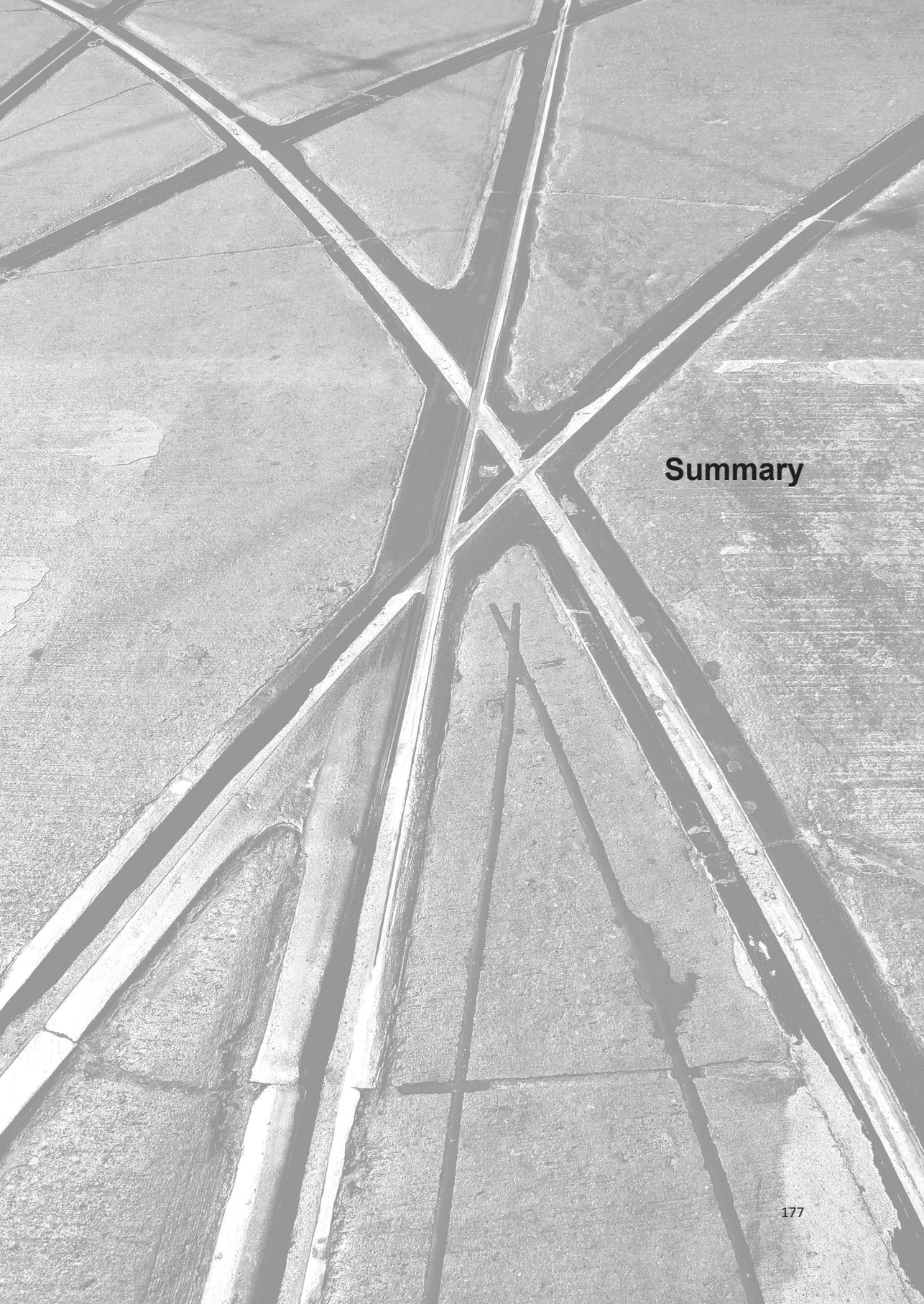
References

1. Van de Velde S, Roshanov P, Kortteisto T, Kunnamo I, Aertgeerts B, Vandvik PO, Flottorp S. Tailoring implementation strategies for evidence-based recommendations using computerised clinical decision support systems: protocol for the development of the GUIDES tools. *Implement Sci* 2016;11:29.
2. Williamson M, Cardona-Morrell M, Elliott JD, Reeve JF, Stocks NP, Emery J, Mackson JM, Gunn JM. Prescribing Data in General Practice Demonstration (PDGPD) project - a cluster randomised controlled trial of a quality improvement intervention to achieve better prescribing for chronic heart failure and hypertension. *BMC Health Serv Res* 2012;12:273.
3. Grol RP. Successes and Failures in the Implementation of Evidence-Based Guidelines for Clinical Practice. *Medical Care* August 2001;39:II-54.
4. Grol RPTM, Grimshaw JM. From best evidence to best practice: effective implementation of change in patients' care. *The Lancet* 2003;362:1225-1230.
5. Slade SC, Kent P, Bucknall T, Molloy E, Patel S, Buchbinder R. Barriers to primary care clinician adherence to clinical guidelines for the management of low back pain: protocol of a systematic review and meta-synthesis of qualitative studies. *BMJ Open* 2015;5:e007265.
6. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ* 1998; 317:465-468.
7. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA : the journal of the American Medical Association* 2012;307:1513-1516.
8. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. *PLoS One* 2013;8:e78962.
9. Rohrbasser A, Mikan S, Harris J. Exploring why quality circles work in primary health care: a realist review protocol. *Syst Rev* 2013;2:110.
10. Nadeem E, Olin SS, Hill LC, Hoagwood KE, Horwitz SM. Understanding the components of quality improvement collaboratives: a systematic literature review. *Milbank Q* 2013;91:354-394.
11. Trietsch J, van Steenkiste B, Hobma S, Frericks A, Grol R, Metsemakers J, van der Weijden T. The challenge of transferring an implementation strategy from academia to the field: A process evaluation of local quality improvement collaboratives in Dutch primary care using the normalization process theory. *J Eval Clin Pract* 2014;20:1162-1171.
12. Chaillet N, Dube E, Dugas M, Audibert F, Tourigny C, Fraser WD, Dumont A. Evidence-based strategies for implementing guidelines in obstetrics: a systematic review. *Obstet Gynecol* 2006;108:1234-1245.
13. Chaillet N, Dube E, Dugas M, Francoeur D, Dube J, Gagnon S, Poitras L, Dumont A. Identifying barriers and facilitators towards implementing guidelines to reduce caesarean section rates in Quebec. *Bull World Health Organ* 2007;85:791-797.
14. Johnson MJ, May CR. Promoting professional behaviour change in healthcare: what interventions work, and why? A theory-led overview of systematic reviews. *BMJ Open* 2015;5:e008592.
15. Wensing M. Implementation science in healthcare: Introduction and perspective. *Z Evid Fortbild Qual Gesundheitsw* 2015;109:97-102.
16. Hobma S. Directed self-learning as approach to continuing professional development. Maastricht: Maastricht University; 2005.
17. Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien Mary A, Wolf F, Davis D, Odgaard-Jensen J, Oxman Andrew D. Continuing education meetings and workshops: effects on professional practice and health care outcomes. In: *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd; 2009.
18. Carter P, Ozieranski P, McNicol S, Power M, Dixon-Woods M. How collaborative are quality improvement collaboratives: a qualitative study in stroke care. *Implement Sci* 2014;9:32.
19. Forsner T, Hansson J, Brommels M, Wistedt AA, Forsell Y. Implementing clinical guidelines in psychiatry: a qualitative study of perceived facilitators and barriers. *BMC Psychiatry* 2010;10:8.

20. Kaufmann-Kolle P, Szecsenyi J, Broge B, Haefeli WE, Schneider A. [Does implementation of benchmarking in quality circles improve the quality of care of patients with asthma and reduce drug interaction?]. *Z Evid Fortbild Qual Gesundheitsw* 2011;105:389-395.
21. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, Gask L, MacFarlane A, Murray E, Rapley T, et al. Understanding the implementation of complex interventions in health care: the normalization process model. *BMC Health Serv Res* 2007;7:148.
22. Irwin R, Stokes T, Marshall T. Practice-level quality improvement interventions in primary care: a review of systematic reviews. *Prim Health Care Res Dev* 2015:1-22.
23. Verstappen WHJM, van der Weijden T, Dubois WI, Smeele I, Hermsen J, Tan FES, Grol RPTM. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med* 2004;2:569-575.
24. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21:387-395.
25. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, Grimshaw JM. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci* 2014;9:14.
26. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implementation science : IS* 2015;10:167.
27. Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing physician behavior: what works? *Am J Manag Care* 2015;21:75-84.
28. Wensing M, Grol R. Single and combined strategies for implementing changes in primary care: a literature review. *Int J Qual Health Care* 1994;6:115-132.
29. Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, Robertson N. Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2010;(3):CD005470.
30. Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, Robertson N, Wensing M, Fiander M, Eccles MP, et al. Tailored interventions to address determinants of practice. *Cochrane Database Syst Rev* 2015;4:CD005470.
31. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O'Brien MA, Johansen M, Grimshaw J, Oxman AD. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;6:CD000259.
32. Veninga CC, Denig P, Zwaagstra R, Haaijer-Ruskamp FM. Improving drug treatment in general practice. *J Clin Epidemiol* 2000;53:762-772.
33. Engels Y, Verheijen N, Fleuren M, Mekkink H, Grol R. The effect of small peer group continuous quality improvement on the clinical practice of midwives in The Netherlands. *Midwifery* 2003;19:250-258.
34. Wensing M, Broge B, Kaufmann-Kolle P, Andres E, Szecsenyi J. Quality circles to improve prescribing patterns in primary medical care: what is their actual impact? *J Eval Clin Pract* 2004;10:457-466.
35. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. *Bmj* 2008;336:1491-1494.
36. Wensing M, Broge B, Riens B, Kaufmann-Kolle P, Akkermans R, Grol R, Szecsenyi J. Quality circles to improve prescribing of primary care physicians. Three comparative studies. *Pharmacoepidemiology and Drug Safety* 2009;18:763-769.
37. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Niessen LW, Grol RP. Short- and long-term effects of a quality improvement collaborative on diabetes management. *Implementation science : IS* 2010;5:94.
38. O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in Australia. 1999;318.
39. Foy R, Eccles MP, Jamtvedt G, Young J, Grimshaw JM, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC Health Serv Res* 2005;5:50.
40. Ho MJ, Venci J. Improving the success of mailed letter intervention programs to influence prescribing behaviors: a review. *J Manag Care Pharm* 2012;18:627-649.

41. Giguere A, Legare F, Grimshaw J, Turcotte S, Fiander M, Grudniewicz A, Makosso-Kallyth S, Wolf FM, Farmer AP, Gagnon MP. Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;10:CD004398.
42. Ivers NM, Tu K, Young J, Francis JJ, Barnsley J, Shah BR, Upshur RE, Moineddin R, Grimshaw JM, Zwarenstein M. Feedback GAP: pragmatic, cluster-randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care. *Implement Sci* 2013; 8:142.
43. van der Velden AW, Kuyvenhoven MM, Verheij TJ. Improving antibiotic prescribing quality by an intervention embedded in the primary care practice accreditation: the ARTI4 randomized trial. *J Antimicrob Chemother* 2015.
44. Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract* 2000;17:192-196.
45. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.
46. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermesen J, Grimshaw J, Grol RPTM. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289:2407-2412.
47. Lagerlov P, Loeb M, Andrew M, Hjortdahl P. Improving doctors' prescribing behaviour through reflection on guidelines and prescription feedback: a randomised controlled study. *Qual Saf Health Care* 2000;9:159-165.
48. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003;3:28.
49. Verstappen WHJM, van der Weijden T, ter Riet G, Grimshaw J, Winkens R, Grol RPTM. Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. *Journal of Clinical Epidemiology* 2004;57:1119-1123.
50. Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijers P, Winkens R, van Steenkiste B, Grol R, Metsemakers J. A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: Study Protocol. *Implement Sci* 2009;4:6.
51. Konstantinou GN. Pragmatic trials: how to adjust for the 'Hawthorne effect'? *Thorax* 2012;67:562.
52. Godman B, Wettermark B, Miranda J, Bennie M, Martin A, Malmström RE. Influence of multiple initiatives in Sweden to enhance ARB prescribing efficiency following generic losartan; findings and implications for other countries. *International Journal of Clinical Practice* 2013;67:853-862.
53. Malchow C, Fiedler G. Effect of observation on lower limb prosthesis gait biomechanics: Preliminary results. *Prosthet Orthot Int* 2015.
54. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ* 2015;351:h4672.
55. Grol R, Thomas S, Roberts R. Development and implementation of guidelines for family practice: lessons from The Netherlands. *J Fam Pract* 1995;40:435-439.
56. Grol R, Dalhuijsen J, Thomas S, Veld Cit, Rutten G, Mekkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998;317:858-861.
57. Burgers JS, Grol RPTM, Zaat JOM, Spies TH, van der Bij AK, Mekkink HGA. Characteristics of effective clinical guidelines for general practice. *Br J Gen Pract* 2003;53:15-19.
58. Teichert M, van der Aalst A, de Wit H, Stroo M, De Smet PA. How useful are prescribing indicators based on the DU90% method to distinguish the quality of prescribing between pharmacotherapy audit meetings with different levels of functioning? *Eur J Clin Pharmacol* 2007;63:1171-1177.
59. Martirosyan L, Voorham J, Haaijer-Ruskamp FM, Braspenning J, Wolffenbuttel BH, Denig P. A systematic literature review: prescribing indicators related to type 2 diabetes mellitus and cardiovascular risk management. *Pharmacoepidemiol Drug Saf* 2010;19:319-334.
60. Boesten J, Harings L, Winkens B, Knottnerus A, van der Weijden T. Defining antimicrobial prescribing quality indicators: what is a new prescription? *Eur J Clin Pharmacol* 2011;67:91-96.
61. van Weel C, Schers H, Timmermans A. Health care in the Netherlands. *J Am Board Fam Med* 2012;25 Suppl 1:S12-17.

62. Valk GD, Renders CM, Kriegsman DM, Newton KM, Twisk JW, van Eijk JT, van der Wal G, Wagner EH. Quality of care for patients with type 2 diabetes mellitus in the Netherlands and the United States: a comparison of two quality improvement programs. *Health Serv Res* 2004;39:709-725.
63. Guzek J, Guzek S, Murphy K, Gallacher P, Lesneski C. Improving diabetes care using a multitiered quality improvement model. *Am J Med Qual* 2009;24:505-511.
64. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, Halperin I, Vachon B, Ramsay T, Manns B, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet* 2012;379:2252-2261.
65. Bird GC, Marian K, Bagley B. Effect of a performance improvement CME activity on management of patients with diabetes. *J Contin Educ Health Prof* 2013;33:155-163.
66. Pronovost PJ, Hudson DW. Improving healthcare quality through organisational peer-to-peer assessment: lessons from the nuclear power industry. *BMJ Qual Saf* 2012;21:872-875.
67. Hudson DW, Holzmueller CG, Pronovost PJ, Gianci SJ, Pate ZT, Wahr J, Heitmiller ES, Thompson DA, Martinez EA, Marsteller JA, et al. Toward improving patient safety through voluntary peer-to-peer assessment. *Am J Med Qual* 2012;27:201-209.
68. Grol R. Kwaliteitsbewaking in de huisartsgeneeskunde. Nijmegen: Radboud University; 1987.



Summary

Summary

This thesis describes the research that was put into a complex intervention aimed at general practitioners with the aim to improve their test ordering behavior and prescribing behavior. Interventions of this type were performed before but never on test ordering and prescribing together. Neither was this type of intervention ever tested under normal practice conditions with much freedom for the participants to adapt the intervention to their own needs and situations. In this chapter the separate sections of this thesis are summarized for a quick overview of the findings.

Chapter 1 Is the general introduction to this thesis. The background of the research is reported together with the questions this thesis addresses.

Several problems are present in healthcare today such as the ever growing costs of healthcare that policymakers need to control. Also physicians can't keep up with all the evidence that is being published. In the Netherlands therefore guidelines are being developed since the 1980's as a means to summarize evidence and translate it to the field. Researchers also started off to research how to effectively translate evidence to field workers. The most promising tools for this seem to be interventions that are multifaceted and contain education, audit and feedback, small peer group work and the use of opinion leaders. However, most of this research was conducted under well controlled circumstances which diminishes the external validity of these findings. Usually these interventions also dealt with only one or two well described clinical topics. As guidelines became available to general practitioners (GPs) they started to discuss prescribing behavior in small peer review groups together with their affiliated local pharmacist, so called pharmacotherapeutic audit groups (PTAM).

In an earlier trial small peer group groups were used for discussing test ordering behavior of GPs. Experiences from this intervention and the availability of an existing infrastructure of PTAM groups led us to designing a complex intervention to support GPs in following guidelines on test ordering and prescribing. The aim of this intervention was to transfer the strategy to the field.

Qualitative research was used to evaluate the transference of the intervention to the field. A specific design was used to research the effect on test-ordering rates and prescribing rates of this intervention. This design is known as a balanced incomplete block design in implementation science.

We searched for answers to the following questions in this thesis on changing test ordering behaviour and prescribing behaviour of GPs:

- Was the intervention executed as planned?
- What were the barriers to and facilitators of the implementation of the strategy?

- Has the level of group performance been improved in the participating groups?
- Did the volumes of tests ordered and drugs prescribed change in the preferred direction, as described in the working agreements of the local quality improvement collaboratives (LQICs), compared to baseline?
- What is the effect of this strategy on GPs' test ordering and prescribing behaviour in terms of interphysician variation and total volumes of tests and prescriptions with respect to specific clinical topics, compared to that among GPs exposed to the same strategy but for other topics?
- Is the balanced incomplete block design suitable for use in implementation science?

Chapter 2 is the research protocol for the intervention. In this chapter earlier work of our research group on test-ordering behavior is described. The results of the previous trials and the design of the present intervention are described in more detail. In the intervention comparative feedback on test-ordering and prescribing behavior of GPs was provided embedded in existing PTAM groups. In these PTAM groups the feedback will be discussed in a peer review process and working agreements will be set. The feedback will be offered during two consecutive sessions; the first on test ordering and the second on prescribing. The feedback offers standardized data on practice level together with data at group level and of the surrounding groups. Hereby these PTAMs will change their names to local quality improvement collaboratives.

The methodology of the study is described in detail. It is a cluster randomized controlled trial (C-RCT) with two arms where both arms receive the same intervention but on different clinical topics. Both arms serve as blind controls for the other research arm. The main outcome measures were the change in volumes of tests ordered and volumes of drugs prescribed per 1000 patients per six months. Additionally the working agreements set by the GPs were analyzed and related to changes in volumes that were observed.

Chapter 3 reports the results of a process evaluation of the quality improvement strategy consisting of comparative feedback and peer review embedded in already existing local quality improvement collaboratives (research questions 1 and 2). All elements of the intervention were of proven effectiveness and feasibility on changing physician behavior in previous trials. However, implementing our combined strategy covering both test ordering and prescribing performance proved to be problematic. The aim of this study was to describe and analyze the challenges of the transferring process. We interviewed 19 regional health officers, pharmacists, laboratory specialists and general practitioners within 6 months after the transfer period. The transcripts of these

interviews were independently coded. These codes were then matched to the dimensions of the normalization process model (NPM). This model can be used to structure factors that matter when normalizing an intervention. Normalizing can be interpreted here as the sustainable implementation of an intervention in a way that the groups will accept it and adopt it as normal.

The result of the evaluation showed that the general idea of the intervention was widely supported by GPs. The process of designing and maintaining the databases containing test ordering data and prescribing data needed to generate feedback as well as the generation of the feedback from these databases was more complex than expected. Also the participants did not assume responsibility for the work and the distribution of the resources that came with it as was intended by us. This resulted in a higher need for support from the research group than anticipated on. The conclusion is that although we used elements for our intervention that were proven to be effective and embedded the intervention in an existing infrastructure of PTAMs this proved to be insufficient to normalize the intervention. The general positive attitude towards our intervention did not change this.

Several elements came up that require special attention when designing such a complex intervention in the future. How to build and manage such large databases, the transference of the responsibility for tasks and the distribution of resources are the main elements that should be taken care of as early as possible when planning such a complex quality improvement strategy. Only copying the barriers and facilitators experienced in a preceding trial is not sufficient. However, despite such a preparation management decisions could change the healthcare field, resulting in the intervention being less fitted to the situation. We learned that multifaceted implementation strategies to change professional behavior seem attractive but they have a pitfall in their inherent complexity.

Chapter 4 describes the design that was used before in similar implementation research and the problems that cling to the use of its terminology in quality of care research. The balanced incomplete block (BIB) design has oftentimes been claimed to be used in designing and evaluating complex interventions in quality of care research. We have doubts about the appropriateness of using this design for this type of interventions. This chapter describes that various articles on health services research claim to have applied the BIB. After careful reading the design used seems to differ from the descriptions of the BIBD in original methodological literature on this design. The design used resembles more of a straight two armed trial or a classical Latin square design. The BIBD is a sophisticated design with the aim to reduce the need for resources in a surrounding where these resources are limited. We conclude that the

use of this design is not suited for evaluating complex interventions in research using humans. Luckily the statistical techniques used in the evaluation of the studies that claimed to have used the BIBD are appropriate. They were in fact appropriate for the evaluation of normal two armed experiments. Therefore the results that were presented are still valid. The use of the term BIBD in quality of care research seems inappropriate. Also more attention should be paid to proper referencing to the original methodological literature.

Chapter 5 describes the results of our cluster-RCT in existing local quality improvement collaboratives (LQICs) in general practice. Twenty-one LQIC groups, including 197 GPs working in 88 practices, entered the trial and were allocated to one of two trial arms. The intervention in both arms comprised of audit and feedback combined with peer review. Each arm was presented a different set of five clinical topics. The collaboratives chose three of the five clinical topics presented to them, according to their preference, and agreed on targets for change. The results of both a crude pre-post analysis of the differences in testing and prescribing rates at the LQIC level, and a per-protocol analysis for the same differences are presented. We did not demonstrate a difference in the mean changes in test ordering or prescribing volumes between intervention and control groups for the crude analysis. It was observed that the LQICs that deviated most from the mean volumes at baseline, changed most in the desired direction on both test ordering and prescribing performance. Therefore we conclude that the beneficial results from earlier work could not be confirmed by our attempt to implement the strategy in the field. It might be interesting to further research on the groups or practices that deviate the most from the mean at baseline. Targeting them for interventions on quality of care could potentially yield the largest effects. It remains unclear however whether this behavior is consistent behaviour for a single practice or group.

Chapter 6 addresses an overview of the results provided in this thesis and the discussion based on these results. We conclude that despite the vigorous design, the use of an existing backbone of LQICs and much room for local tailoring we did not succeed in effectively changing behavior on test ordering or prescribing by GPs. The uptake of the strategy by the field was also not as expected. It showed that the burden of running the intervention and the administration that came was relatively high. The administrative burden perceived by the Dutch GPs was already high due to a large healthcare reform. Would the strategy be implemented in all LQIC groups nationwide the costs of running the intervention would be very high and probably exceed the savings. Also the barriers and facilitators that we expected and anticipated on proved

to be incorrect. As this intervention is one of the first where positive findings from earlier well controlled interventions are tested under normal more or less uncontrolled circumstances some questions remain and arise from this thesis. Further research is needed to test whether the results from implementation research in well controlled settings can be replicated when tested under usual circumstances. The effects that we observed in the practices deviating most from the mean at baseline are interesting. This raises the question whether these GPs should better targeted instead of the whole field. The cut-off point for the definition of deviation and whether this behavior is consistent within a practice requires specific research.



Samenvatting

Samenvatting

Dit proefschrift beschrijft de resultaten van een complexe implementatiestrategie die als doel heeft om het diagnostisch aanvraag- en voorschrijfgedrag van huisartsen te verbeteren. Interventies van dit type die eerder zijn onderzocht waren gericht op het veranderen van diagnostisch aanvraag- of voorschrijfgedrag van artsen, maar nooit op beide gedragingen tegelijk. Ook is dit type strategie nooit eerder getest in de dagelijkse zorg waarbij veel vrijheid werd gegeven aan de deelnemers om de strategie aan te passen aan hun eigen behoeften en omstandigheden. In dit hoofdstuk worden de verschillende onderdelen van dit proefschrift samengevat ten behoeve van een snel overzicht van de bevindingen van ons onderzoek.

Hoofdstuk 1 is de algemene inleiding van dit proefschrift en beschrijft de achtergrond van het onderzoek.

De hedendaagse gezondheidszorg kent verschillende problemen, zoals onder andere de toenemende kosten van de gezondheidszorg die de beleidsmakers proberen te beheersen. Daarnaast kunnen artsen nieuwe kennis en inzichten die worden gepubliceerd nauwelijks toepassen door de grote hoeveelheid ervan. In Nederland zijn daarom sinds 1980 richtlijnen ontwikkeld als een middel om bevindingen uit de wetenschap samen te vatten en te vertalen naar het veld. Onderzoekers zochten naar effectieve manieren om wetenschappelijke bevindingen over te brengen aan de artsen. Meest veelbelovend lijken gecombineerde interventies bestaande uit educatie, audit en feedback, intercollegiale toetsing in kleine groepen en het gebruik van rolmodellen. Het grootste deel van dergelijk onderzoek werd echter uitgevoerd onder goed gecontroleerde omstandigheden waardoor de externe validiteit van deze bevindingen relatief laag is. Ook behandelden deze interventies meestal slechts één of twee nauw afgebakende klinische onderwerpen. Tegelijkertijd met de ontwikkeling van richtlijnen begonnen huisartsen, samen met hun lokale apotheker, met het bespreken van hun voorschrijfgedrag in kleine groepen, zogenaamde farmacotherapeutische overleggen (FTO's). Uitgaande van de bestaande infrastructuur van FTO's en eerdere ervaringen met intercollegiale toetsing over diagnostiek, het zogenaamde Diagnostisch Toets Overleg (DTOs), hebben we een complexe implementatiestrategie ontworpen om huisartsen in het veld te ondersteunen in het navolgen van richtlijn-aanbevelingen over aanvraag- en voorschrijfgedrag. Hierbij hadden we de ambitie om de strategie over te dragen aan het veld zelf. De haalbaarheid van het uitvoeren van de strategie door het veld zelf hebben we geëvalueerd door middel van kwalitatief onderzoek. Het effect op het aanvraag- en voorschrijfgedrag hebben we geëvalueerd door middel van een

specifiek design, in de implementatie-literatuur bekend als het zogenaamde balanced incomplete block design (BIBD).

De vragen die we willen beantwoorden in dit proefschrift over diagnostisch aanvraag- en voorschrijfgedrag van huisartsen zijn:

1. Werd de implementatiestrategie uitgevoerd zoals gepland?
2. Wat waren de belemmeringen en de bevorderende factoren in de uitvoering van de strategie?
3. Is het niveau van presteren van de deelnemende FTO-DTO groepen verbeterd? Zijn de volumina van de aangevraagde tests en voorgeschreven medicijnen veranderd in de gewenste richting, zoals beschreven in de werkafspraken van de groepen zelf, dit in vergelijking tot de Ausgangssituatie?
4. Wat is het effect van deze strategie op het diagnostisch aanvraag- en voorschrijfgedrag van huisartsen in termen van interdokter variatie en het totale volume van diagnostische aanvragen en voorschriften van medicatie met betrekking tot specifieke klinische onderwerpen? Dit in vergelijking met huisartsen die dezelfde strategie aangeboden kregen maar over andere onderwerpen.
5. Is het BIBD geschikt voor gebruik in onderzoek naar implementatiestrategieën?

Hoofdstuk 2 beschrijft het onderzoeksprotocol voor de implementatiestrategie. In dit hoofdstuk wordt het eerdere werk van onze onderzoeksgroep ten aanzien van FTO en DTO beschreven. De resultaten uit deze eerdere studie en het ontwerp van de huidige interventie worden verder uitgewerkt. De implementatiestrategie bestaat uit het verstrekken van feedback met vergelijkende spiegelinformatie over aanvraag- en voorschrijf volumina van de participerende huisartsen. De spiegelinformatie wordt vervolgens besproken in het bestaande FTO (intercollegiale toetsing), waarin afspraken voor toekomstig beleid worden gemaakt. De spiegelinformatie wordt in twee sessies aangeboden over hetzelfde onderwerp. De eerste sessie gaat over het diagnostisch aanvraag gedrag en de tweede sessie gaat over het voorschrijven van medicatie. De spiegelinformatie is samengevat op praktijkniveau en wordt vergeleken met de cijfers van de groep en de omliggende huisartsgroepen.

Daarnaast wordt de methodologie van de studie beschreven. Deze studie betreft een cluster gerandomiseerd gecontroleerd experiment met twee onderzoekarmen waarbij alle huisarts-groepen in beide armen dezelfde verbeter-interventie zoals zojuist beschreven aangeboden kregen, maar per onderzoekarm over verschillende klinische onderwerpen. Beide armen dienden als geblindeerde controles voor de andere onderzoekarm. De belangrijkste uitkomstmaten waren de verandering in de volumina van aangevraagde tests en voorgeschreven geneesmiddelen per 1000 patiënten per half jaar. Daarnaast werden de werkafspraken die de huisartsen maakten geanalyseerd

en gerelateerd aan de veranderingen in de gestandaardiseerde volumina die werden waargenomen.

In **Hoofdstuk 3** worden de resultaten beschreven van de procesevaluatie van de implementatiestrategie, bestaande uit vergelijkende feedback en intercollegiale toetsing ingebed in de bestaande FTO groepen (onderzoeksvragen 1 en 2). Alle elementen van de strategie zijn in eerdere onderzoeken effectief en haalbaar gebleken in het veranderen van het gedrag van artsen. Onze uitvoering van de gecombineerde strategie die zowel diagnostisch aanvraag gedrag als het voorschrijfgedrag van huisartsen probeerde te veranderen bleek echter problematisch. Het doel van dit onderdeel van de studie was om de problemen in het proces van implementatie te beschrijven en te analyseren. We interviewden 19 regionale medisch managers in de eerste lijn, apothekers, laboratorium-specialisten en huisartsen binnen 6 maanden na afloop van de interventie. De transcripties van deze interviews werden onafhankelijk gecodeerd. Deze codes werden daarna ingepast in de dimensies van het normalization process model (NPM). Dit model kan worden gebruikt om factoren die van belang zijn om te komen tot normalisatie van een interventie te structureren. Normaliseren staat hierbij voor het effectief en langdurig implementeren van een strategie, zodanig dat deze door de groep als standaard (normaal) uitgevoerd wordt.

Uit het resultaat van de evaluatie bleek dat huisartsen de idee achter de interventie op grote schaal ondersteunden. Het opzetten en onderhouden van de voor de feedback benodigde databases met aanvraag- en voorschrijfgegevens en het genereren van de feedback hieruit bleek complexer dan vooraf verwacht. Daarnaast bleek dat de deelnemers niet de verantwoordelijkheid namen voor het werk en de verdeling van de middelen, zoals door ons bedoeld. Hierdoor was er een grotere behoefte aan ondersteuning door de onderzoeksgroep dan verwacht en bedoeld vooraf. Ondanks dat de elementen waaruit onze strategie opgebouwd was in eerder onderzoek effectief bleken en de strategie uitgevoerd werd in de bestaande infrastructuur van FTO groepen bleek dit onvoldoende om de interventie normaliseren. De algemene positieve houding ten opzichte van onze interventie heeft dit helaas niet kunnen veranderen.

Uit de resultaten van deze procesevaluatie komen verschillende punten naar voren die speciale aandacht nodig hebben bij het ontwerpen van dergelijke complexe implementatiestrategieën in de toekomst. Ten eerste dient vooraf veel aandacht gegeven te worden aan het opzetten en beheren van grote databases voor feedback. Ten tweede moet de overdracht van de verantwoordelijkheid voor taken en de verdeling van de middelen zo vroeg mogelijk worden overdacht bij het plannen van een dergelijke complexe strategie. Alleen het overnemen van de weerstanden en bevorderende factoren, zoals ervaren in eerdere onderzoeken, blijkt onvoldoende. Een

dergelijke voorbereiding kan echter niet voorkomen dat door onverwachte veranderingen in het management van de gezondheidszorg de strategie minder goed past dan voorzien. Wij hebben geleerd dat complexe strategieën om professioneel gedrag te veranderen aantrekkelijk lijken, maar dat deze complexiteit ook een valkuil is.

Hoofdstuk 4 beschrijft de onderzoeksopzet en de bezwaren die kleven aan de terminologie die ervoor wordt gehanteerd in de wetenschappelijke literatuur. Het balanced incomplete block design (BIBD) wordt vaak beweerd gebruikt te zijn bij het ontwerpen en evalueren van complexe interventies in kwaliteit van zorg onderzoek. Wij hebben twijfels over de geschiktheid van het gebruik van dit design voor dergelijke interventies. Dit hoofdstuk beschrijft dat verschillende artikelen over gezondheidszorgonderzoek beweren het BIBD gebruikt te hebben. Bij nalezen lijkt de gebruikte onderzoeksopzet te verschillen van de omschrijving van het BIBD in de originele methodologische literatuur over dit ontwerp. De gebruikte onderzoeksopzet lijkt eerder een normaal twee armig experiment of eventueel een Latin square opzet. Het BIBD is een geavanceerd en complex design met als doel om onderzoek mogelijk te maken in een omgeving waarin beperkte middelen beschikbaar zijn. We concluderen dat het gebruik van dit design niet geschikt is voor het evalueren van complexe interventies in mens gebonden onderzoek. Gelukkig zijn de statistische technieken die gebruikt zijn bij de evaluaties in de studies die beweerden het BIBD toegepast te hebben wel degelijk geschikt. Het betrof in feite statistische technieken voor de evaluatie van normale tweearmige experimenten. Derhalve zijn de resultaten die gepresenteerd zijn in deze onderzoeken nog steeds geldig. Het gebruik van de term BIBD in onderzoek gericht op verbetering van de kwaliteit van zorg lijkt oneigenlijk. Ook zou meer aandacht moeten worden besteed aan juiste verwijzingen naar originele methodologische literatuur.

Hoofdstuk 5 beschrijft de resultaten van ons cluster-experiment in bestaande FTO groepen in de huisartsgeneeskunde in Nederland. Eenentwintig FTO groepen, bestaande uit 197 huisartsen werkzaam in 88 praktijken, namen deel en zijn aan één van de twee onderzoeksarmen toegewezen. De implementatiestrategie in beide armen bestond uit audit en feedback in combinatie met intercollegiale toetsing. De groepen in beide armen kregen echter een andere set van vijf klinische onderwerpen aangeboden. Iedere FTO groep koos drie van de vijf klinische onderwerpen, op basis van hun eigen voorkeur, en stelden doelen voor verandering van beleid binnen hun eigen praktijken voor. De resultaten van zowel de intention-to-treat voor-na analyse van de verschillen in volumina van aangevraagde diagnostiek en de voorgeschreven medicatie op praktijkniveau, en een per-protocol analyse van dezelfde verschillen worden in dit

hoofdstuk gepresenteerd. We hebben gemiddeld gezien geen verschil gevonden in de verandering van volumina van diagnostisch aanvraag- en voorschrijfgedrag tussen interventie- en controlegroep in de intention-to-treat analyse. Wel zagen wij dat de huisartsgroepen die het meest afweken van de gemiddelde volumina bij aanvang van de studie, de grootste verandering lieten zien in de gewenste richting. Op basis van deze resultaten concluderen wij dat de positieve resultaten uit eerder werk niet konden worden bevestigd na onze poging om de strategie te implementeren in het bestaande netwerk van FTO groepen. Het zou interessant zijn om verder onderzoek te doen naar effecten van verbeterinterventies uitsluitend gericht op praktijken of FTO groepen die het meest van het gemiddelde afwijken bij een nulmeting. Het is echter vooralsnog onduidelijk of dergelijk afwijkend gedrag van een praktijk op een nulmeting voorspellend is voor afwijkend gedrag op andere tijdstippen en op andere domeinen.

Hoofdstuk 6 bevat de discussie van de studieresultaten die beschreven zijn in dit proefschrift. We concluderen dat ondanks de het gebruik van een gecombineerde strategie, het gebruik van het bestaande netwerk van FTO groepen en veel ruimte voor lokaal maatwerk, we er niet in geslaagd zijn om diagnostisch aanvraag- en voorschrijfgedrag door huisartsen effectief te veranderen. Ook verliep de incorporatie van de strategie in de bestaande FTO structuur door het veld niet zoals beoogd. De administratieve last van het uitvoeren van de interventie was relatief hoog, zeker gezien de administratieve last die veel huisartsen de laatste jaren ervaren door de veranderingen in de Nederlandse gezondheidszorg (transparantie, declaraties). Zou deze strategie landelijk in alle FTO groepen uitgerold worden dan zouden de kosten in tijd en geld voor het uitvoeren van de strategie niet opwegen tegen de besparingen.

De belangrijkste redenen voor het falen van onze strategie zijn een algemeen gevoel van overbelast zijn van huisartsen als gevolg van een grote gezondheidszorg hervorming die kort na het begin van de werving werd geïnitieerd. Daarbovenop bleken wij een onjuiste of op zijn minst achterhaalde visie te hebben op de te verwachten belemmerende en bevorderende factoren. Aangezien deze interventie een van de eersten is waar positieve bevindingen uit eerder goed gecontroleerde evaluatiestudies worden getest onder normale, min of meer ongecontroleerde omstandigheden blijven een aantal vragen onbeantwoord en worden enkele nieuwe vragen opgeworpen. Verder onderzoek is nodig om te testen of de resultaten van implementatieonderzoek dat uitgevoerd is onder goed gecontroleerde omstandigheden kan worden gerepliceerd indien het getest wordt onder normale omstandigheden. De effecten die we waarnamen in de praktijken die het meest afweken van de gemiddelde uitgangswaarden zijn interessant. Het roept de vraag op of toekomstig onderzoek zich beter op deze huisartsen zou kunnen richten in plaats van op het hele veld. De

afkapwaarde voor de definitie van “afwijkend” en de vraag of dit gedrag consistent is binnen een praktijk vereist verder onderzoek.

.



Valorisatie

Valorisatie

In dit proefschrift worden de resultaten weergegeven van een complex onderzoek naar het effect van een strategie met als doel om het diagnostisch aanvraag gedrag en het voorschrijfgedrag van huisartsen te verbeteren. De wetenschappelijke resultaten zijn richtinggevend voor de ontwikkelingen in de dagelijkse praktijk, met name voor beleidsmakers en huisartsen. Hier worden globaal de bevindingen uit het onderzoek beschreven in relatie tot de ontwikkelingen die momenteel gaande zijn. Aan de orde komen daarbij weerstanden de veranderingen die ingezet is van het farmacotherapie overleg (FTO) en het diagnostisch toets overleg (DTO), de obstakels hiervoor en de implicaties voor het kwaliteitswerk in kleine groepen door huisartsen. . Voor het slagen van deze veranderingen zijn randvoorwaarden aan te wijzen die voortkomen uit dit proefschrift. Deze randvoorwaarden, onder andere op het gebied van databeheer zullen als laatste benoemd en besproken worden.

Werken in kleine groepen

Ons onderzoek was gericht op werken in kleine groepen die een ruime eigen inbreng hadden in de te bespreken onderwerpen. Dat past bij de benadering van (huis)artsen vanuit beleidsmakers die momenteel aan het verschuiven is van een top-down benadering naar een bottom-up benadering. Dit is ingegeven doordat implementatie van richtlijnen beter wordt geacht te verlopen als dit bewerkstelligd wordt in kleine groepen in plaats van opgelegd van boven. Dit komt overeen met hetgeen beschreven is in de evidence scan van the health foundation.¹ De beroepsgroep lijkt voor te sorteren op het werken in kleine groepen waarin artsen in kleine groepen reflecteren op hun handelen en keuzes en hen hier zelf beleid over te laten formuleren op basis van spiegelinformatie. Dit past eveneens in de onlangs ingezette campagne “het roer moet om” waardoor nadrukkelijk meer aandacht komt voor behoud en bevordering van de autonomie van artsen zelf. Het lijkt logisch om te kiezen voor het al jaren goed functionerende systeem van FTO groepen als plaats waar deze nieuwe benadering als eerste ingevoerd kan gaan worden.

Het diagnostisch toets overleg

Het werk van Verstappen toont aan dat het DTO een effectieve manier is om overdiagnostiek door huisartsen tegen te gaan.^{2,3} Uit dit proefschrift blijkt echter dat deze positieve effecten niet reproduceerbaar zijn als de groepen de methode aangereikt krijgen en er zelf mee aan de slag moeten. De deelnemende huisartsen waren, zo blijkt uit de resultaten beschreven in hoofdstuk 3, enthousiast over de

algemene idee om te starten met een DTO. Ook was de algemene opinie over het werken met spiegelinformatie in een kleine groep positief. Desondanks is na de interventie beschreven in dit proefschrift het DTO in de meeste groepen niet verder voortgezet. De groepen gaven aan veel twijfel te hebben gehad aan de juistheid van de spiegelinformatie het opzetten en uitvoeren van een DTO werd als een zware belasting ervaren. Ook het ontbreken van een veldpartij die de leiding heeft genomen is een beperkende factor geweest voor het voortzetten van de strategie. Gezien de positieve resultaten van het onderzoek van Verstappen en de resultaten uit het hier beschreven onderzoek lijkt er voldoende basis om door te gaan met de invoering van een DTO naast het FTO. Voor het slagen hiervan is het nodig om enkele randvoorwaarden die naar voren zijn gekomen uit dit onderzoek goed in te vullen.

De financiering van deelname aan een DTO voor huisartsen door de zorgverzekeraars zou mogelijk een grote impuls kunnen zijn voor de invoering van het DTO. Daarnaast zijn laboratoria het DTO inmiddels gaan zien als een kans om hun relatie met de huisartsen in hun regio te verstevigen. Het Nederland Huisartsen genootschap (NHG) is daarnaast een richtlijn aan het ontwikkelen waarin aangegeven zal staan waaraan een goed DTO zal moeten voldoen.

Een centraal punt in de opzet van het DTO zal het gebruik van spiegelinformatie zijn. Deze spiegelinformatie zal in eerste instantie bestaan uit data over aanvragen gedaan bij het klinisch chemisch, hematologisch of microbiologisch laboratorium. Gegevens over beeldvormend onderzoek en functie-onderzoek zoals röntgenfoto's en endoscopieën bleken in dit onderzoek niet beschikbaar te maken in geaggregeerde vorm doordat ze niet in een uniform format opgeslagen worden. Dit zal in de nabije toekomst ook nog niet mogelijk zijn.

Het farmacotherapeutisch overleg

In Nederland bestaan al enkele decennia FTO groepen. Meerjarige financiering en ondersteuning van deze groepen hebben ertoe geleid dat ze zijn uitgegroeid tot een zelfstandig kwaliteitsinstrument. Nadat de financiering voor deelname aan een FTO is stop gezet zijn de meeste groepen toch doorgegaan met de periodieke besprekingen van farmacotherapie beleid in FTO's. In dit onderzoek hebben wij gebruik gemaakt van de bestaande structuur van FTO's en het vertrouwen dat er is tussen de leden onderling. Voor een verdere doorgroei van het FTO is echter een professionalisering nodig van de erkend kwaliteitsconsulenten (EKC).

Uit dit onderzoek bleek dat de toevoeging van spiegelinformatie aan het bestaande FTO door de huisartsen zeer positief ontvangen werd. De spiegelinformatie die gebruikt zal worden moet bestaan uit procesindicatoren en uitkomstindicatoren met als doel om de kwaliteit van de geboden zorg te verbeteren. Om de steun van de huisartsen te

behouden kunnen de gebruikte indicatoren hierbij enkel voor intern gebruik binnen de groepen beschikbaar worden gesteld.

Integratie van DTO en FTO

Ondanks de positieve attitude van de deelnemende huisartsen ten opzichte van het idee om te starten met een DTO en een FTO met gebruikmaking van spiegelinformatie, bleek het moeilijk om de groep verantwoordelijk te maken voor het proces van implementatie van die strategie in de eigen groep. Er was duidelijk behoefte aan een leider die het initiatief nam. De leiders die door ons getraind waren bleken onvoldoende voorbereid op hun taak. Door langduriger begeleiding en herhaalde trainingen was dit te ondervangen geweest.

Inhoudelijk gezien is er geen twijfel dat bij het streven naar werken in kleine groepen het een logische keuze is om het DTO te koppelen aan een FTO over hetzelfde onderwerp. Het model dat gebruikt is in onze interventie lijkt zeer geschikt om te gebruiken als een integratie nagestreefd zou worden. Hierbij hebben de groepen eerst een DTO over een bepaald klinisch onderwerp en op een later moment een FTO over hetzelfde onderwerp. De verwachting van beleidsmakers is echter dat er voorlopig geen integratie zal komen van DTO met FTO. De implementatie van het DTO zal een geleidelijk proces zijn waarvoor in 2016 het startsein gegeven lijkt met onder andere de introductie van financiering hiervoor. Vanwege de complexiteit van een integraal DTO-FTO zoals beschreven is in dit proefschrift zullen veel groepen eerst het DTO goed ingevoerd moeten hebben voordat zij de volgende stap richting integratie kunnen zetten.

Databases

Uit ons onderzoek bleek duidelijk dat er grote weerstand bestond tegen het werken met cijfers. Huisartsen gaven vaak aan geen vertrouwen te hebben in de cijfers en voerden onder andere aan dat de data voor zowel DTO als FTO incompleet waren. De data gebruikt voor het FTO werd daarnaast met argusogen gezien omdat de bron niet vertrouwd werd. Dit bleek een grote handicap bij de uitvoering van de interventie. In dit onderzoek is gebruik gemaakt van twee apart gebouwde databases met FTO data, afkomstig van de verzekeraars, en DTO data, afkomstig van de laboratoria. Hierdoor konden wij problemen als onvolledige datasets voorkomen. Het bleek echter aanvankelijk een enorme inspanning te kosten om de databases gevoed te laten worden met nieuwe data. Het beste zou zijn om te komen tot twee landelijke databases met DTO en FTO gegevens. De betrouwbaarheid en uniformiteit van de data zal hierdoor toenemen. Momenteel is er nog geen partij die in staat is om een

dergelijke dataset te beheren met voldoende mandaat van alle huisartsen. Er zijn enkele onderzoeksinstituten met geschikte databases. Verstrekken van spiegelinformatie voor DTO en FTO door hen zal momenteel echter op juridische en praktische bezwaren stuiten. Ook moet de bron waar de data vandaan komt zorgvuldig gekozen worden. Het lijkt logisch om de data direct bij de huisarts uit het huisartseninformatie systeem (HIS) te extraheren. Ook hiervoor geldt dat hiervoor een aantal praktische en juridische zaken vooraf opgelost moeten worden. Zo moet rekening gehouden worden met de privacy en het beroepsgeheim en dient er een methode ontwikkeld te worden om eenvoudig, zonder veel administratieve last voor de huisartsen, data uit het HIS te extraheren.

Leiderschap tonen

Uit ons onderzoek blijkt dat voor het slagen van de implementatie van DTO en het revitaliseren van het FTO het hebben van een leider in de regio en in iedere groep een voorwaarde is. De regionale leiders voelden zich onvoldoende voorbereid en de leiders in de groepen namen onvoldoende het voortouw bij een overwegend afwachtende houding van de groep. Zowel de leider in de groep als in de regio kan optreden als kartrekker. Beiden hebben echter een andere rol; de leider in de groep moet de groep enthousiast maken en houden, de leider in de regio moet de gehele regio aanjagen en aansturen. Vanwege de complexiteit van de verandering die nodig is zullen de leiders in de groep nadrukkelijk aandacht moeten schenken aan de verwachtingen, weerstanden en twijfels die er leven in de groep. Pas als dit in voldoende mate besproken en opgelost is kan er ruimte zijn voor het gebruik van spiegel informatie. De leiders in de groep zullen de vragen over de betrouwbaarheid van de spiegelinformatie adequaat moeten kunnen beantwoorden. De leiders in de regio moeten gaan fungeren als aanjager en facilitator van de groepen en de leiders in de groepen. Zij zullen groepen actief moeten begeleiden bij de eerste stappen op weg naar een DTO en de aanlevering van de juiste data. Het lijkt logisch dat de Erkend Kwaliteits Consulente (EKC) deze taak op zich neemt, samen met de lokale klinisch chemicus. Echter op dit moment is de EKC hiervoor niet voldoende geëquipeerd en is de financiering van de EKC nog niet afdoende formeel geregeld. Voorafgaand aan het uitrollen van het DTO in de bestaande FTO groepen zal de rolverdeling van iedereen helder moeten zijn.

Implementatie, borging en follow-up

De implementatie van het DTO, stapsgewijs en gekoppeld aan het FTO, zal tijd kosten. De groepen die aan dit onderzoek hebben meegewerkt kunnen waarschijnlijk beschouwd worden als voorlopers. Het lijkt realistisch om te anticiperen op een

meerjarig traject, zoals bij de introductie van het FTO, waarbij steeds meer groepen zullen starten met een DTO. Belangrijk hierbij is dat de randvoorwaarden vooraf duidelijk zijn en waar nodig ingevuld. Ook zal het waarschijnlijk zo zijn dat als de laatste groepen starten met het DTO, de eerste groepen al toe zijn aan een integratie van hun DTO en FTO. De uiteindelijke introductie en borging van het DTO, en later een DTO-FTO, hangt ook in belangrijke mate af van andere ontwikkelingen in het veld. De invoering van populatiebekostiging, de discussie over het nut van het gebruik van indicatoren en maatregelen als stopregels of spertijden zijn allen gericht op de kwaliteit van zorg en beheersing van de kosten, net als de invoering van het DTO(-FTO). Na de start van de implementatie van het DTO is het belangrijk om te blijven meten, op uitkomst indicatoren en op proces indicatoren. Er zal een veldpartij verantwoordelijk moeten zijn voor de begeleiding van de invoering van het DTO en de latere integratie in het FTO. Samen met de partij die de databases beheert en spiegelinformatie verstrekt zullen zij het mandaat van de huisartsen moeten krijgen om deze taak naar behoren uit te voeren.

Het huidige systeem van herregistratie zal aangepast moeten worden waarbij punten door het werken met spiegelinformatie behaald moeten worden.

Samenvatting

Dit proefschrift heeft laten zien dat implementatie van een DTO-FTO met spiegelinformatie mogelijk is. De effecten zijn echter maar klein en de inspanningen om te komen tot een effectief overleg zijn groot. In potentie blijft er sprake van een zeer goede methode om richtlijnen te implementeren in het veld. Hiervoor zijn wel enkele randvoorwaarden aanwijsbaar die ingevuld moeten zijn voor effectieve implementatie en borging:

1. Een gemeenschappelijk door huisartsen opgesteld en gedragen visie ten aanzien van kwaliteitsbeleid en de plaats van DTO-FTO in kleine groepen daarin.
2. Berouwbaar data moet voorhanden zijn (extractie uit HISsen).
3. Financiering en ondersteuning moeten meerjarig gegarandeerd worden.
4. EKC moeten een forse impuls krijgen om tot leiderschap te komen.
5. Regionale partijen zullen leiding moeten nemen in het ondersteunen van de groepen om data te verkrijgen ten behoeve van feedback.

Huisartsen zullen de winst van het nieuwe beleid moeten ervaren.

References

1. de Silva D, Bamber J. improving quality in general practice. In., vol. 23: The Health Foundation; 2014.
2. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermesen J, Grimshaw J, Grol RPTM. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians: A Randomized Trial. *JAMA* 2003;289(18):2407-2412.
3. Verstappen WHJM, van der Weijden T, Dubois WI, Smeele I, Hermesen J, Tan FES, Grol RPTM. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med* 2004, 2(6):569-575.



Dankwoord

Dankwoord

Schrijven is een vak, dat is mij duidelijk geworden. Ik ben blij dat het eindelijk zo ver is gekomen dat ik mijn dankwoord op papier kan zetten. Tijdens mijn opleiding tot huisarts besloot ik dat ik onderzoek wilde gaan doen. Via Bas Maiburg en Geert-Jan Dinant kwam ik bij Trudy van der Weijden terecht. Dit was het startpunt voor een ingewikkeld traject van huisarts (in opleiding) zijn én promovendus. Dit traject had ik nooit succesvol kunnen afronden zonder de hulp van velen. Sommigen hebben mij actief geholpen, anderen kritische geluiden laten horen op de momenten dat dat nodig was en weer anderen juist de ruimte geboden om dit werk af te maken. De lijst van mensen die ik moet bedanken is lang, zo lang dat ik niet iedereen bij naam kan noemen. Enkelen verdienen het echter uitdrukkelijk genoemd te worden.

Mijn promotores Richard Grol, Job Metsemakers en Trudy van de Weijden: Het was voor mij een hele prettige samenwerking gedurende dit lange traject.

Trudy, Ik ben jou heel veel dank verschuldigd voor jouw niet aflatende geduld. Ik heb veel van jou geleerd, jij hebt mij gestimuleerd om steeds opnieuw door te gaan en bood mij kansen om mezelf te verdiepen. De mogelijkheid om al vroeg in mijn onderzoeksloopbaan naar Canada te gaan voor een masterclass bij David Sackett was hiervoor exemplarisch. Af en toe zal je moedeloos worden zijn van de trage vooruitgang en alle tegenslagen die we in dit onderzoek tegenkwamen. Als ik echter gefrustreerd over een probleem jouw kamer in liep wist jij steeds weer van een probleem een uitdaging te maken. Zonder jouw voortdurende motivatie had ik wellicht niet doorgezet. Als ik je een dubbeltje moest geven voor iedere "heks" die jij vond, was ik nu straatarm.

Richard, je hebt me telkens verstoeld doen staan met jouw scherpe analyses. Je haalde iedere tegenstrijdigheid of onduidelijkheid uit mijn artikelen, stimuleerde om verder te denken en kwam met oplossingen voor problemen. Pas later in de trein terug naar huis bedacht ik me hoeveel werk jouw opmerkingen in de kantlijn eigenlijk inhielden. Echter je had het steeds bij het juiste eind en mijn werk is er beter van geworden. Bedankt dat ik gebruik heb mogen maken van jouw kennis en kunde op het gebied van implementatie en schrijven.

Job, vanaf het begin ben je betrokken geweest bij dit project, de ene keer meer op de achtergrond en dan weer meer direct betrokken. Het was prettig om jou erbij te hebben omdat je naast jouw ervaring als onderzoeker ook ervaring uit het veld kon inbrengen. Bedankt voor de momenten dat ik bij je binnen kon lopen en sparren over mijn onderzoek.

Ik had geluk met mijn begeleidingsteam bestaande uit Wim Verstappen, Rob Janknegt, Paul Muijrsers en Ron Winkens. Zij brachten een goede mix van praktijk ervaring en theoretische kennis in. Wim, het eindpunt van jouw onderzoek was het vertrekpunt van mijn onderzoek. Ik heb veel en dankbaar gebruik gemaakt van jouw achtergrond op het gebied van de interventie en het veld tijdens de opzet van mijn onderzoek. Rob, jouw kennis als apotheker en begeleider van FTO-groepen bracht veel praktische kennis mee. Dank voor je betrokkenheid en enthousiasme bij het op de rails krijgen van dit onderzoek. Paul, jouw contacten met de juiste mensen bij CZ hielpen enorm. Door jouw rust en relativiseringsvermogen was het altijd prettig samen te werken. Ron, de energie en het enthousiasme die jij hebt zijn bewonderenswaardig. In jouw drukke agenda als praktijkhoudend huisarts en actief onderzoeker heb je daarnaast tijd gemaakt voor het opzetten van dit onderzoek, dit waardeer ik enorm.

Ben van Steenkiste, je bent in een later stadium betrokken geraakt bij mijn onderzoek en op het moment dat ik het het hardst nodig had. Jouw pragmatische blik, relativiserings-vermogen en aanstekende lach en hielpen me vaak weer vooruit.

Bjorn Winkens, statistiek blijft een ingewikkelde materie voor doeners als ik. De lastigste problemen echter, tekende jij uit zodat zelfs ik het snapte....tot je weer de deur uit was. Gelukkig was het voor jou geen enkel probleem als ik dan later opnieuw kwam checken of ik het nu toch echt goed begrepen had.

Alfons Schroten, veel tijd hebben we gestoken in het zoeken naar het juiste format voor de databases. Het steeds weer op tijd binnenkrijgen van nieuwe data bleef een uitdaging. Samen met Diane van de Vorstenbosch van Meetpunt Kwaliteit waren jullie verantwoordelijk voor het op tijd afleveren van de juiste spiegelinformatie bij de groepen. Diane en Alfons bedankt voor jullie inzet en enorme flexibiliteit.

Naast de spiegelinformatie kregen de groepen ook samenvattingsteksten van de richtlijnen. Deze zijn beoordeeld door veel, meestal toch al druk bezette, artsen. Voor zover jullie niet al elders genoemd zijn: dank voor het lezen en corrigeren van de teksten.

De leden van de beoordelingscommissie, Prof dr. Geert-Jan Dinant, Prof. dr. Frank Buntinx, dr. Loes van Bokhoven, Prof. dr. Michel Wensing en Prof. dr. Jako Burgers wil ik graag bedanken voor hun bereidheid mijn proefschrift te lezen en te beoordelen.

Arnoud Frericks, gedurende mijn gehele onderzoek hebben we op gezette tijden contact gehad. Het was altijd prettig om met jou van gedachten te wisselen. Jouw

kennis als DTO-coördinator in Etten-Leur was zeer welkom, bedankt voor jouw inbreng en de fijne samenwerking.

Het IVM wil ik bedanken voor het gebruik dat wij mochten maken van hun meetinstrument en voor het indelen van de groepen naar niveau.

Zonder “mijn” onderzoeks-assistentes was ik nog veel langer bezig geweest met regelen en analyseren. Anuska Muyres maakte in de eerste jaren deel uit van het team, regelde veel en nam mij daardoor werk uit handen. Na haar vertrek nam Paddy Hinssen de ondersteuning van mijn onderzoek op zich. Paddy, je hebt slechte nachten gehad en je bent wanhopig geworden van de analyses...., ik weet het. Jouw niet aflatende volharding en optimisme zijn bewonderenswaardig. Ik ben je eendeloos dankbaar voor al het werk dat je voor me hebt verricht.

Mijn “roommates” Jolien, Anemieke en Marika: bijna tegelijk begonnen we bij HAG aan onze promotieonderzoeken. Alle vier met een volledig andere achtergrond en een heel ander onderzoek. Het werken met jullie samen op één kamer maakte mijn werk op de universiteit altijd tot een welkome afwisseling met de dagelijkse praktijk. Bedankt voor alle gezellige momenten en de lol die we samen gehad hebben.

Ine Siegelaer dank voor al jouw praktische hulp. Ik ga jouw grote gevoel voor humor missen.

Tiny Wouters hartelijk bedankt voor het verzorgen van de layout van mijn proefschrift. Dit boekje was anders echt bij een stapel artikelen met een nietje erdoor gebleven.

Zonder Paul Hulshof was het moeilijk geweest om mijn huisarts opleiding af te ronden. Bedankt dat je me de kans hebt gegeven om bij jou het laatste deel van mijn opleiding af te ronden. Ik heb veel van je geleerd, daar pluk ik nog iedere dag de vruchten van.

De huisartsen, apothekers, klinische chemici en medisch coördinatoren die hebben deelgenomen aan dit onderzoek: jullie zijn, met ruim 250, teveel om allemaal persoonlijk genoemd te worden. Ik had het echter niet kunnen doen zonder jullie belangeloze inzet. De resultaten die dankzij jullie behaald zijn kunnen nu gebruikt worden om het DTO-FTO beter in te voeren.

Mijn collega praktijkhouders Piet De Bruyckere en Sandra de Vree, dank voor jullie interesse en de ruimte en tijd die jullie me gaven om mijn promotie af te maken. We vullen elkaar goed aan en ik hoop dat we nog lang samen mogen werken.

Michel en Martijn, natuurlijk zijn jullie mijn paranimfen! Bedankt voor jullie vragen, nieuwsgierigheid en interesse in mijn onderzoek. Voor alle familie, vrienden, burens en kennissen die, soms met enige aarzeling of het wel het juiste moment was, interesse toonden; Dank jullie wel. Fijn dat jullie er waren en zijn.

Diana, je bent mijn toetssteen en rots in de branding. We hebben hele drukke jaren achter de rug waarin we veel ballen tegelijk in de lucht hebben gehouden. Je hebt me ongelooflijk veel ruimte gegeven en gestimuleerd om mijn ambities na te jagen. Ook deze bal is in de lucht gebleven en kan nu neergelegd worden. Samen zijn we een fantastisch team, met een even zo fantastische toekomst. Dank dat je er altijd bent, dat je mijn vrouw bent en voor al jouw steun.

Mees, Stijn en Teun bedankt voor zowel jullie geduld als ongeduld. Dat laatste motiveerde weer om dit boekje te ronden. Vanaf nu is papa weer vaker thuis en doe ik weer mee. Mijn upgrade is voltooid.



Curriculum vitae

Curriculum vitae

Jasper was born in Nijmegen on 4 January 4 1975. He was raised in Hilversum and Bilthoven where he finished secondary school in 1994. The same year he moved to Maastricht to start studying health sciences. In 1995 he started to study medicine which he finished in 2002.

After graduating med school he started working at the Maasland hospital in Sittard as a resident on the gynecology & obstetrics department. In 2004 he started at the Wilhelmina childrens hospital / University clinics in Utrecht to work as resident in the gynecology & obstetrics department. Later that year he moved back to Sittard to fill in his old position as resident. In 2005 he switched career to the broadly oriented general practice. In 2007 he finished his training and started working as general practitioner. In 2008 he obtained a position as GP in the healthcentre Terwinselen in Kerkrade. Since 2009 he is co-owner of the GP-praxis and provides care for the residents of Kerkrade. During his vocational training Jasper talked to Geert-Jan Dinant and Trudy van der Weijden about his ambition to start a PhD study. He started in May 2006 with his first orientation on the field of implementation science and improvement of patient care. His PhD thesis was accepted in May 2016.

Jasper is married to Diana and they have three sons Mees, Stijn and Teun. He lives in Maastricht and still works in Kerkrade.

