Maastricht University

Towards optimal test ordering in primary care

Citation for published version (APA):

Verstappen, W. H. J. M. (2004). *Towards optimal test ordering in primary care*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. https://doi.org/10.26481/dis.20040917wv

Document status and date:
Published: 01/01/2004

DOI:
10.26481/dis.20040917wv

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

Please check the document version of this publication:

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

Link to publication

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

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

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

www.umlib.nl/taverne-license

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 25 Apr. 2024

CHAPTER X Summary

CHAPTER I introduces the subject of this thesis, the test ordering behaviour of general practitioners (GPs). The numbers of tests ordered by GPs are increasing and many of these tests appear unnecessary according to established evidence-based guidelines. Furthermore, inter-doctor variation seems to be large. This thesis describes the variation in test ordering behaviour in primary care, provides a systematic literature review on the strategies used by others to influence physicians' test ordering behaviour, and discusses the effects and costs of an innovative strategy we developed to improve GPs' test ordering behaviour. The strategy was systematically developed on the basis of the findings of the literature review. The multifaceted strategy had an iterative character and included the following elements: personalised graphical feedback, guideline dissemination and continuous small group quality improvement meetings. An important feature of the graphical feedback reports was a comparison between the behaviour of individual GPs and that of their colleagues. Mutual feedback by working in pairs, discussing guidelines, and drawing up plans for change were important features of the small group quality improvement meetings.

The meetings were organised in local GP groups. Local GP groups are an existing part of the infrastructure of Dutch GPs working together in a specific region. One of their tasks is to organise care during out-of-office hours, while continuing medical education is another important activity in many of these local groups. Co-ordination of the feedback and supervision of the group meetings was provided by a diagnostic centre, a facility where GPs can order laboratory, imaging and function tests without referring patients for specialist care. One of the tasks of the medical coordinator of the diagnostic centre is to give feedback to GPs on their test ordering behaviour.

The multifaceted, innovative approach was implemented in five regions catered for by five diagnostic centres, all working with a problem-oriented test ordering form. It was evaluated in 40 local GP groups by means of a multicentre trial with randomisation at local GP group level.

CHAPTER II describes a survey study of the variation in the test ordering behaviour of the GPs that participated in the test ordering trial, which tried to establish professional-related and context-related determinants of GPs' inclination to order tests, by means of a cross-sectional analysis. The baseline data of the trial, which involved 19 laboratory and 8 imaging tests, combined in a sum score per GP per year, were analysed to assess determinants of inter-doctor variation. In a multivariable multilevel regression analysis, these data were linked with survey data on professional characteristics such as knowledge about and attitude towards test ordering, and with data on context-related factors such as practice type or experience with feedback on test ordering data. The response to the survey was 97 %. Test ordering data were available for 229 GPs in 40 local GP groups from five regions. We found that the total median number of tests per GP per year was 998 (interquartile range 663 to 1500), with large differences between the regions ($p<0.001$). Factors significantly associated with smaller number of tests ordered were, at professional level, 'individual involvement in developing guidelines' (yes versus no) and at context level 'group practice' (versus two-person or single-handed practices) and

'more than one year of experience working with a problem-oriented laboratory order form' (yes versus no). GPs who met these three criteria ordered 27%, 18%, and 41% fewer tests, respectively, than their colleagues. We concluded that, in addition to professional-oriented determinants, context-related factors are strongly associated with the numbers of tests ordered. Further studies on GPs' test ordering behaviour should include local and regional factors.

Chapter III reports on a systematic literature review of strategies to influence test ordering behaviour, applying rigorous Cochrane Collaboration methods. It was hypothesised that changing the absolute rate of test use (which in most cases meant reducing the general overuse of diagnostic tests) and improving the appropriateness of test use (usually by means of explicit guidelines for certain disease-defined patient categories) are different issues that need different strategies. The second hypothesis was, that multi-faceted strategies would generally have a greater impact than single strategies. Finally, it was hypothesised that studies evaluating strategies involving context-oriented interventions would have a greater impact than exclusively professional-oriented interventions. A total of 98 studies with 118 comparison groups were included. Overall results were heterogeneous, due to differences in the type or intensity of the intervention or the setting, or due to methodological differences between studies. Modifying the overuse of tests and improving the appropriateness of test ordering behaviour may require different strategies. In addition to professional-oriented interventions, it seems important to consider the use of interventions that focus on organisational factors. It is not clear whether single strategies have less impact than multifaceted strategies, but it seems important to focus the intervention on both the professional and the context. Audit and feedback seem effective in decreasing absolute test rates as well as in improving the appropriateness of test use. Reminders by computer aided decision support were found to improve the appropriateness of test use, while outreach visits, patient-mediated interventions and small group quality improvement deserve more attention.

Chapter IV evaluates the strategy of combining feedback, guideline dissemination and small group quality improvement on the basis of a multicentre randomised controlled trial with a balanced, incomplete block design. The primary outcome measure was the total number of tests ordered for three different clinical problems per GP per six months. Arm I consisted of 13 groups receiving the strategy on three clinical problems, viz., cardiovascular diseases, upper abdominal complaints and lower abdominal complaints, while arm II consisted of 14 groups that received the same strategy, but concentrating on three other clinical problems, viz., COPD / asthma, general malaise / vague complaints and degenerative joint complaints (see chapter I, figure 1). The ordering volume of all tests related to the six clinical problems was monitored in both arms. The GPs were blinded for the intervention on the three clinical problems included in the other arm. In agreement with existing national, evidence-based guidelines, decreases in the total numbers of tests ordered as well as in the numbers of tests ordered per clinical problem and for some specified inappropriate tests were regarded as quality improvements. Analysis of covariance showed that

in arm I, the decrease in the total numbers of tests relating to cardiovascular diseases, upper abdominal complaints, and lower abdominal complaints was greater than in arm II, the difference being 67 tests more per GP per six months (p =0.01). For the GPs in arm II the mean change in the numbers of tests for COPD / asthma, general malaise / vague complaints and degenerative joint complaints was greater than that in arm I, the difference being 28 tests (p=0.22). In both arms, there was a reduction in the ordering of specified inappropriate tests, although the reduction was not significant for the GPs in arm II. The new strategy, focusing on guidelines and interaction and feedback between GPs, thus seems an effective tool for improving GPs' diagnostic testing.

Chapter V assesses the added value of small peer group quality improvement meetings for improving test ordering behaviour compared to one of the elements of the strategy, viz., simple feedback, on its own. This research question was evaluated by comparing arm I (see chapter IV) with a third arm including GPs receiving feedback on the same three clinical problems as in arm I (cardiovascular diseases, upper abdominal complaints, and lower abdominal complaints). The complete strategy was applied in 13 GP groups with 85 GPs (arm I), while 14 GP groups with 109 GPs received feedback only (arm III) (see chapter I, figure 1). Analysis of covariance showed that in arm I the decrease in the mean total number of tests (51 fewer tests per GP per six months) was far more substantial than that in the feedback arm (p=0.0049). Five tests deemed 'inappropriate' for the clinical problem of ' upper abdominal complaints' showed a greater decrease in arm I than in the feedback arm, the difference being 13 tests per GP per six months (p=0.0015). Inter-doctor variation decreased more in arm I. This implies that if audit and feedback are to be effective, they need to be integrated in an interactive, educational environment.

Chapter VI provides a framework for cost evaluations of quality improvement strategies. Cost analyses were done on the trial arms reported in Chapter V, that is the complete strategy, involving feedback, guidelines, and small group quality improvement, versus feedback only. Regular costs, development costs, and scientific costs were determined. Costs per GP of the new strategy were € 92.70 per six months in the total strategy arm, € 17.10 per six months in the feedback arm. An analysis of covariance was performed with the mean costs per GP per six months after the intervention as the dependent variable, and the costs of tests at baseline and the district as independent variable. The total strategy arm achieved a mean cost reduction of € 301 per GP per six months (p=0.001), while the feedback only strategy GP saved € 161 per GP per six months. Within the proposed framework, it is imperative to account for non-monetary benefits. We concluded that our strategy is a useful quality instrument. In line with the cost analysis framework for this kind of behavioural intervention, it seems useful to implement this strategy on a larger scale

Chapter VII evaluates the use and applicability of a multifaceted strategy to improve GPs' test ordering behaviour by means of a prospective process evaluation during the first and second years of the trial. All 193 GPs participating in arm I and arm II received the planned six feedback reports. Data from 156 quality meetings of 26 local GP

groups showed a participation rate of 81% (95% CI: 77%-85%) in the first year and 73% (95% CI: 68%-77%) in the second year. The three key points of the quality meetings, viz., mutual feedback by working in pairs, drawing up individual plans for change and drawing up group plans for change, were performed satisfactorily in both intervention years. In the first year, GPs expressed their level of satisfaction in a score of 7.55 on a scale of 0 - 10 (95% CI 7.46-7.64); the average score in the second year was 7.51 (95% CI 7.30-7.74). We concluded that the strategy is a feasible tool for continuing improvement of GPs' test ordering behaviour, which fits in well with local and regional quality improvement efforts.

To assess the value of balanced incomplete block designs in quality improvement research, and their capacity to control for the Hawthorne effect, **Chapter VIII** evaluates the study design of our trial. Local GP groups were randomised into to three arms. The GP groups in arm I received the total strategy to improve the quality of their test ordering, focusing on tests used for cardiovascular diseases and upper and lower abdominal complaints. GP groups in arm II received an identical intervention, but focusing on tests used for COPD / asthma, general malaise / vague complaints, and degenerative joint problems. GP groups in arm III received a minimal (feedback) intervention on the same tests as the GPs allocated to arm I (chapter I, figure 1). The numbers of tests related to all six clinical problems were monitored in all arms. The GPs were blinded for the interventions in the other arms. Three 2-arm comparisons were made, two within the block design, between arm I and arm II, and one with a classical design, between arm II and arm III. The block design involved analysing intervention-specific effects on changes in the number of tests ordered controlling for any Hawthorne effect. Since the GPs in both arms of the block design were subject to the same level of intervention, the Hawthorne effect was assumed to be equal in both arms. To gain insight into other potential threats to the study's validity, data on tests ordered for COPD / asthma, general malaise / vague complaints and degenerative joint problems were compared for the GPs in arm II, who had received the complete intervention with the GPs in arm III, who had only received a minimal intervention on the other three clinical problems. In the classical design the effect could to some extent be attributed to the Hawthorne effect. We concluded that the block design had a surplus value compared with the classical design. Clever use of block designs may further our understanding of non-specific effects in quality improvement research.

Chapter IX presents the general discussion and conclusions of the entire research project and the lessons to be learnt from it for the intended nation-wide implementation. Of course, innovative strategies like these are not the ultimate solution for all aspects of quality improvement in test ordering performance, and further improvement may require additional strategies. Outcome measures were volume data: the total number of tests. Unfortunately, we were unable to use clinical data. Furthermore, no long-term effects were studied. Nevertheless, the new strategy seems an innovative and practicable, efficient and cost-efficient quality instrument which can be usefully integrated within local and regional quality improvement programmes in an attempt to consistently improve GPs' test ordering behaviour.

CHAPTER XI Samenvatting

Dit proefschrift behandelt het diagnostisch aanvraaggedrag van huisartsen: het klinisch handelen van huisartsen met betrekking tot het aanvragen van aanvullend diagnostisch onderzoek, op het gebied van laboratorium, beeldvormend en functie onderzoek. In vele Westerse landen, inclusief Nederland, vragen huisartsen steeds meer diagnostische testen aan, terwijl volgens de evidence-based richtlijnen een deel daarvan overbodig is. Het proefschrift besteedt verder aandacht aan de grote interdoktervariatie tussen huisartsen wat betreft het aanvragen van diagnostische tests en een literatuur review behandelt strategieën om het aanvraaggedrag van artsen te beïnvloeden.

Hoofdstuk I geeft een globaal overzicht van dit proefschrift. Om het aanvraaggedrag van huisartsen te verbeteren, dus het realiseren dat meer aanvragen volgens bestaande richtlijnen worden aangevraagd, werd op systematische wijze een innovatieve strategie ontwikkeld, genaamd het DTO: Diagnostisch Toets Overleg. Deze meervoudige strategie bestond uit een combinatie van persoonlijke, grafische feedback rapporten, richtlijnen verspreiding en intercollegiale toetsingsbijeenkomsten. De vergelijking van het aanvraaggedrag van de individuele huisarts met zijn collega's uit de HAGRO en uit de regio was een belangrijk kenmerk van de schriftelijke feedbackrapporten.
De belangrijke kenmerken van de intercollegiale toetsingsbijeenkomsten waren de open bespreking van de feedback rapporten in tweetallen aan het begin van de bijeenkomsten, discussies over de nationale richtlijnen en het maken van individuele en groepsvoornemens op het gebied van aanvullende diagnostiek. De strategie heeft een continu karakter omdat het belangrijk is te evalueren of individuele en groepsvoornemens inderdaad leiden tot daadwerkelijke veranderingen van het aanvraaggedrag.
De meeste aandacht in dit proefschrift gaat uit naar de effecten van deze nieuwe strategie op het klinisch aanvraaggedrag van huisartsen en welke kosten en kostenbesparingen deze interventie met zich mee bracht. Deze effecten en kosteneffecten werden door middel van een gerandomiseerde studie onderzocht in een grote huisartsenpopulatie van ongeveer driehonderd huisartsen, samenwerkend in 40 huisartsengroepen (HAGRO's) in vijf regio's in Nederland. Behalve waarneming voor elkaar, scholen veel HAGRO's gemeenschappelijk na, b.v. in veel HAGRO's vindt tegenwoordig toetsing over prescriptiecijfers plaats (Farmacotherapeutisch Overleg: FTO). De coördinatie van de strategie, feedback en de supervisie, organisatie van de toetsingsgroepen lag bij de verschillende diagnostische centra. Een diagnostisch centrum is een instituut, dat meestal verbonden is aan een ziekenhuis waar huisartsen laboratorium, beeldvormend en functie onderzoek kunnen aanvragen zonder hun patiënten te verwijzen. Ongeveer een kwart van de ziekenhuizen in Nederland heeft momenteel een dergelijk diagnostisch centrum. Het geven van feedback aan adherente huisartsen over hun aanvraaggedrag is een van de taken van de medisch coördinator van het diagnostisch centrum.

Hoofdstuk II behandelt de variatie in het diagnostisch aanvraaggedrag tussen huisartsen. Professionele en contextgerelateerde determinanten die deze variatie zouden kunnen verklaren werden onderzocht. Daarvoor werd een cross-sectionele analyse gedaan van de som van 19 laboratorium en 8 beeldvormende onderzoeken, verzameld in de

vijf diagnostische centra in 1997. De samenstelling van de onderzoeks-populatie maakte analyses op een drietal niveaus mogelijk: huisarts / praktijk niveau, HAGRO-niveau en regio. In een multivariate, multilevel analyse werden deze aanvraagdata gekoppeld aan gegevens uit een enquête onder de deelnemende huisartsen over hun professionele houding ten opzichte van het aanvragen van diagnostische onderzoeken en met contextgerelateerde gegevens zoals het praktijktype of de ervaring met feedback vanuit een diagnostisch centrum. 229 Huisartsen konden in deze studie worden ingesloten. De respons op de enquête was 97%. Het totale aantal testen per huisarts per jaar was 998 (P25-P75: 663 tot 1500), met grote verschillen tussen de vijf regio's. Op professioneel niveau was 'actieve betrokkenheid bij het maken van richtlijnen' (ja/nee) en op praktijk niveau 'groepspraktijk' (vergeleken met solo- of duo-praktijken) en op regionaal niveau 'meer dan 1 jaar ervaring met het probleemgeoriënteerd laboratoriumformulier' (ja/nee), geassocieerd met respectievelijk 27%, 18% en 41% minder aanvragen. De conclusie luidt dat behalve de professionele determinanten, ook ander context-gerelateerde determinanten van invloed zijn op het aanvraaggedrag. Toekomstige studies zullen zeker rekening moeten houden met locale en regionale factoren.

Hoofdstuk III beschrijft de resultaten van een systematische literatuur review, volgens de richtlijnen van de Cochrane Collaboration, van strategieën om het aanvraaggedrag van (huis-)artsen te beïnvloeden. Een drietal hypothesen werden onderzocht. Ten eerste: of het veranderen van het absolute aantal testen, meestal door overdiagnostiek te verminderen, andere strategieën zou vergen dan pogingen om het juist gebruik van diagnostische testen (meestal door expliciete richtlijnen) te bewerkstelligen. Ten tweede is het de vraag of meervoudige strategieën mogelijk meer effect zouden hebben dan enkelvoudige en als laatste of interventies die zich richtten op de context een meerwaarde zouden hebben vergeleken met interventies die zich alleen richtten op de professional. Achtennegentig studies met 118 vergelijkingsgroepen werden geincludeerd.

De resultaten waren niet eenduidig, omdat de interventies niet steeds vergelijkbaar waren en er veel methodologische verschillen bestonden tussen de studies. Behalve algemeen geaccepteerde regels zoals het zorgen dat de te onderzoeken strategie aansluit bij de praktijk en met name rekening houdt met de weerstand tegen verandering die bij professionals vaak bestaat, is ook het doel van de interventie van belang: vermindering van het overmatig diagnostisch handelen vs. meer aanvraaggedrag volgens de richtlijnen. Deze laatste twee doelen vergen inderdaad verschillende strategieën. Enkelvoudige strategieën bleken niet minder impact hebben dan meervoudige.

Het lijkt wel belangrijk om te focussen op zowel de professional als de context. Intercollegiale toetsing en feedback lijken zowel het aantal overbodige testen als de 'kwaliteit van het diagnostisch handelen' positief te beïnvloeden. Computerondersteunende reminders verbeteren de kwaliteit van het aanvraaggedrag.

Het blijkt steeds belangrijker ook belangrijk organisatorische interventies te onderzoeken. Verder lijken strategieën die gebruik maken van patiëntenoordelen en intercollegiale toetsing zeker meer aandacht behoeven in wetenschappelijk onderzoek omdat ze in potentie een positieve invloed hebben.

In **Hoofdstuk IV** wordt de DTO-strategie geëvalueerd in een multicentre gerandomiseerd experiment met een gebalanceerd, incomplete blockdesign. De primaire uitkomstmaat was het aantal testen dat een huisarts aanvraagt per half jaar voor drie verschillende klinische beelden. De studiepopulatie bestond uit 26 HAGRO's. Arm I bestond uit 13 groepen die de totale interventie (feedback, richtlijnen en intercollegiale toetsing) kregen over testen behorende bij de drie klinische beelden cardiovasculaire ziekten, bovenbuikklachten en onderbuikklachten. Arm II bestond uit 14 HAGRO's die dezelfde interventie ondergingen met betrekking tot drie andere klinische beelden COPD/astma, vage klachten en degeneratieve gewrichtsafwijkingen. (Zie hoofdstuk I, figuur 1) Van alle deelnemende huisartsen werden de aantallen testen van alle zes klinische beelden gemonitored.

De huisartsen uit de ene arm waren blind voor het feit dat de andere groep dezelfde interventie onderging maar met betrekking tot drie andere klinische beelden. Volgens de bestaande nationale, evidence-based richtlijnen werd een daling van het totaal aantal testen waarop geintervenieerd werd en een daling per klinisch beeld opgevat als een verbetering van de kwaliteit van het aanvraaggedrag. Verder werden enkele 'overbodige' testen gedefinieerd die volgens de richtlijnen niet meer door huisartsen hoefden te worden aangevraagd. Covariantie analyses lieten zien dat voor huisartsen in arm I de daling in aantallen testen voor de klinische beelden cardiovasculaire ziekten, bovenbuikklachten en onderbuikklachten per huisarts per half jaar gemiddeld 67 meer was dan voor huisartsen in arm II ($p=0.01$). Van huisartsen in arm II daalde het aantal testen voor de klinische beelden COPD/astma, vage klachten en degeneratieve gewrichtsafwijkingen met 22 meer dan voor huisartsen in arm I maar die verandering was niet significant ($p=0.22$). Ook de overbodige testen daalden in beide armen, hoewel die daling voor de huisartsen in arm II niet significant was. De conclusie was dat de nieuwe strategie, die zich richtte op het gebruik van richtlijnen en sociale interactie en feedback tussen huisartsen, een effectief kwaliteitsinstrument kan zijn om het aanvraaggedrag van huisartsen te verbeteren.

In **Hoofdstuk V** wordt de meerwaarde onderzocht van de richtlijnen en intercollegiale toetsing op het gebied van de verbetering van het diagnostisch aanvraaggedrag, vergeleken met klassieke feedback. Het design was een multicenter trial met randomisatie op HAGRO-niveau. De totale strategie werd in 13 groepen met 85 huisartsen uitgevoerd (arm I, dezelfde als uit hoofdstuk IV), terwijl de feedback strategie in 14 groepen met 109 HAGRO's werd gedaan (arm III). Deze huisartsen kregen feedback over drie dezelfde klinische beelden (cardiovasculaire ziekten, bovenbuikklachten en onderbuikklachten), waarover de huisartsen in arm I de DTO-strategie ondergingen (Zie hoofdstuk I, figuur 1). Volgens de richtlijnen kon een absolute daling van het aantal testen opgevat worden als kwaliteitsverbetering.
Covariantie analyses lieten een significante daling van gemiddeld 51 testen per huisarts zien vergeleken met de feedback arm (arm III). Vijf 'overbodige' testen voor het klinisch beeld bovenbuikklachten gaven een significante gemiddelde daling van 13 testen meer per huisarts per half jaar dan bij de huisartsen die alleen feedback kregen. De interdoktervariatie daalde meer in de arm die de totale interventie kreeg

dan in de feedback arm.Vergeleken met het alleen maar toezenden van feedbackrapporten, verbeterde de DTO-strategie het aanvraaggedrag van huisartsen duidelijk meer en meer consistent. Dat betekent dat toetsing en feedback effectiever zijn als ze geïntegreerd zijn in een interactieve en educatieve omgeving.

In **Hoofdstuk VI** wordt een raamwerk gegeven voor kostenevaluaties van kwaliteitsbevorderende strategieën. Een kostenevaluatie werd gedaan met dezelfde trialarmen als in hoofdstuk V. Lopende kosten, ontwikkelingskosten en researchkosten werden vastgesteld. Per huisarts in de totale interventie arm (arm I) kostte de nieuwe strategie € 92.70 per half jaar, en in de feedback arm (arm III) waren de kosten voor de feedback strategie € 17.10 per huisarts per half jaar. Covariantie analyses met de gemiddelde kostenreductie per huisarts als onafhankelijke variabele en de gemiddelde kosten per huisarts bij de nulmeting en de regio als onafhankelijk variabele, gaven een significante hogere kostenreductie van € 144 per huisarts per half jaar voor de totale interventie arm vergeleken met de feedback arm (p=0.001). In het raamwerk behoren ook niet-geldelijke voordelen meegenomen te worden. De conclusie was dat de DTO-strategie een waardevol kwaliteitsinstrument is en dat bovendien de kosten en kostenreducties van deze nieuwe strategie het wenselijk maken verder te onderzoeken hoe deze op grotere schaal te implementeren.

In **Hoofdstuk VII** wordt het gebruik en de toepasbaarheid van de DTO-strategie beschreven met behulp van een prospectieve procesevaluatie tijdens de interventieperiode van twee jaar.

Alle 193 huisartsen van arm I en II kregen de geplande 1158 feedbackrapporten. Gegevens van 156 toetsingsbijeenkomsten gaven een opkomstpercentage van 81 % in het eerste jaar (95% BI: 77%-85%) en 73% (95%BI: 68%-77%) in het tweede jaar. De drie belangrijkste elementen van de toetsingsbijeenkomsten: paarsgewijze bespreking van de rapporten, relateren van het aanvraaggedrag aan de richtlijnen en het maken van individuele en groepsvoornemens werden in beide jaren voldoende uitgevoerd. In het eerste jaar gaven de huisartsen de totale strategie een 7.55 (95% BI: 7.46-7.64) op een 10-puntsschaal en 7.51 (95% BI: 7.30-7.74). Het DTO bleek implementabel in de dagelijkse praktijk en lijkt goed inpasbaar in locale en regionale nascholings- en toetsingsprogramma's.

In **Hoofdstuk VIII** wordt de meerwaarde van het blok design bepaald ten opzichte van een klassieke design met twee armen, o.a. in het omgaan met non-specifieke effecten zoals het Hawthorne-effect. Deze studie werd gebaseerd op de totale 3-armige studie. HAGRO's uit arm I kregen de totale strategie over testen, behorend bij de klinische beelden hart- vaatziekten, boven- en onderbuikklachten. HAGRO's uit arm II kregen de complete interventie over testen, behorend bij de klinische beelden COPD/asthma, algemene malaise en moeheid en degeneratieve gewrichtsafwijkingen.
HAGRO's uit arm III kregen de minimale feedback interventie over testen behorend bij de klinische beelden hart- vaatziekten, boven- en onderbuikklachten (Zie hoofdstuk I, figuur 1). In alle armen werden van alle huisartsen alle testen behorend bij een van de zes klinische beelden geregistreerd. Huisartsen waren blind voor de interventie in

de andere armen. De huisartsengroepen uit de eerste twee armen waren controlearm van elkaar. Drie 2-armige vergelijkingen waren mogelijk, twee binnen het blokdesign en een klassiek design tussen arm II en arm III w.b. de testen behorend bij de klinische beelden COPD/asthma, algemene malaise en moeheid en degeneratieve gewrichtsafwijkingen.
In het blokdesign werden interventie-specifieke effecten geanalyseerd, waarbij gecontroleerd werd voor het Hawthorne-effect. Omdat de huisartsen uit de armen van het blokdesign dezelfde mate van interventie ondergingen maar wel over verschillende klinische beelden werd het Hawthorne-effect gelijkelijk verdeeld over deze twee armen. In het blok design trad geen verbetering op voor de testen waarop niet geintervenieerd werd. In het klassieke design kon het effect voor een deel worden toegeschreven aan het Hawthorne-effect en dus had het blok design een duidelijke meerwaarde ten opzichte van het klassieke design. Een juist gebruik van het blok design in kwaliteitsonderzoek kan de kennis van de invloed van non-specifieke effecten in kwaliteitsonderzoek verbeteren.

Hoofstuk IX tenslotte gaat over de algemene conclusies van het hele project. Conclusies uit de determinantenstudie, het literatuur review en de klinische en kosteneffecten van de DTO-strategie worden nogmaals kritisch beschouwd. Het is belangrijk te constateren dat in dit onderzoek de lange termijn effecten niet onderzocht konden worden en dat vooral kwantitatieve uitkomstmaten gebruikt werden. Het bleek (nog) niet mogelijk om klinische data te gebruiken. Natuurlijk is de DTO- strategie niet de ultieme oplossing om het aanvraaggedrag van huisartsen blijvend te verbeteren. Waarschijnlijk zijn ook andere strategieën mogelijk en nodig. Toch wordt geconcludeerd dat het ontwikkeld kwaliteitssysteem het diagnostisch aanvraaggedrag van huisartsen op een effectieve, kostenefficiënte en, in de dagelijkse praktijk toepasbare manier, kan verbeteren. Deze strategie kan zeker ook gebruikt worden voor andere vormen van intercollegiale toetsing bijvoorbeeld over verwijscijfers of prescriptiecijfers. Niet alleen huisartsen kunnen er hun voordeel doen mee doen, ook andere samenwerkende professionals kunnen (elementen uit) deze methode toepassen. Een bredere invoering van het DTO wordt aanbevolen.