Towards Evidence-based Government Supervision in Healthcare

SANDRA OUDE WESSELINK

Towards Evidence-based Government Supervision in Healthcare

SANDRA OUDE WESSELINK

ISBN
978-94-62330-39-9
Digital version
www.oudewesselink.net/proefschrift

Towards Evidence-based Government Supervision in Healthcare
Thesis, Erasmus University Rotterdam

Design and lay-out
Irma Theunisse, www.irmatheunisse.nl
Print
Gildeprint – Enschede

Thanks to
Apotheek Biermasz

The studies presented in this thesis were financially supported by a grant from the Dutch Healthcare Inspectorate. This thesis was printed with financial support of the Department of Public Health Erasmus MC Rotterdam and the Erasmus University Rotterdam.

Towards Evidence-based Government Supervision in Healthcare

Op weg naar evidence-based overheidstoezicht op de gezondheidszorg

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 8 september 2015 om 15:30 uur

door

Sandra Frederieke Oude Wesselink

geboren te Zwolle

ERASMUS UNIVERSITY ROTTERDAM

voor mijn opa´s en oma´s

Promotiecommissie

Promotoren:

Prof.dr. J.P. Mackenbach

Prof.dr. P.B.M. Robben

Overige leden:

Prof.dr. P.J.E. Bindels

Prof.dr. F.J.G. Janssens

Prof.dr. C. Wagner

Copromotor:

Dr. H.F. Lingsma

CHAPTER 1

Government supervision in healthcare

In healthcare, low quality of care can have serious consequences for patients and their relatives, but individual patients are generally unable to check the quality and safety of care. That is one of the reasons why the government regulates public and private parties in the field of healthcare. In the Netherlands, an independent governmental control agency, the Dutch Healthcare Inspectorate, is responsible for this supervision on quality and safety of care. For various reasons, activities of the inspectorate receive public and political attention on a regular basis. These reasons are for example that healthcare organisations desire less regulatory burden, politicians aim to reduce the budgets of the inspectorate and the public asks for maximum safety and highest quality of care.[1] While continuously trying to deal with these contradictory of requests for both reduction and increase of supervision, the inspectorate aims to improve quality and safety of care.

QUALITY OF CARE

In the early 1990s the Institute of Medicine defined quality of care as: "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge".[2] This definition, containing an aspect of outcome and an aspect of process, is in line with the definition of Donabedian.[3] According to his model, quality of care can be assessed by examining structures, processes and outcomes of care.[3] Structures are the attributes of the setting in which care occurs, for example facilities and qualification of personnel. The term 'processes of care' refers to what is done when care providers are giving care, for example the practitioner's activities in making a diagnosis and recommending or implementing treatment. Outcomes of care include health status of patients and populations, for example mortality or functional outcome, but also patient's satisfaction with care.

Quality of care can be improved using various strategies[4]:

- Provider education
- Provider reminder systems and decision support
- Patient education
- Organisational change
- Financial incentives
- Audit and feedback

"Audit and feedback" refers to providing care providers with a summary of clinical performance. This summary must include external assessment[5], which can be divided into models of peer review, accreditation, and inspection. In this thesis we will focus on external assessment through inspections, namely government supervision.

GOVERNMENT SUPERVISION AND QUALITY OF CARE

In this thesis we will focus on government supervision that aims to protect patients and the public by controlling or influencing the behaviour of healthcare organisations or care providers.[6 7]

Healthcare systems differ greatly across Europe and the world. Therefore the role, organisation, and even presence of government supervision also differ between countries.[8] In the Netherlands, government supervision in healthcare is conducted by the Dutch Healthcare Inspectorate (later: inspectorate), an independent agency of the Ministry of Health, Welfare, and Sport. The inspectorate supervises a sector in which 1.3 million people work, in 40,000 institutions and companies. It enforces 26 laws, including for example the Care Institutions Quality Act.[9] The ultimate aim of the inspectorate is to improve population health by improving quality of care. The primary instruments of the inspectorate are advice and encouragement. If these do not achieve the desired result, the inspectorate can implement corrective action by, for example, increasing the supervision or by limiting the ability to practice a profession. The inspectorate supervises by using a combination of three methods: supervision in response to calamities, risk-based supervision, and theme-based supervision. Supervision in response to calamities or emergencies means that healthcare organisations have the obligation to report calamities or emergencies to the inspectorate. The inspectorate evaluates these events and in case of structural shortcomings action is taken.[10] Risk-based supervision is based on indicators, which are used to identify care providers or institutions at risk of having low quality of care. Care providers provide quality indicators to the inspectorate for further risk analysis.[11] Theme-based supervision is directed at a specific issue in care, which is sometimes requested by the minister or parliament. Typically, the current state of healthcare with regard to that specific topic is investigated and subsequently improvements are requested. In theme-based supervision, the inspectorate collaborates with other stakeholders such as patients, care providers, healthcare organisations and health insurance companies.[12] In this thesis we will focus on theme-based supervision.

POTENTIAL EFFECTS OF GOVERNMENT SUPERVISION AND STUDYING THESE EFFECTS: EVIDENCE-BASED SUPERVISION

Little is known about the actual effects of the supervision programmes of healthcare inspectorates on quality of care and specifically health outcomes. Therefore the Health Council of the Netherlands recommended in its report[13] to aim for evidence-based supervision.[14] Evidence-based supervision is defined as the conscientious, explicit, and judicious use of current best evidence in supervision

by professional inspectors. This definition is derived from Sackett's commonly used definition of evidence-based medicine.[15] Recently, the OECD reported core principles on regulatory enforcement and inspections.[16] Their first principle involves evidence-based enforcement. The report states that "regulatory enforcement and inspections should be evidence-based and measurement-based: deciding what to inspect and how should be grounded on data and evidence, and results should be evaluated regularly". Despite the objective of evidence-based supervision, a review of the Dutch and international scientific literature has shown that research into the effects of supervision is still in its infancy.[13] The researchers found that while no quantitative studies had been conducted, a few qualitative studies have taken place in Australia, the United States and the United Kingdom.[13] Recently, some quantitative studies on supervision have been published.[17 18] To further improve evidence-based supervision an Academic Collaborative Centre on supervision of the Dutch Healthcare Inspectorate was started in 2011. In this centre the Dutch Healthcare Inspectorate collaborates with four research institutes. The aims of this collaboration are 1) to professionalise supervision by evaluating the current practice and effectiveness studies, 2) to contribute to the development of supervision methods and instruments, and 3) to expand and distribute scientific knowledge on supervision.[19]

Information on the effects of supervision may help inspectorates in decision making and further improve their working methods. Research may also provide society with information on how successful supervision programmes have been.[13] Decision making based on evidence regarding the effectiveness of supervision programs leads to evidence-based supervision. Therefore, more research is needed on the effects of government supervision on quality of care and specifically on health outcomes.

First, however, it must be determined how the effect of supervision should be assessed. By definition, the inspectorate cannot directly affect patient's health, but can only have an effect through healthcare providers. This limitation is also noted in the effect chain that is used by the Dutch Healthcare Inspectorate.[9] This scheme depicts their activities and the potential effects of these activities (Figure 1). The first step is the effect of the inspectorate on the compliance of those inspected, with respect to guidelines and legislation. This is a direct result of the supervision programmes. A second step is the effect of improved compliance by care providers on health outcomes. However, not all increases in compliance will directly lead to improved health outcomes. This issue is well-known in quality of care research, where it has been found that the correlation between structures and processes and the outcomes of care is often weak, even for strongly evidence-based processes.[20] The change in process of care is usually not from 0 to 100%

compliance, but for example from 50 to 70%. Consequently, only 20% of the patients potentially benefit from the improved care. The effect of an evidence-based process of care is typically in the magnitude of 30% lower risk of the unfavourable outcome. Therefore, even substantial changes of differences in processes of care may have only small effects on health outcome. However, for the inspectorate, to influence health outcomes, this association is crucial. Therefore, to be able to draw conclusions regarding the effect of supervision on health outcomes, it is important to have insight into the relation between structures and processes of care, and health outcomes.

Aim

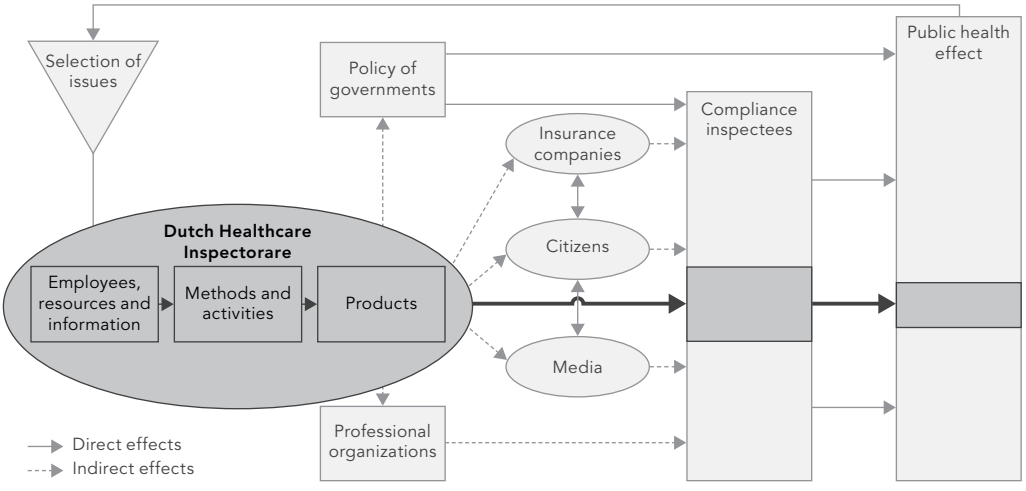
The general aim of this thesis is to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes. We evaluated two cases of government supervision: supervision on integrated diabetes care and supervision on quit-smoking counselling by midwives. We addressed two specific research questions:

1. What are the effects of these two government supervision programmes on structures and processes of care as well as on health outcomes?
2. In these cases of government supervision, what is the relation between structures and processes of care, and health outcomes?

POTENTIAL STUDY DESIGNS AND CASES FOR EVALUATING GOVERNMENT SUPERVISION

This study aims to generate empirical evidence of the effects of government supervision on quality of care

Figure 1 *Effect chain of Dutch Healthcare Inspectorate*



and specifically health outcomes. We first provide an overview of the potential study designs that can be used to generate this evidence. Studying the effect of government supervision is challenging because each supervision programme consists of many components which together form a complex intervention. Furthermore, the formal context of government supervision often poses barriers to creating adequate control conditions. Therefore, not all classic epidemiological study designs can feasibly be applied in the context of supervision.

The research described in this thesis is performed within the paradigm of health services research.[21] The classification of research designs within this paradigm possibly does not correspond to classifications that are used in other research fields, such as social sciences. We acknowledge that many classifications of research methods and designs are possible and none of them will perfectly describe our research on the effects of government supervision on quality of care.

Quantitative methods

Quantitative designs can be divided into experimental, quasi-experimental and observational designs, with the first providing the highest level of evidence and the last the lowest level of evidence with regard to the effectiveness of the intervention.[22]

Experimental designs

In a randomised experimental design the inspectees, such as GPs or hospitals, are randomly divided into two groups.[23] One group receives the supervision, for example site visits of the inspectorate.[24] The other group receives no supervision. This design is called a randomised controlled trial (RCT). In cluster-RCTs, the randomised research units are larger units, such as regions, organisations or practices.[25]

RCTs provide strong evidence, because the possibility of confounding due to differences between the intervention and control group is eliminated by the randomisation. However, RCTs may not always be feasible in the context of supervision. Randomisation is for example undesirable when the inspectorate wants to directly address care providers with a low quality of care. Creating a control group may then conflict with the aim of the inspectorate to reach as many care providers as possible as quickly as possible, to eliminate the risk of low quality of care. Furthermore, the interconnections between care providers in the Netherlands often make it challenging to isolate groups (intervention and control) of care providers for research purposes.

Quasi-experimental designs

When randomisation is practically or ethically impossible, quasi-experimental designs can be considered.[22] Quasi-experiments are experiments in which research units are not assigned to conditions randomly, but differences in exposure to the intervention occur naturally over time or between groups.[26]

There are several designs that can be classified as quasi-experimental studies.

When a control group is available and the intervention is not assigned randomly, the analysis should be adjusted for differences between groups that existed before the intervention. One method of adjustment involves using a propensity score. The propensity score is the conditional probability of assignment to a particular treatment, based on covariates. Adjustment for this score removes the bias that is due to all observed covariates.[27]

When no control group is available, changes over time can be studied. The simplest design is a before-after study. In before-after studies data is collected at two points in time, before and after the intervention. In this design it is difficult to discriminate between the intervention effect and a secular trend.[22] This is less of a problem in designs that use interrupted time series, where data is collected at multiple time points before and after the intervention. This allows for a more reliable estimation of the intervention effect while taking into account the underlying secular trends. Nevertheless, it is often difficult to be certain that no other changes, coinciding in time with the intervention, explain the observed changes.

In quasi-experimental studies without a control group, conclusions on causality should be drawn with caution, as it is difficult to know what would have happened without the intervention. Overall, quasi-experimental designs will always be more prone to bias then RCTs, as unmeasured confounding can never be excluded with absolute certainty.

Observational designs

In observational designs the structural features of experiments are missing, i.e. there is no sudden change in exposure to the intervention.[22] An example is a cross-sectional study of the effects of supervision, where differences in the quality of care or health outcomes are compared between regions or healthcare organisations, in which different forms or intensities of supervision have been applied. However, variations in supervision are usually not random and may actually be based on variations in quality of care. These pre-existing differences remain unobserved in cross-sectional studies, where confounding variables are only measured after the intervention. These studies may therefore result in confounding by indication.[28]

In summary, all strong designs have features that may be challenging for evaluating supervision. Thus, the choice of a research design for the evaluation of a supervision programme requires a trade-off between feasibility and the strength of the design for causal inference.

Qualitative methods

When a study aims to identify the mechanisms behind the effect of supervision, a qualitative method might be useful.[22] We will consider three types of qualitative studies. In programme theory analyses, the programme theory of the supervision programme is reconstructed and made explicit. In ethnographic

research, the daily practice and behaviour of inspectors and inspectees is studied. And finally, in case studies generic mechanisms are explored in one or more specific cases that are thought to represent a larger universe of patients, providers and settings. Types of data collection in qualitative research include document analysis, (participating) observations and interviews. Although it is theoretically possible to establish causal relationships with qualitative research,[29] doing so will remain difficult for the complex relation between supervision programmes and quality of care. However, qualitative research may provide more insight in the mechanisms of the different elements of a supervision programme.

Selection of cases for evaluation

The above-mentioned opportunities and limitations of epidemiological study designs for the evaluation of supervision were also described by the Health Council of the Netherlands.[13] They formulated the following set of methodological criteria that should be considered when choosing the design and topic for studies to measure the effect of supervision programmes:

1. Description of problem, aim, intervention and effects: the extent to which the problem, aim and content of the supervision intervention and the intended effects can be described.
2. Effect mechanism: the extent to which a description of the effect of the supervision intervention is available or can be constructed afterwards and whether this mechanism is plausible.
3. Randomisation of intervention and control group: the possibilities of randomisation of intervention and control groups, the number of units that can be randomised and the ethical-political and financial possibilities of randomisation; the comparability of the groups.
4. Effect measures: the extent to which effect can reliably be measured, independent data source and independent researcher.
5. Data before/after comparison or trend data: the availability of data for a before-after comparison or possibilities to collect these data, more study measurements before and after the supervision programme.
6. Bias: the extent to which information on confounding variables is available or can be collected.

These criteria were used to select the supervision programmes for evaluation and to design the studies. The programmes were selected from the work plan of the inspectorate of 2011. In the selection feasibility also played a role; for example the programme had to start within the research period. The selected programmes were two theme-based supervision programmes; one on integrated diabetes care and the other on quit-smoking counselling by midwives.

DIABETES MELLITUS TYPE 2

Diabetes mellitus type 2 is a chronic metabolic disease that occurs when the body cannot effectively use the insulin it produces. Diabetes mellitus type 2 (later: diabetes), which was formerly known as non-insulin-dependent or adult-onset diabetes, develops gradually and progressively.[30] Therefore, diabetes may remain undetected for many years and the diagnosis is often made when a complication appears or a routine blood or urine glucose test is done. Over time, diabetes can damage the heart, blood vessels, eyes, kidneys and nerves.[31] Diabetes is a major public health problem in Europe and the U.S.[32] In 2011, 5% of the Dutch population had diabetes.[33]

The main goal of treatment of diabetes is achieving blood glucose control. Further treatment includes blood pressure control and foot care. People with diabetes also have to eat a healthy diet, engage in regular physical activity, maintain a normal body weight and avoid tobacco use to prevent further complications of the disease. Over time most people with diabetes will require oral drugs or insulin. Complications of diabetes can be avoided by screening and treatment for retinopathy which causes blindness, blood lipid control to regulate cholesterol levels and screening for early signs of diabetes-related kidney disease.[30] Effective treatment can prevent or delay the development of complications. However, for example in Europe, the provision of care is far from optimal and target levels of health outcomes are often not reached.[34] Therefore several quality-improvement initiatives are needed to improve the quality of treatment in diabetes care.[35]

As the Dutch Minister of Health wanted to improve diabetes care, an action programme was started to develop integrated care for diabetes patients through the creation of care groups.[36] Care groups are organisations that provide integrated diabetes care. They consist of 3 to 250 general practices, which are paid with a bundled payment.[37] Bundled payment means paying a single fee for all medical services involved in an episode of care.[38] In this case, care groups receive a yearly fee for each diabetes patient in the care group. Care groups consist of multiple healthcare providers, situated within and outside the general practice; they are often owned by general practitioners. General practitioners and practice nurses provide diabetes care. Practice nurses are mainly involved in performing check-ups for diabetes patients. Other care providers are contracted by the care group. Care groups are comparable with accountable-care organisations in the United States.[39] However, accountable-care organisations have a much broader scope, which includes hospital care. Legal requirements are much more extensive in the U.S. than in the Netherlands.

Supervision of integrated care for diabetes patients

The supervision programme that was evaluated focused on integrat-

ed care for diabetes mellitus type 2 patients. The inspectorate started this programme that aimed to improve diabetes care in care groups. The supervision programme targeted directors of care groups and took place in 2011 and 2012. Topics of the programme were selected based on a risk analysis. The risk analysis included literature research and discussions within the inspectorate and with other interest groups. Site visits were the main activity of the supervision. The following topics were selected by the inspectorate and assessed during each site visit: multidisciplinary patients' files, individual care plans, prevention and self-management, continuity of care, quality and transparency.[40] [41] The programme was assigned to 20 care groups, randomly selected from all 100 care groups. The site visits were announced by mail to the care groups and the topics of the supervision were provided. Care groups were inspected for one day by two inspectors. During these visits the inspectors talked with directors, general practitioners and practice nurses. After the site visits, an individualized report with recommendations for improvements was sent to the care groups, on average three months after the visit. Next, the inspectorate aggregated the results and recommendations in a national report, which was sent to all care groups in the Netherlands.

SMOKING DURING PREGNANCY

The second case in our research involves smoking during pregnancy. It has been known for decades that cigarette smoking reduces health in general and specifically during pregnancy. However, six percent of women in the Netherlands smoke during pregnancy.[42] Among lower-educated women, the prevalence of smoking during pregnancy is around 14%. Maternal smoking is associated with a higher risk of foetal mortality and of adverse birth outcomes such as stillbirth, preterm birth, small for gestational age, intrauterine growth restriction, and congenital heart defects.[43]

In the Netherlands pregnant women can choose where they would like to give birth, and many do so at home.[44] For low-risk pregnancies and deliveries midwives may provide care on their own during gestation, childbirth and the postpartum period. Midwifery education is a 4-year vocational training (at a university of applied science) at a bachelor's degree level.[45] Primary care midwives work in private practices, either as self-employed practitioners or as employees in someone else's practice. Self-employed practitioners work alone or in partnership with one or more other midwives. Most practices work with teams of 3 to 5 midwives caring for one pregnant woman, with one team member assisting at delivery and the team sharing information about the woman via an electronic patient file. Each practice has a midwife on call 24/7.

Midwives can play a key role in provision of quit-smoking counselling of pregnant smokers; in the Netherlands primary care midwives see 80% of all pregnant women at an

early stage of pregnancy in midwifery practices.

Most midwives who provide quit-smoking counselling use V-MIS, a minimal intervention strategy that increases the quit-smoking rate in pregnant smokers.[46] It targets the clients of midwifery practices and is based on the stages of change theory.[47] Midwives use V-MIS during their normal consultations with pregnant smokers or plan a separate consultation to provide counselling. V-MIS comprises seven steps. In step 1, the midwife identifies the smoking behaviour of the woman and partner. In step 2, the midwife attempts to enhance the motivation to quit. In step 3, the midwife and woman discuss barriers for successful quitting and how to mobilise social support for quitting. In step 4, the midwife and woman agree on a quit date. In step 5, the midwife discusses and provides additional self-help materials. In step 6, the midwife provides aftercare if necessary. In step 7, the midwife supports the woman to prevent relapse after delivery.[46]

Supervision on primary care midwives providing quit-smoking counselling

As perinatal mortality was higher in the Netherlands than in other European countries[48] and research showed that the quality of quit-smoking counselling provided by midwives was low,[49] the inspectorate started a programme aimed at improving the provision of quit-smoking counselling to pregnant women by primary care midwives. The supervision programme consists of different elements: assessments with questionnaires and a personal report, announcement of a deadline by which midwives should provide counselling and assessments with site visits and a personal report. We aimed to evaluate this supervision programme on its effectiveness.

In 2010 the inspectorate randomly distributed a questionnaire to a sample of midwifery practices. The goal of the questionnaire was to gain insight into the current state of provision of counselling in midwifery practices. After the questionnaire, the inspectorate also visited 10 midwifery practices. Results of the questionnaire and visits as well as points for improvement were sent to practices in a personal report. From the aggregated results of the questionnaire, the inspectorate concluded that large improvements in provision of counselling were needed. The inspectorate started a multifaceted inspection to improve counselling in practices. The results of the questionnaire have been published in an article in the professional journal of midwives in the Netherlands.[50] Furthermore, the inspectorate has stated that professional norms should be followed. A deadline was announced, indicating when all practices were expected to comply with these professional norms. All practices received an enforcement letter to inform them about the professional norms and date from which these norms would be enforced. Finally, the inspectorate announced site visits to practices to verify whether professional norms were imple-

mented. After the deadline the inspectorate inspected a sample of all midwifery practices with site visits. During these visits inspectors checked whether midwives were complying with the guideline. Site visits were announced by mail to practices and topics of supervision were provided. Two inspectors took half a day to inspect each practice. Following the site visits, all the inspected practices received a personal report with feedback on their counselling and a time frame for implementing the required improvements.

OUTLINE OF THIS THESIS

This thesis describes a series of studies conducted to answer our research questions. The first part focuses on the effects of supervision programmes. Chapter 2 describes a randomised controlled trial that tested for the effects of supervision on integrated diabetes care. In chapter 3 we present an RCT, an interrupted time-series design and a before-after study on the effect of supervision on midwives' quit-smoking counselling. Chapter 4 describes a qualitative exploration of the effect of supervision on midwives' quit-smoking counselling.
In part two, relations between structures and processes of care and health outcomes are studied. Chapter 5 presents a cross-sectional study on the relation between guideline adherence in integrated diabetes care and health outcomes. Chapter 6 describes a quasi-experimental study on the effect of quit-smoking counselling on the smoking behaviour of pregnant women and birth weight. Chapter 7 provides the general discussion of this thesis. It summarises the main results from the studies and addresses the research questions of this thesis. Finally, chapter 7 also provides a discussion of the methodological considerations of this thesis, interpretation of the results and implications for further research and policy development.

REFERENCES

1. Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid). Supervising public interests, towards a broader perspective on government supervision (Toezien op publieke belangen, naar een verruimd perspectief op rijkstoezicht). Den Haag, 2013.
2. Lohr KN. Applications of health status assessment measures in clinical practice: Overview of the third conference on advances in health status assessment. *Medical Care* 1992;30(5):MS1-MS14.
3. Donabedian A. The quality of care: How can it be assessed? JAMA 1988;260(12):1743-48.
4. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood)* 2005;24(1):138-50.
5. Shaw C. External assessment of health care. BMJ 2001;322(7290):851.
6. Robben PBM, Bal R, Grol R. Governmental supervision by the Dutch Health Care Inspectorate (Overheidstoezicht door de Inspectie voor de Gezondheidszorg). Den Haag: WRR, 2012.
7. Walshe K. *Regulating healthcare: a prescription for improvement*?: McGraw-Hill International, 2003.
8. Gno D, den Breejen E, Putters K, et al. Supervising the quality of care in changing healthcare systems - An international comparison. Rotterdam: Erasmus MC University Medical Center Rotterdam, 2008.

9. Dutch Healthcare Inspectorate (Inspectie voor de Gezondheidszorg). Multi-annual term policy plan 2012-2015 (Meerjarenbeleidsplan 2012-2015). Utrecht, 2011.
10. Huisman A, Robben P, Kerkhof A. An examination of the Dutch Health Care Inspectorate's supervision system for suicides of mental health care users. *Psychiatr Serv* 2009;60(1):80-85.
11. Borghans I, Hekkert KD, den Ouden L, et al. Unexpectedly long hospital stays as an indicator of risk of unsafe care: an exploratory study. BMJ *Open* 2014;4(6):e004773.
12. Neefjes FCJ, Bal R, Robben PBM. Developments of standards through theme-based regulation (Normontwikkeling door Thematisch Toezicht). *Tijdschrift voor Toezicht* 2011;4(2).
13. Health Council of the Netherlands (Gezondheidsraad). Towards evidence-based supervision. Research on effects of supervision by the Dutch Health Care Inspectorate (Op weg naar evidence based toezicht. Het onderzoek naar effecten van toezicht door de Inspectie voor de Gezondheidszorg). Den Haag, 2011.
14. Robben P. *Supervision in a glass house: The effectiviness of supervision on the quality of health care* (*Toezicht in een glazen huis: Effectiviteit van het toezicht op de kwaliteit van de gezondheidszorg*), 2010.
15. Sackett DL, Rosenberg W, Gray JA, et al. Evidence based medicine: what it is and what it isn't. BMJ 1996;312(7023):71-72.
16. OECD. Regulatory enforcement and inspections, OECD best practice principles for regulatory policy, 2014.
17. Klerks MCJL, Ketelaars CAJ, Robben P. Unannounced, compared with announced inspections: A systematic review and exploratory study in nursing homes. *Health Policy* 2013;111(3):311-19.
18. Tuijn S, Janssens F, Robben P, et al. Reducing interrater variability and improving health care: a meta-analytical review. *J Eval Clin Pract* 2012;18(4):887-95.
19. Vedder A. Factsheet on the Academic Collaborative Centre for Supervision (Factsheet Academische Werkplaats Toezicht) Utrecht, 2013.
20. Lingsma HF, Steyerberg EW, Scholte op Reimer WJM, et al. Statin treatment after a recent TIA or stroke: is effectiveness shown in randomized clinical trials also observed in everyday clinical practice? *Acta Neurol Scand* 2010;122(1):15-20.
21. Bowling A. *Research Methods in Health: Investigating Health and Health Services*: Open University Press, 2014.
22. Grol R, Baker R, Moss F. *Quality Improvement Research: Understanding the Science of Change in Health Care*: BMJ Publishing Group, 2003.
23. Vandenbroucke JP, Hofman A, van Stiphout W. *Grondslagen der epidemiologie*: Elsevier/Bunge, 1999.
24. van Dishoeck AM, Oude Wesselink SF, Lingsma HF, et al. Transparency: can the effect of governmental surveillance be quantified? (Transparantie: is het effect van toezicht te meten?). *Ned Tijdschr Geneeskd* 2013;157(16).
25. Mate KS, Ngidi WH, Reddy J, et al. A case report of evaluating a large-scale health systems improvement project in an uncontrolled setting: a quality improvement initiative in KwaZulu-Natal, South Africa. BMJ *Qual Saf* 2013;22(11):891-98.
26. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*: Wadsworth Cengage learning, 2002.
27. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70(1):41-55.
28. de Koning JS, Klazinga NS, Koudstaal PJ, et al. The role of 'confounding by indication' in assessing the effect of quality of care on disease outcomes in general practice: results of a case-control study. BMC *Health Serv Res* 2005;5(1):10.
29. Maxwell JA. Using qualitative methods for causal explanation. *Field methods* 2004;16(3):243-64.
30. WHO. Diabetes factsheet, 2014.
31. American Diabetes Association. Standards of medical care in diabetes - 2013. *Diabetes Care* 2013;36 (suppl. 1): S11-S66.
32. Cowie CC, Rust KF, Byrd-Holt DD, et al. Prevalence of diabetes and high risk for diabetes using A1C criteria in the US population in 1988–2006. *Diabetes Care* 2010;33(3):562-68.
33. Baan CA, Poos MJJC. How common is diabetes mellitus and how many people die from it? (Hoe vaak komt diabetes mellitus voor en hoeveel mensen sterven eraan?), 2013.
34. Stone MA, Charpentier G, Doggen K, et al. Quality of Care of People With Type 2 Diabetes in Eight European Countries Findings from the Guideline Adherence to Enhance Care (GUIDANCE) study. *Diabetes Care* 2013.
35. Dutch Healthcare Inspectorate (Inspectie voor de Gezondheidszorg). State of healthcare, integrated care chronically ill (Staat van de Gezondheidszorg 2003, Ketenzorg bij chronisch zieken). Den Haag: Ministry of Health, Welfare and Sport, 2003.
36. Taskforce Diabetes Program (Taakgroep Programma Diabeteszorg). Diabetes care better (Diabeteszorg beter), 2005.
37. de Jong-van Til JT, Lemmens LC, Baan CA, et al. The organization of care groups in 2011 (De organisatie van zorggroepen anno 2011). Bilthoven (NL): National Institute for Public Health and the Environment (RIVM), 2012.

38. de Bakker DH, Struijs JN, Baan CA, et al. Early results from adoption of bundled payment for diabetes care in the Netherlands show improvement in care coordination. *Health Aff (Millwood)* 2012;31(2):426-33.
39. Struijs JN, Baan CA. Integrating care through bundled payments - lessons from The Netherlands. *N Engl J Med* 2011;364(11):990-91.
40. Dutch Health Care Inspectorate (Inspectie voor de Gezondheidszorg). Implementation care standard diabetes after four years of insufficient progress (Implementatie zorgstandaard diabetes na vier jaar onvoldoende gevorderd). Utrecht: Ministry of Health, Welfare and Sport, 2012.
41. Ketelaars C. Risk analysis on integrated care policy. *Int J Integr Care* 2011;11(Suppl).
42. Lanting CI, Van Wouwe JP, Van Den Burg I, et al. Smoking during pregnancy (Roken tijdens de zwangerschap). *Ned Tijdschr Geneeskd* 2012;156(50):2123.
43. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* 2004;6(Suppl 2):S125-S40.
44. Dutch professional midwifery organisation (Koninklijke Nederlandse Organisatie van Verloskundigen), Dutch expertise centre on tobacco control (STIVORO). Guide for policy smoking cessation counselling in midwife practices (Handreiking Ontwikkeling Beleid voor Begeleiding Stoppen met Roken in Verloskundigenpraktijken), 2012.
45. Wiegers TA, Warmelink JC, Spelten ER, et al. Work and workload of Dutch primary care midwives in 2010. *Midwifery* 2013.
46. de Vries H, Bakker M, Mullen PD, et al. The effects of smoking cessation counseling by midwives on Dutch pregnant women and their partners. *Patient Educ Couns* 2006;63(1):177-87.
47. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983;51(3):390-95.
48. Mohangoo AD, Buitendijk SE, Hukkelhoven CWPM, et al. High perinatal mortality in the Netherlands compared to other European countries: Peristat II study (Hoge perinatale sterfte in Nederland vergeleken met andere Europese landen: de Peristat-II-studie). *Ned Tijdschr Geneeskd* 2008;152(50):2718-27.
49. Segaar D. Adoption and implementation of smoking cessation support in healthcare [PhD Thesis]. Maastricht University, 2006.
50. Reulings PGJ, van der Lans SMGA, Wentzel HR, et al. Grab your chance, smoking-cessation couseling by midwives (Pak je kans: stoppen met roken begeleiding door verloskundigen). *Tijdschrift voor Verloskundigen* 2011;april.

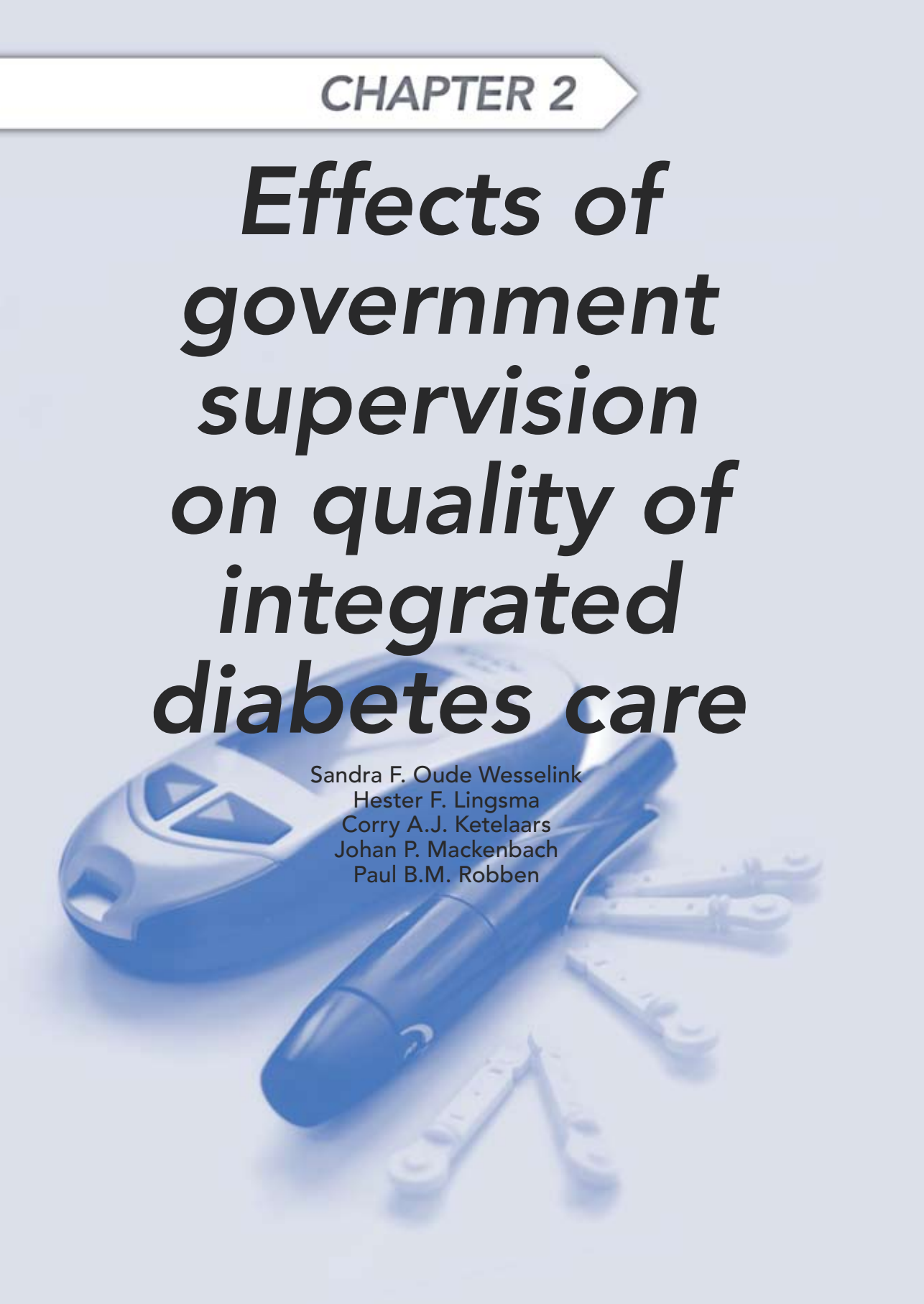
PART I

Effect of government supervision on quality of care

CHAPTER 2

Effects of government supervision on quality of integrated diabetes care

Sandra F. Oude Wesselink
Hester F. Lingsma
Corry A.J. Ketelaars
Johan P. Mackenbach
Paul B.M. Robben



ABSTRACT

Introduction

The quality of integrated diabetes care is important for reducing the burden of diabetes. Therefore we have evaluated the effect of a supervision program on the quality of integrated diabetes care in the Netherlands in the 2011-2012 period.

Methods

In this cluster RCT the supervision program was assigned to randomly selected care groups providing care to diabetes patients. The supervision program included announcements of inspections, site visits and sending individualized reports. Indicators of effectiveness were derived from the structures, processes and outcomes of care. These indicators were collected from patients' files, before and after the supervision program. Hierarchical linear and logistic regression models were used to analyze data from 356 patients of 10 intervention and 8 control care groups.

Results

Structures and processes of care did not improve more in the intervention groups than in the control care groups. Moreover, health outcomes did not improve more in the intervention groups than in the control care groups. Although structures of care improved over time in the total population of intervention and control care groups, there were no changes in process of care or health outcomes.

Conclusions

In this cluster RCT we could not demonstrate improvements in quality of integrated diabetes care resulting from the supervision program. Although structures of care did improve over time, other quality-improvement initiatives are necessary to substantially strengthen integrated care for diabetes patients.

Adapted from: Oude Wesselink SF, Lingsma HF, Ketelaars CAJ, Mackenbach JP, Robben PBM. Effects of government supervision on quality of integrated diabetes care: A cluster randomized controlled trial. Medical Care, 2015.
(in press)

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a major public health problem in Europe and the U.S, as it is one of the most costly and prevalent chronic conditions in the general population.[1] In 2011, 5% of the Dutch population had diabetes 2, leading to 1,7 billion euro costs for healthcare.[3] Early detection and effective treatment of hyperglycemia leads to a reduction of the burden of diabetes.[4] However, optimal treatment according to existing guidelines is not consistently implemented in clinical practice.[5] Quality-improvement initiatives are needed to improve the quality of treatment in diabetes care.[6]

There are various approaches to improving quality of care. One approach is external assessment, which can use models of peer review, accreditation, and inspection.[7] This study focuses on inspections, which are required under national or regional statutes. The standards for these inspections are derived from regulation and existing guidelines.[7] Inspectorates can use various instruments, such as site visits and performance indicators.[8] The main focus lies on the competence of professional staff, compliance with professional standards, and outcomes for service users.[8] In the Netherlands (EU) healthcare supervision is delegated to the national Dutch Healthcare Inspectorate (later: inspectorate) (Appendix 1).

The inspectorate reported in 2003 that the quality of care for chronically ill patients does not meet the requirements. They concluded that the many care providers involved in care for chronically ill patients did not collaborate sufficiently.[9] In response, the Minister of Health started an action program to develop integrated care for diabetes patients through the creation of care groups.[10] Care groups are organizations that provide integrated diabetes care to patients and consist of several affiliated general practices (Box 1). After implementation of integrated care for diabetes patients, the inspectorate started its supervision program on care groups in 2011.[11] The objective of all the previously mentioned activities was to improve patients' health.[10 11] However, the actual impact of these activities on patient health is unknown.[12]

Therefore we designed a cluster randomized controlled trial (RCT), to provide the strongest possible evidence of the effectiveness of the supervision program. The aim of this study was to evaluate the effect of the government supervision program on the quality of care for care groups providing integrated diabetes care.

METHODS

Design

Several methodological problems can arise when evaluating a supervision program. For example, the time intervals between consecutive activities of the inspectorate are typically short, so that evaluation measurements need to follow shortly after the program, to isolate the effect of one specific intervention. This supervision program is

evaluated because of the good researchability, with respect to randomization and a control group. Therefore we used a cluster randomized controlled trial in which we randomly assigned care groups to the supervision program or to no supervision program. The main outcome measures were structures, processes and health outcomes of care.

Study population

Care groups are organizations that provide integrated diabetes care to patients (Box 1). The supervision program was randomly assigned to 20 of the 100 care groups in the Netherlands, leaving 80 care groups as control groups. For the study we randomly approached 17 intervention groups and 22 control care groups. After exclusion of non-eligible care groups, we were left with 14 intervention care groups and 19 control care groups. Of these, 10 intervention care groups and 8 control care groups participated in this study (response rate of 71% intervention group, 42% control group) (Appendix 2). Each participating care group identified one or two practices for this study, resulting in 16 intervention practices and 15 control practices. From each practice one practice nurse participated, because they provide most care to diabetes patients. This resulted in 31 participating practices nurses in total. Practice nurses in our study have a registered nursing degree or are practice assistants who have followed a two-year, bachelor level practice-nurse education. Selection of patients within care groups was random and anonymous, resulting in between 5 to 10 patients per practice per measurement. We selected only patients who showed up for a check-up up to one month before the data collection. For baseline and post-intervention measurement, we selected different patients using a randomization procedure. Selecting the same patients was not possible because of privacy and practical reasons. Data collection took place in general practices, for the baseline measurement between June and December 2011 and for the post-intervention measurement between May and July 2012. Practice nurses, together with research assistants,

Box 1 *Care groups*

Care groups are organizations that provide integrated diabetes care to patients. Care groups consist of 3 to 250 general practitioners, which are paid with a bundled payment 35. Bundled payment means paying a single fee for all medical services involved in an episode of care 36; in this case, paying a yearly fee for each diabetes patient in the care group.

Care groups consist of multiple health care providers and are often owned by general practitioners. General practitioners and practice nurses provide diabetes care. Practice nurses are mainly involved in performing checkups for diabetes patients. Other care providers are contracted by the care group.

Care groups are comparable with accountable care organizations 24. However, accountable care organizations in the United States have a much broader scope, which includes hospital care. Legal requirements are much more extensive in the U.S. than in the Netherlands. On average, 1.8 general practitioners are working in each general practice in the Netherlands.

selected the patients based on the time schedule and extracted the data from medical records. In addition, the practice nurse was asked to answer questions about guideline adherence in a questionnaire.
Participants were not compensated in this study. The local ethics committee of Erasmus University Medical Centre waived ethical approval for this particular analysis. Written informed consent was obtained from all participating practice nurses.

Evaluated intervention: Program of the Dutch Healthcare Inspectorate

The aim of the supervisory program was to improve diabetes care in care groups, because a previous report showed that care for diabetes patients did not meet the requirements.[9] The supervision program was targeted to directors of care groups and took place in 2011 and 2012. This supervisory program was the first program on care groups, as they are new organizations. Topics of the program were selected based on a risk analysis, including literature research and discussions within the inspectorate and with other organizations. Site visits were the main activity of the supervision. The following topics were selected by the inspectorate and assessed during each site visit: multidisciplinary patients' file, individual care plan, prevention and self-management, continuity of care, quality and transparency.[11 13] Care groups were visited between October and December 2011. The program was assigned to 20 care groups, randomly selected from all 100 care groups. The site visits were announced by mail to the care groups and the topics of the supervision were provided. Care groups were visited for one day by two inspectors. During these visits they talked with directors, general practitioners and practice nurses, to get an overview of how the care group was organized. Besides these conversations, they also checked policy documents of the care group and reviewed few patient files. After the site visits, an individualized report with recommendations for improvements was sent to the board of the care groups, on average within 3 months after the visit. Recommendations were related to the topics of this supervisory program, for example public disclosure of quality of care information. All inspectors hold at least a university master degree and were trained internally as inspectors. Most of them worked as care provider in the past.

Study variables and definitions

The outcome parameter of the study was quality of care, measured with structures and processes of care, and health outcomes. Additionally, patient characteristics were collected for adjustment of the effect in the analysis.
The patient characteristics consisted of demographic factors and clinical factors. Demographic factors were age at data collection, sex, and a socio-economic status (SES) indicator. The SES was based on the neighborhood of the general practice where the patient is treated, because this is mostly

close to the home of the patient. This score was obtained from a government agency (Netherlands Institute for Social Research)[14] and ranges from -10 to +10. A higher score represented a higher SES. All patient characteristics, except SES, were collected from patients' files. Clinical factors were years since diagnosis (between diagnosis and data collection) and comorbidities (defined with International Classification of Primary Care (ICPC) codes).[15] Comorbidities unrelated to T2DM were based on general national guidelines[16] and comorbidities related to T2DM were derived from the National Diabetes Guideline (Appendix 3).[17]

Health outcomes were weight, BMI, systolic blood pressure, diastolic blood pressure, fasting glucose, HbA1c, LDL cholesterol, urine albumin, glomerular filtration rate (GFR) and smoking. Weight (kg), length (cm), BMI (kg/m2), blood pressure (mmHg) and smoking behavior (yes/no) were assessed in general practices. Glucose (mmol/L), HbA1c (mmol/mol), LDL cholesterol (mmol/L), urine albumin (mg/L) and GFR (ml/min) were assessed in cooperation with diagnostic centers. All outcomes were obtained from patients' files. The most recent measurements were used and measurements from before 2009 were not used.

Processes of care were measured at patient level, using information from the patients' medical records, such as yearly measurement of previously mentioned health outcomes. Yearly measurement was defined as a difference between two measurements <366 days. Subsequently, we used the proportion of patients who had yearly measurements per practice for the analysis. In the end all process indicators were added up to compose an aggregated score for quality of care in terms of processes. This score ranged from 0 for lowest quality of care to 11 for

Table 1 *Measures of data collection, divided into structures of care, processes of care and health outcomes*

Structure elements	points	Process elements	points	Outcome elements
Care plan	0-1	Yearly assessment of:		Weight
Lifestyle support	0-1	Weight	0-1	BMI
Medical equipment	0-1	BMI	0-1	Systolic blood pressure
Patients' files	0-1	Blood pressure	0-1	Diastolic blood pressure
Multimorbidity	0-1	Glucose	0-1	Glucose
Self management	0-1	HbA1c	0-1	HbA1c
Tasks and responsibilities	0-1	LDL cholesterol	0-1	LDL cholesterol
Communication and referral	0-1	Albumin in urine	0-1	Albumin in urine
Participation in integrated care	0-1	GFR	0-1	GFR
No show patients	0-1	Smoking behaviour	0-1	Smoking
Patient experience feedback system	0-1	Foot status	0-1	
		Eye status	0-1	
Total structure score	*0 - 11*	*Total process score*	*0 - 11*	

highest quality of care in terms of processes (Table 1).

Structures of care were measured at practice level, by asking practice nurses whether they: use care plans, provide yearly lifestyle support to each patient, have a policy to check medical equipment, regulate of access to patients' files, organize care for multimorbidity patients, are educated in self-management, have a policy for the distribution of tasks, have a policy for communication and referral with other care providers, have agreements about participation in integrated care, have a policy for no-show patients, and systematic patient experience feedback system. All these aspects were related to the recommendations made by the inspectorate. Further, for the structure indicators a total score for quality of care in terms of structures was composed, ranging from 0 for the lowest quality of care to 11 for the highest quality of care in terms of structures (Table 1).

Statistical analysis

From a power calculation we retrieved the minimum of 10 intervention care groups and 10 control care groups, and per care group 2 practices that provided data of 10 patients (400 patients in total).

Baseline patient characteristics and health outcomes were described on patient level. Baseline structures and processes of care were described on practice level, to assess differences between practices.

The effect of this supervision program was evaluated using hierarchical linear and logistic regression models with 2 levels (practices, patients). In hierarchical models, the clustering of patients within practices is taken into account.[18]

The dependent variable was the structure, process or health outcome measure being studied. Independent variables were program (yes/no), and pre- or post-intervention, to assess whether there were general changes over time. This analysis corrected also for baseline differences. Patient characteristics were added as covariates to adjust the program and the time effect for possible differences in patient population. Practice was included in the model as a random intercept, with the exception of analysis of structure measures, where only practice information was used and no random intercept was included.

For each of the study outcomes a separate model was fitted and cases with missing dependent variables were excluded from the analysis. We imputed year since diagnosis, the only covariate with many missing values, with linear regression analysis based on seven covariates (sex, age, related and unrelated comorbidities, SES, structure and process quality score).

From the regression models, betas (in case of a continuous outcome) or odds ratios (in case of a binary or ordinal outcome) and 95% confidence interval were derived.

For regression analyses we used statistical software package SAS version 9.3 (PROC MIXED, PROC GLIMMIX and PROC LOGISTIC) (SAS Institute Inc, Cary, NC) and for other analyses we used SPSS version 21.0 (IBM Inc., Somers, NY).

Table 2 *Baseline patient characteristics, health outcomes and practice characteristics (structure and process), stratified by intervention and control group*

		Intervention			Control		
		n	median	IQR	n	median	IQR
Patient characteristics							
Age (years)		88	68	56-76	88	65	60-73
Sex (% males)		88	49%		88	53%	
Number of related co morbidities [1]		88	0	0-1	88	1	0-1
Number of unrelated co morbidities [2]		88	0	0-1	88	0	0-1
SES indicator [3]	*Low*	88	48%		88	42%	
	Middle		34%			35%	
	High		18%			23%	
Health outcomes							
Weight (kg)		86	83	70-94	87	83	75-94
Length (cm)		84	167	159-177	83	168	161-174
BMI (kg/m²)		84	29	26-33	83	30	28-34
Systolic blood pressure (mmHg)		88	131	123-143	88	136	125-146
Diastolic blood pressure (mmHg)		88	76	70-84	88	79	73-85
Glucose (mmol/L)		85	7.2	6.3-8.4	88	7.5	6.6-8.6
HbA1c (mmol/mol)		88	49	45-56	86	50	46-55
LDL cholesterol (mmol/L)		87	2.5	1.9-2.9	85	2.5	2.0-2.8
Albumin in urine (mg/L)		85	6	3-13	78	6	3-18
GFR (ml/min)		88	67	60-87	83	74	60-90
Smoking (% smokers)		88	21%		86	21%	
Structures of care (care providers measures, mean)							
Care plan		16	0.46	0-0.95	14	0.24	0-0.50
Lifestyle support		16	0.94	0.81-1.00	15	0.97	1.00-1.00
Medical equipment		16	0.78	0.50-1.00	15	0.80	0.50-1.00
Patients' files		16	0.44	0.25-0.50	15	0.42	0.25-0.50
Multimorbidity		16	0.73	0.42-1.00	15	0.76	0.33-100
Self-management		16	1.00	1.00-1.00	15	0.90	1.00-1.00
Tasks and responsibilities		16	0.94	1.00-1.00	15	0.87	1.00-1.00
Communication and referral		16	0.88	1.00-1.00	15	0.93	1.00-1.00
Participation in integrated care		16	0.94	1.00-1.00	15	0.93	1.00-1.00
No show patients		16	1.00	1.00-1.00	15	0.93	1.00-1.00
Systematic patient experience feedback system		16	0.31	0-1.00	15	0.73	0-1.00
Total structures of care score (points)		16	8.4	7.8-9.3	15	8.5	7.6-9.3
Processes of care (patient measures, mean)							
Yearly assessment of							
Weight		16	0.88	0.89-1.00	15	0.85	0.75-1.00
BMI		16	0.61	0.05-1.00	15	0.61	0.30-1.00
Blood pressure		16	0.99	1.00-1.00	15	0.95	1.00-1.00
Glucose		16	0.92	1.00-1.00	15	0.89	0.83-1.00
HbA1c		16	0.94	0.88-1.00	15	0.80	0.67-1.00
LDL cholesterol		16	0.60	0.29-0.95	15	0.48	0.25-0.71
Albumin in urine		16	0.57	0.31-0.80	15	0.38	0.17-0.56
GFR		16	0.71	0.52-0.96	15	0.51	0.25-0.78
Smoking behaviour		16	0.96	1.00-1.00	15	0.88	0.75-1.00
Foot status		16	0.28	0-0.50	15	0.29	0.20-0.44
Eye status		16	0.28	0-0.43	15	0.24	0.17-0.33
Total processes of care score (points)		16	7.7	7.1-8.7	15	6.9	5.9-8.5

[1] in total 11 ICPC codes
[2] in total 62 ICPC codes
[3] Based on postal codes of general practice (low=national lowest tertile, middle=national middle tertile, high=national highest tertile)

RESULTS

In the baseline measurement 176 patients were included, 88 in the intervention group and 88 in the control group (Table 2). Median age was 68 years in the intervention group and 65 years in the control group. In general, patient characteristics and health outcomes at

Table 3 *Crude differences between intervention and control group at baseline and post-intervention and estimated adjusted program and time effects on structures and processes of care, and on health outcomes, analyzed with linear and logistic regression models*

	Baseline		Post-intervention		Time effect		Intervention effect	
	Inter-vention	Control	Inter-vention	Control	OR*	(95% CI)	OR*	(95% CI)
Structures of care								
Care plan	0.46	0.24	0.61	0.39	2.19	(1.21-3.96)	1.15	(0.52-2.54)
Medical equipment	0.78	0.80	0.88	0.87	1.63	(0.82-3.26)	1.30	(0.50-3.39)
Patients' files	0.44	0.42	0.52	0.50	3.05	(1.60-5.82)	0.66	(0.27-1.59)
Multimorbidity	0.73	0.76	0.81	0.89	1.98	(1.03-3.79)	0.78	(0.32-1.87)
Communication and referral	0.88	0.93	0.94	0.87	0.83	(0.33-2.13)	3.28	(0.77-13.96)
Systematic patient experience feedback system	0.31	0.73	0.38	0.80	1.75	(0.88-3.46)	0.63	(0.25-1.61)
Total structure score (β)	*8.4*	*8.5*	*8.9*	*9.3*	*0.88*	*(0.55;1.21)*	*-0.35*	*(-0.82;0.12)*
Processes of care								
Yearly assessment of:								
Weight	0.88	0.85	0.92	1.00	+∞ [1]	(0.28->999)	0.00 [1]	(<0.00-5.85)
BMI	0.61	0.61	0.78	0.73	2.21	(0.91-5.37)	3.11	(0.75-12.86)
Blood pressure	0.99	0.95	0.98	0.96	0.94	(0.17-5.30)	0.56	(0.03-11.68)
Glucose	0.92	0.89	0.90	0.99	7.87	(0.70-89.08)	0.09	(0.01-1.60)
HbA1c	0.94	0.80	0.90	0.93	2.86	(0.87-9.44)	0.15	(0.03-0.89)
LDL cholesterol	0.60	0.48	0.72	0.51	0.83	(0.431-1.61)	2.18	(0.84-5.68)
Albumin in urine	0.57	0.38	0.52	0.50	1.20	(0.63-2.31)	0.67	(0.27-1.66)
GFR	0.71	0.51	0.71	0.62	1.18	(0.61-2.28)	0.95	(0.36-2.48)
Smoking behaviour	0.96	0.88	0.89	0.77	0.37	(0.14-0.99)	0.76	(0.14-4.18)
Foot status	0.28	0.29	0.32	0.28	0.91	(0.44-1.90)	1.31	(0.46-3.70)
Eye status	0.28	0.24	0.28	0.21	0.73	(0.34-1.56)	1.47	(0.53-4.07)
Total process score (β)	*7.7*	*6.9*	*7.9*	*7.5*	*0.26*	*(-0.33;0.85)*	*-0.11*	*(-0.94;0.71)*
Health outcomes								
Weight (β)	83	83	80	82	-1.87	(-6.53;2.80)	2.69	(-3.86;9.24)
BMI (β)	29	30	28	29	-0.66	(-2.20;0.88)	0.53	(-1.58;2.64)
Systolic blood pressure (β)	131	136	135	136	0.43	(-3.91;4.77)	0.65	(-5.41;6.70)
Diastolic blood pressure (β)	76	79	73	80	-0.49	(-2.95;1.98)	-1.96	(-5.40;1.49)
Glucose (β)	7.2	7.5	7.1	7.6	0.22	(-0.34;0.77)	-0.24	(-1.03;0.54)
HbA1c (β)	49	50	49	52	1.64	(-1.12;4.39)	-1.39	(-5.23;2.45)
LDL cholesterol (β)	2.5	2.5	2.5	2.4	-0.001	(-0.23;0.23)	0.045	(-0.28;0.37)
Albumin in urine (β)	6	6	7	6	3.81	(-18.5;26.1)	-3.24	(-34.2;27.8)
GFR (β)	67	74	72	71	-0.01	(-5.40;5.38)	1.40	(-6.06;8.86)
Smoking	21%	21%	15%	19%	1.06	(0.48;2.37)	1.36	(0.44;4.21)

* Odds ratios, unless indicated differently

[1] Analysis not valid, because the results of the second measurement of the control group showed 100% yearly measurement.

Note: All models control for baseline differences between practices, age, sex, years since diagnose, number of related and unrelated comorbidities and social economic status

baseline were comparable between the intervention group and the control group. The percentage of missing health outcomes was low, at most 7%.

The second part of Table 2 shows practice characteristics (i.e., structures and processes of care, respectively) at baseline. Structure scores represent the proportion of practices that positively answered the questions. For example, 46% of the intervention practices reported that they use care plans. The process score should be interpreted as the proportion of patients who were checked annually. The results show for example that 88% of the patients had a yearly weight assessment and IQR of proportions per practice was 0.89 to 1.00. Overall, structures and processes of care were suboptimal in both intervention and control group practices. Most practice characteristics were comparable between the intervention and control group at baseline. The study population was comparable to similar study populations in other research.[19 20]

The third column of Table 3 shows the estimated time effects, which represent the differences between the pre and post-intervention measurement in the total population. The difference in time was on average 250 days between the two measurements and 200 days between the site visit and post-intervention measurement (data not shown). The following structure indicators improved statistically significantly over time: use of care plans (OR=2.2; 95% CI=1.2-4.0), regulation of access to patients' files (OR=3.1; 95% CI=1.6-5.8), and organized care for multimorbidity patients (OR=2.0; CI=1.0-3.8). The total structure score improvement was 0.88 points (95% CI=0.60-1.2), on a scale of 0 to 11, which is an 8% improvement. The process score increased by 0.26 points over time, but this difference was not statistically significant. Likewise, other processes of care and health outcomes did not improve over time.

In the fourth column of the table the program effects are shown, that reflect the difference between the intervention and control groups at the post intervention measurement. The supervision program did not improve structures and processes of care and health outcomes statistically significantly.

DISCUSSION

Summary of main findings

Structures and processes of care did not improve more in the intervention groups than in the control care groups. Moreover, health outcomes did not improve more in the intervention groups than in the control care groups. Although structures of care improved over time in the total population of intervention and control care groups, there were no changes in process of care or health outcomes.

Interpretation

We could not demonstrate an effect of the supervision program on quality of care in care groups providing integrated diabetes care. Three explanations are possible: [1]

there was no effect of the supervision program; [2] the control group improved as well, which limited the contrast between the groups; or [3] the effect of the program was not captured in this study design. Although we have no clear evidence for any one of these possibilities, there are several considerations. The fact that both the intervention group and the control group improved over time is indicative of a spillover effect of the supervision program. It might be that both groups benefited from the supervision.

Dutch diabetes care experts stated that two other developments might have affected the quality of care during this period, namely benchmarking and contracts with insurance companies (personal communication). Benchmarking between care groups started in 2011 and was facilitated by an umbrella organization of care groups[21-23] Additionally, care groups now have to negotiate with insurance companies to gain funding for diabetes treatment.[24] Quality of care plays an important role in the negotiation process with the insurance companies. Quality-improvement initiatives of care groups had already started when the inspectorate actually visited the care groups. Thus, the results of these initiatives could have been harvested during this research period. This could explain why this study found improvements in structures of care over time. However, we have no data to actually proof this hypothesis.

This supervision program can be considered to be a complex intervention.[25] It is not only the site visits that might have an effect on the quality of care, but the announcements of the supervision program to professional organizations and discussion of legal aspects of supervision [26 27] might also have influenced care groups. The supervision was also discussed at conferences.[28 29] The control care groups received no information from the inspectorate; therefore we cannot define the point in time when the control care groups were informed that they would not be inspected. Some of the control care groups probably knew this before the post-intervention measurement, because for example some managers worked for two or more care groups. Furthermore, the nature of the inspectorate is to involve several stakeholders from professional groups to increase the effect of the supervision.[30] Such actions are likely to affect not only the intervention group but the control group as well. This could also explain the improvements found in structures of care over time.

Another possible reason that we did not detect any effect of supervision involves the time span between the intervention and the post-intervention measurement. Improvements in health require a certain period of time to occur. The time between the inspectorate's site visits and the post-intervention measurement was on average 200 days, because of the timing of the supervisory program. In other research, health outcomes improved after 3 years.[31] To address this limitation, we deliberately chose to use intermedi-

ate health outcomes. Nevertheless the time span might have been too short to detect improvements.

Supervision programs can only influence care providers and not directly patients and health outcomes (Appendix 4). This supervision program targeted directors, instead of GPs or practice nurses, which might dilute the effect of this supervision on health outcomes. Furthermore, the treatment enforced by the inspectorate includes non-evidence-based elements.[17] This might also explain why health outcomes did not improve in this study. Nevertheless, the Dutch Healthcare Inspectorate strives to evaluate the effect of its supervision, preferably on health outcomes.

Strengths and limitations

To our knowledge this is the first study that evaluated the effects of a supervision program on health outcomes in an RCT.[12] A control group is almost never available, because the inspectorate wants to reach the highest number of care providers possible. Furthermore, supervision is seldom assigned at random.[32] Randomization is considered unethical,[33] since a lack of supervision increases the risk of patients receiving low quality of care. As care groups are new organizations and no information about quality of care was available at the time of planning the program, random assignment of supervision was accepted.

A limitation of the study is the low response rate, which might have caused selection bias. Non-response was more common in the control group; more care groups in the control group did not understand the purpose of the supervision program and therefore did not want to participate in our study. However, intervention and control care groups are comparable with respect to quality of care and study population and therefore we have no indications that selection bias took place. Care groups identified participating practices. It is possible that they selected the practices that they thought were providing the best quality of care. This might have resulted in an overestimation of the quality of the care provided. However, as we expect this mechanism to be the same in the intervention and control groups, we assume that our results are not biased. In addition, only patients who recently showed up for a check-up were included. Similarly this might have resulted in an overestimation of the quality of care, but not in a biased comparison between the intervention and control group.

Fewer practices than expected agreed to participate in this study, which resulted in fewer patients being included. Therefore, this study was somewhat underpowered. However, since the point estimates do not suggest any trend, we assume that our negative results are not solely caused by insufficient power.

A strength of our study is the use of outcomes on all levels of quality of care, namely structure, process and health outcomes of care.

Implications and general conclusion

The enforcement through supervision of evidence-based health

interventions should be encouraged, as this will increase the potential improvements in health outcomes. Since the inspectorate aims at promoting population health, the effects on population health should be considered in advance, instead of focusing on possible risks. In addition, the supervision program should target the care process itself and not the board of a care group or its directors. Other possibilities that can be explored to improve supervision are to improve the competency of the inspectors and to select supervision topics bottom-up instead of top-down.

Although an RCT is the optimal design for studies aiming to gain more insight into the effectiveness and efficiency of supervision [34], the current manner of conducting supervision makes it challenging to use the this design. Further research should generate more evidence on the effectiveness of supervision, if possible using RCTs, but a stepped wedge design could also be considered.[32] Evidence on the effects of supervision may help inspectorates in decision making and to further improve their working methods and may provide society with information on how successful supervision programs have been.[12]

In summary, in this cluster RCT we could not demonstrate improvements in quality of integrated diabetes care resulting from the supervision program. Although structures of care did improve over time, other quality-improvement initiatives are necessary to substantially strengthen integrated care for diabetes patients.

REFERENCES

1. Cowie CC, Rust KF, Byrd-Holt DD, et al. Prevalence of diabetes and high risk for diabetes using A1C criteria in the US population in 1988–2006. *Diabetes Care* 2010;33(3):562-68.
2. Baan CA, Poos MJJC. How common is diabetes mellitus and how many people die from it? (Hoe vaak komt diabetes mellitus voor en hoeveel mensen sterven eraan?), 2013.
3. National Institute for Public Health (Rijksinsituut voor Volksgezondheid en Milieu). Cost of Illness in the Netherlands (Kosten van Ziekten), 2013.
4. Turner RC, Holman RR, Cull CA, et al. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998;352(9131):837-53.
5. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, et al. A longitudinal study examining adherence to guidelines in diabetes care according to different definitions of adequacy and timeliness. *PloS one* 2011;6(9): e24278.
6. Stone MA, Charpentier G, Doggen K, et al. Quality of Care of People With Type 2 Diabetes in Eight European Countries Findings from the Guideline Adherence to Enhance Care (GUIDANCE) study. *Diabetes Care* 2013.
7. Shaw C. External assessment of health care. BMJ 2001;322(7290):851.
8. Boyne G, Day P, Walker R. The evaluation of public service inspection: A theoretical framework. *Urban Studies* 2002;39(7):1197-212.
9. Dutch Health Care Inspectorate (Inspectie voor de Gezondheidszorg). State of Health Care, Integrated care chronically ill (Staat van de Gezondheidszorg 2003, Ketenzorg bij chronisch zieken). Den Haag: Ministry of Health, Welfare and Sport, 2003.
10. Taskforce Diabetes Program (Taakgroep Programma Diabeteszorg). Diabetes care better (Diabeteszorg beter), 2005.

11. Dutch Health Care Inspectorate (Inspectie voor de Gezondheidszorg). Implementation care standard Diabetes after four years of insufficient progress (Implementatie zorgstandaard Diabetes na vier jaar onvoldoende gevorderd). Utrecht: Ministry of Health, Welfare and Sport, 2012.
12. Health Council of the Netherlands (Gezondheidsraad). Towards evidence-based supervision. Research on effects of supervision by the Dutch Health Care Inspectorate (Op weg naar evidence based toezicht. Het onderzoek naar effecten van toezicht door de Inspectie voor de Gezondheidszorg). Den Haag, 2011.
13. Ketelaars CAJ. Integrated care requires integrated supervision. *Int J Integr Care* 2011;11.
14. Netherlands Institute for Social Research (Sociaal en Cultureel Planbureau). Status scores.
15. Wood M, Lamberts H. International Classification of primary care: prepared for the World Organisation of national colleges, academies and academic associations of general practitioners/family physicians (WONCA) by the ICPC Working Party: Oxford: Oxford University Press, 1987.
16. Hoeymans N, Schellevis FC. Selection of chronic diseases (Selectie van chronische ziekten), 2008.
17. Dutch Diabetes Federation (Nederlandse Diabetes Federatie). NDF *Care Standard* (NDF *Zorgstandaard*). Amersfoort 2007.
18. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*: Springer Verlag, 2009.
19. Van Hateren KJJ, Drion I, Kleefstra N, et al. A prospective observational study of quality of diabetes care in a shared care setting: trends and age differences (ZODIAC-19). BMJ *Open* 2012;2(4).
20. Sidorenkov G, Voorham J, de Zeeuw D, et al. Treatment quality indicators predict short-term outcomes in patients with diabetes: a prospective cohort study using the GIANTT database. BMJ *Qual Saf* 2013;22(4):339-47.
21. Advisory council integrated care (Adviesgroep Ketenzorg), LVG. Transparency Integrated Care Diabetes Mellitus, reporting care groups in 2010 (Transparantie Ketenzorg Diabetes Mellitus Rapportage zorggroepen 2010), 2012.
22. National Organization for Integrated Care (Landelijke Organisatie Ketenzorg). Transparency Integrated Care Diabetes Mellitus and COPD, reporting care groups in 2011 (Transparante Ketenzorg Diabetes Mellitus en COPD, rapportage zorggroepen 2011), 2013.
23. National Organization for Integrated Care (Landelijke Organisatie Ketenzorg). Transparency Integrated Care Diabetes Mellitus, COPD and vascular risk management, reporting care groups in 2012 (Transparante ketenzorg Diabetes Mellitus, COPD en VRM, rapportage zorggroepen over 2012), 2013.
24. Struijs JN, Baan CA. Integrating care through bundled payments—lessons from The Netherlands. N *Engl J Med* 2011;364(11):990-91.
25. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: The new Medical Research Council guidance. *Int J Nurs Stud* 2013;50(5):587-92.
26. Kien NUN, Maas MM. Is theme based supervision on care groups by the Inspectorate necessary? (Is thematisch toezicht door Inspectie onder zorggroepen noodzakelijk?). *De Eerstelijns* 2011;3(9):18-19.
27. Ketelaars C. The Care Institutions Quality Act: focus on quality (Kwaliteitswet Zorginstellingen: sturen op kwaliteit). *De Eerstelijns* 2011;3(10):11.
28. Ketelaars CAJ, Scheijven HLG. Complying to the standard of integrated diabetic care. *Int J Integr Care* 2012;12(Suppl3).
29. Ketelaars C. Risk analysis on integrated care policy. *Int J Integr Care* 2011;11(Suppl).
30. Robben P. *Supervision in a glass house: The effectiviness of supervision on the quality of health care* (*Toezicht in een glazen huis: Effectiviteit van het toezicht op de kwaliteit van de gezondheidszorg*), 2010.
31. Struijs JN, de Jong-van Til JT, Lemmens LC, et al. Three years of bundled payment for diabetes care in the Netherlands. Bilthoven (NL): National Institute for Public Health and the Environment (RIVM), 2012.
32. Van Dishoeck AM, Oude Wesselink SF, Lingsma HF SE, et al. Transparency: can the effect of governmental surveillance be quantified? (Transparantie: is het effect van toezicht te meten?). *Ned Tijdschr Geneeskd* 2013;157(16).
33. Edwards SJL, Braunholtz DA, Lilford RJ, et al. Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 1999;318(7195):1407.
34. Robben P. Evaluation and improvement of supervision (Evalueren van verbeteren van toezicht). *Tijdschrift voor Toezicht* 2011;2(2):72-74.
35. De Jong-van Til JT, Lemmens LC, Baan CA, et al. The organization of care groups in 2011 (De organisatie van zorggroepen anno 2011). Bilthoven (NL): National Institute for Public Health and the Environment (RIVM), 2012.
36. De Bakker DH, Struijs JN, Baan CA, et al. Early results from adoption of bundled payment for diabetes care in the Netherlands show improvement in care coordination. *Health Affairs* 2012;31(2):426-33.

APPENDIX 1

Dutch Healthcare Inspectorate

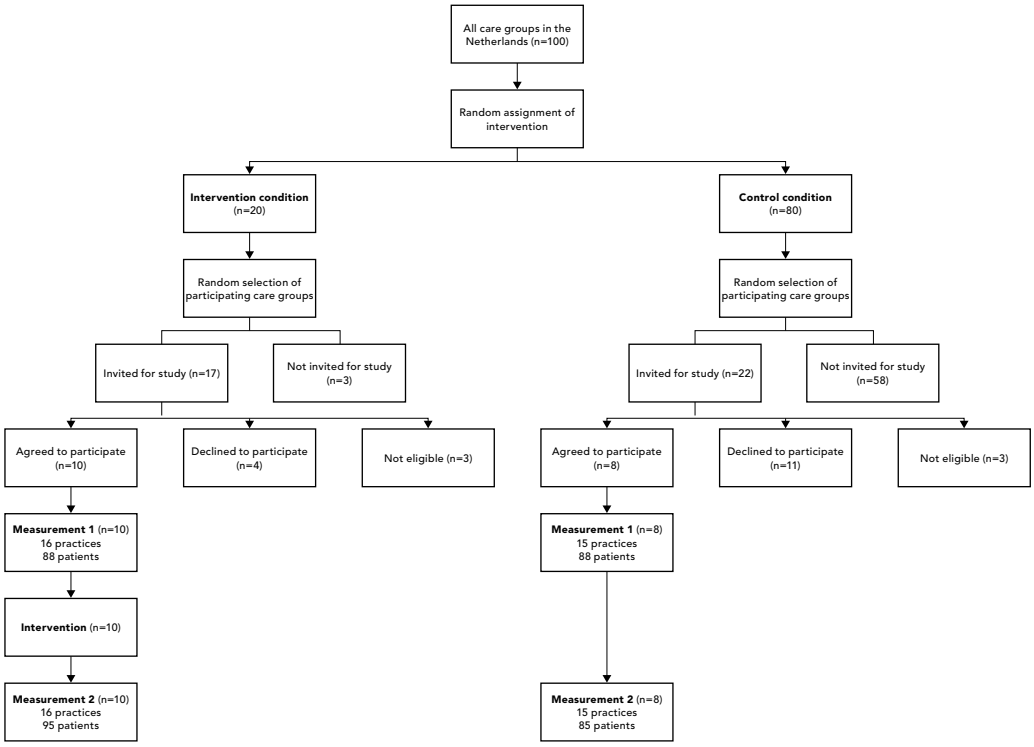
The Dutch Healthcare Inspectorate (Netherlands, EU) is an independent agency of the Ministry of Health, Welfare, and Sport. It aims to improve population health and is expected to efficiently supervise a sector where 1.3 million people work for 40,000 institutions and companies. Its primary instruments are advice and encouragement. If these do not achieve the desired result, it can implement corrective action by, for example, increasing the supervision or by limiting the ability to practice a profession. The inspectorate enforces 25 laws, including for example the Care Institutions Quality Act. The supervision is performed by using a combination of three methods:

1. theme-based regulation, directed at specific issues in care, which are sometimes requested by the minister or parliament
2. regulation in response to calamities or emergencies that indicate structural shortcomings in care provision
3. risk-based supervision to assess the quality of healthcare by means of indicators.

The program described in this study is an example of theme-based regulation.

APPENDIX 2

Flow chart of program and study participants



Note. Reasons for refusal were the following: too busy with providing care to patients (n=4), no compensation for time loss due to research (n=2), do not agree with purpose of research (n=1), currently involved in other research (n=1) and unknown (n=7)

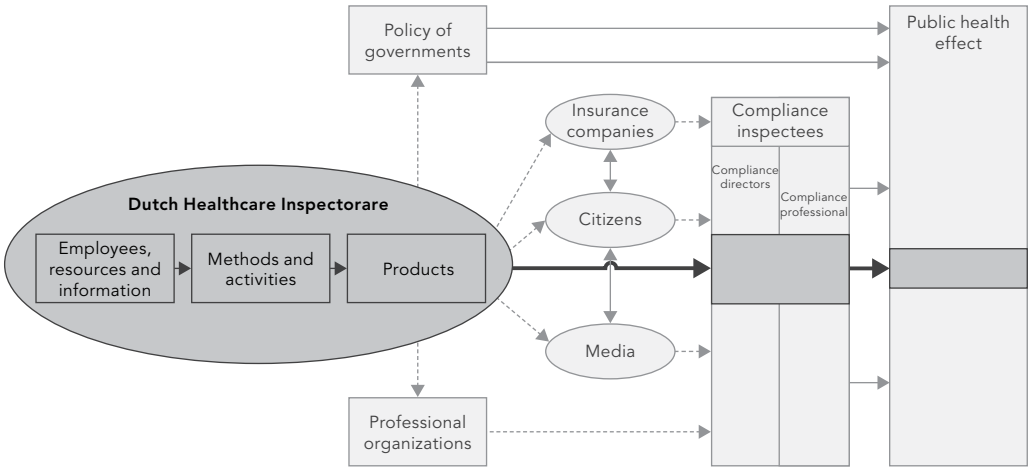
APPENDIX 3

List of related and unrelated comorbidities

Related comorbidities	ICPC code
Angina pectoris	K74
Acute myocardial infarction	K75
Hypertension	K86 and K87
Transient ischemic attack	K89
Stroke	K90
Intermittent claudication	K92
Aneurysm aortae	K99
Diabetic neuropathy	N92
Depression	P03 and P76
Unrelated comorbidities	
Tuberculosis	A70
HIV/AIDS	B90
Cancer	A79, B72, B73, D74, D75, D76, D77, L71, N74, R84, R85, S77, T71, U75, U76, U77, W72, X75, X76, X77, Y77, Y78
Peptic ulcer	D85, D86
Ulcerative colitis	D94
Visual disturbance	F83, F84, F92, F93, F94
Hearing impairment	H84, H85, H85
Congenital heart defect	K73
Heart failure	K77
Chronic neck and back problems	L83, L84, L85, L86
Rheumatoid arthritis	L88
Osteoarthritis	L89, L90, L91
Osteoporosis	L95
Congenital neurological disorder	N85
Multiple sclerosis	N86
Parkinson's disease	N87
Epilepsy	N88
Chronic alcohol abuse	P15
Dementia	P70
Schizophrenia	P72
Anxiety disorder, other neurosis, PTSS	P74, P79
Anorexia nervosa	T06
Mental retardation	P85
COPD	R91, R95
Asthma	R96
Eczema	S87, S88

APPENDIX 4

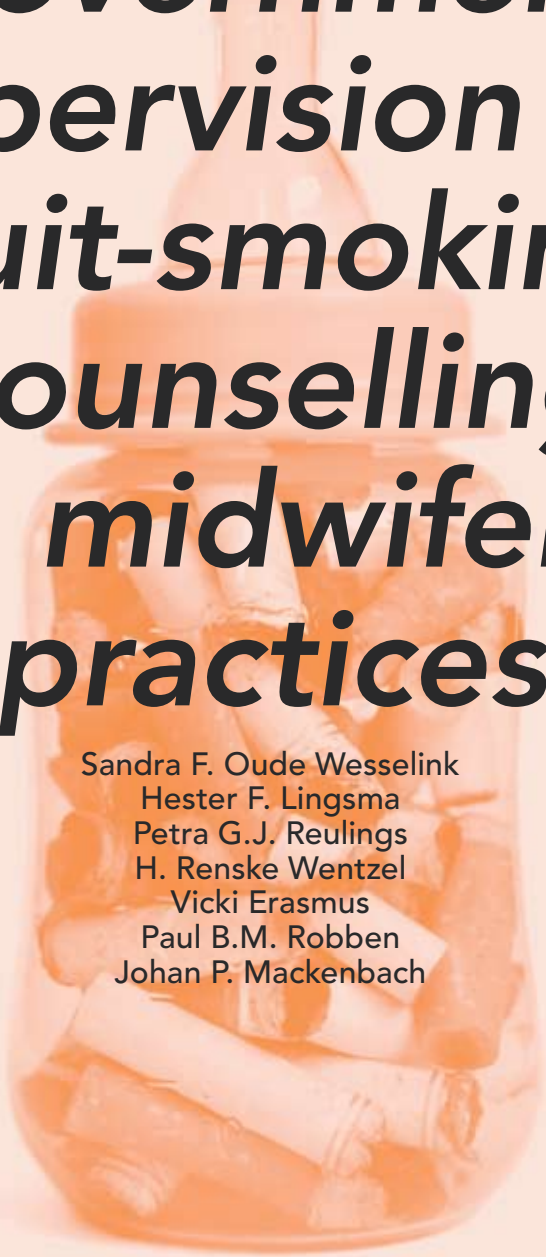
Effect of supervision in steps, adapted from Dutch Healthcare Inspectorate



CHAPTER 3

Effects of government supervision on quit-smoking counselling in midwifery practices

Sandra F. Oude Wesselink
Hester F. Lingsma
Petra G.J. Reulings
H. Renske Wentzel
Vicki Erasmus
Paul B.M. Robben
Johan P. Mackenbach



ABSTRACT

Introduction

Smoking-cessation counselling during pregnancy is important to prevent smoking-related harm in pregnant smokers and their children. Therefore we evaluated the effects of an Inspectorate's supervision programme on the provision of smoking-cessation counselling by midwifery practices in the Netherlands. The supervision programme consisted of 3 elements: A) A deadline was announced by which all practices should comply with professional norms on such counselling (2011); B) A set of randomly selected practices were assessed using a questionnaire and a personal feedback report (2010); C) Another set of randomly selected practices were assessed through a site visit and a personal feedback report (2012).

Methods

Programme A was evaluated in a before-after study, Programmes B and C were evaluated in a randomized controlled trial (RCT) with only a post-intervention measurement. Primary outcome was provision of smoking-cessation counselling through a minimal-intervention strategy (V-MIS). Linear and logistic regression models were used to analyse data from 233 primary-care midwifery practices.

Results

A) After announcement of the deadline, Dutch midwifery practices reported significantly more provision of smoking-cessation counselling. For example, the use of V-MIS increased substantially from 28% to 80%; B) In practices that were assessed with a questionnaire, the provision of counselling improved partially compared to controls; C) The provision of counselling did not differ between practices that were visited and their controls. While the training participation rate in counselling by midwifery practices did not differ between the intervention and control groups, the rate increased significantly in all practices after the start of the supervision programme.

Conclusions

The provision of smoking-cessation counselling improved spectacularly in Dutch midwifery practices. Despite some limitations of our study, the Inspectorate's supervision programme is likely to have contributed to the improvements in provision of counselling.

Adapted from: Oude Wesselink SF, Lingsma HF, Reulings PGJ, Wentzel HR, Erasmus V, Robben PBM, Mackenbach JP. Does Government Supervision Improve Stop-Smoking Counseling in Midwifery Practices? Nicotine & Tobacco Research, 2015.
http://ntr.oxfordjournals.org/content/17/5/572

INTRODUCTION

Six percent of women in the Netherlands smoke during pregnancy.[1] Among lower educated women, the prevalence of smoking during pregnancy is around 14%. Maternal smoking is associated with a higher risk of foetal mortality and of adverse birth outcomes such as stillbirth, preterm birth, small for gestational age, intrauterine growth restriction, and congenital heart defects.[2] As perinatal mortality was higher in the Netherlands than in other European countries, there was room for improvement in perinatal care.[3]

Midwives can play a key role in provision of smoking-cessation counselling of pregnant smokers: in the Netherlands primary care midwives see 80% of all pregnant women at an early stage of pregnancy in midwifery practices. However, research showed that the quality of counselling is low.[4] This situation should thus be improved.

There are various approaches to improving quality of care. One is external assessment, which can use models of peer review, accreditation, and inspection.[5] This study focuses on the inspections, required under national or regional statutes, where standards are derived from regulation and existing guidelines.[5] Inspectorates can use various instruments, such as site visits and performance indicators.[6] The main focus lies on the competence of professional staff, compliance with professional standards, and outcomes for service users.[6] In the Netherlands, healthcare supervision is delegated to the national Dutch Healthcare Inspectorate (later: Inspectorate) (Supplementary Box S1).

In 2010, the Inspectorate started a supervision programme on primary care midwives who provide care to pregnant smokers. It focused on the use of V-MIS, an evidence-based intervention for provision of smoking-cessation counselling.[7] V-MIS is a minimal-intervention strategy that targets midwifery practices and is based on the stages of change theory.[8] In the 2010–2012 period, the Inspectorate promoted the use of V-MIS by a supervision programme, intended to improve the quality of counselling and reduce smoking rates during pregnancy. However, the impact of this supervision programme has not yet been evaluated.

When evaluating a supervision programme in a real-life setting, several methodological problems can arise.[9–11] Often a control group is lacking, for example, in case of national coverage of the supervision program. This was the case here. If a control group is available, a pre-intervention measurement can be unfeasible, because the measurement process itself may change people's behaviour. This effect, also called the Hawthorne effect, occurs in people who are aware of being observed.[12] People who are aware of being observed will adapt their behaviour. If researchers ask about specific aspects of quality of care for research purposes, then that will raise awareness about quality of care, and care providers might change their behaviour before the programme starts. These problems

should be considered when evaluating supervision programmes.
To provide the strongest possible evidence to answer our research question, we tried to overcome these problems by combining various study designs. We wished to establish the effect of the government supervision programme on the quality and quantity of the provision of smoking-cessation counselling, specifically the use of V-MIS, by midwifery practices.

METHODS

Evaluated Intervention: Programme of the Dutch Healthcare Inspectorate

The aim of the programme of the Dutch Healthcare Inspectorate was to improve the provision of smoking-cessation counselling to pregnant women by primary care midwives. The supervision programme consisted of three elements: announcement of a deadline (Programme A), assessments with questionnaires and personal report (Programme B) and assessments with site visits and personal report (Programme C).
In 2010, the Inspectorate randomly distributed a questionnaire to a sample of midwifery practices (113 of 500) (Programme B). Goal of the questionnaire was to gain insight in current state of provision of counselling in midwifery practices. Results of the questionnaire and points for improvements were sent to practices in a personal report. Self-reported data highly agrees with on-site inspections, as shown in previous research.[13]
From aggregated results of the questionnaire, the Inspectorate concluded that large improvements in provision of counselling were needed. The Inspectorate started a multifaceted inspection to improve counselling in practices (Programme A). The results of the questionnaire were published in an article in the professional journal of midwives in the Netherlands.[14] Furthermore, the Inspectorate stated that professional norms should be followed. A deadline was announced, in consultation with the Dutch professional midwifery organization, when all practices were expected to comply with these professional norms, which was January 1, 2012. All practices received an enforcement letter to inform them about the professional norms and date from which these norms would be enforced. Furthermore, the Inspectorate presented the norms and deadline at a meeting with regional midwives representatives. Preconditions for smoking cessation counselling received special attention. Preconditions included education of midwives, smoking-cessation counselling policy in the practice, insight in support possibilities from other care providers and registration of smoking behaviour of all pregnant women. The Dutch expertise centre on tobacco control (later: expertise centre) and the professional midwifery organization made the preconditions available for practices. Finally the Inspectorate announced site visits to practices to verify whether professional norms were implemented.
After the deadline, the Inspectorate

inspected a sample (21 of 500) of all midwifery practices with site visits (Programme C). During these visits, inspectors checked whether midwives complied with aforementioned guidelines. This was done by inspecting policy documents, training certificates and registration forms, and by assessing use of V-MIS and assessing midwives' knowledge of where they could refer women for support to stop smoking. Site visits were announced by mail to practices and topics of supervision were provided. Two inspectors inspected the practices in half a day. Smoking-cessation counselling required 10% of the time of the site visit, rest of the time was used to address other topics regarding quality of midwifery care. After the site visits, all visited practices received a personal report with feedback on their counselling and the timeframe by when improvement should be implemented (Supplementary Figure S1).

Design

We used three designs to separately evaluate the three elements of the supervision programme. Due to national policy, the design evaluating Programme A (the deadline) was a before-after study. All practices were exposed to the programme and measured before and after the programme. The design evaluating Programmes B (the questionnaire) and C (the site visit) was a randomized controlled trial (RCT) without pre-intervention measurements. Pre-intervention measurements were unfeasible in the evaluation of Programmes B and C, since these measurements would raise awareness about smoking-cessation counselling before the programme actually started. Finally, we evaluated all programmes in one time series design focusing on education of midwives.

Study Population

Evaluation of Programme A consisted of 113 practices (measurement I and II) (Supplementary Figures S1 and S2). Evaluation of Programme B consisted of 113 intervention midwifery practices (the same group as in Programme A), which were randomly chosen in 2010, and supplemented with a random sample of 100 control practices (measurement II) (selected in 2012). In evaluation of Programme C, the intervention and control group were selected with a stratified randomization, which resulted in 14 intervention practices and 38 control practices (measurement III and IV) (selected in 2012). We included intervention practices that received either questionnaires or site visits, excluding practices who received both. Therefore we could only include 14 visited practices instead of 21, as seven practices that received both excluded. Several other practices needed to be excluded because they underwent large organizational changes during the study period (e.g., a large part of the staff and owner changed). The time series evaluation included all registered midwifery practices.

Data Collection

Data were collected with online study questionnaires. Practices

were invited by mail, followed by an e-mail with a link to the online study questionnaire. From each practice one midwife was requested to fill in the study questionnaire, preferably the midwife with expertise on smoking cessation. Measurement I took place in June and July 2010. Measurement II and III took place in 2012 between July and September and measurement IV took place in March and April 2013. Education data was obtained from the education centre responsible for training midwives on smoking-cessation counselling.

Participants received a small compensation in the form of a voucher of 10 euros when the entire study questionnaire was completed. The local ethics committee of Erasmus University Medical Centre waived ethical approval for this particular analysis. An electronic informed consent was obtained from midwives who filled in the study questionnaire.

Study Variables and Definitions

The primary outcome parameter was quality and quantity of smoking-cessation counselling, defined as the use of V-MIS. First was asked whether V-MIS was used and then all steps of the V-MIS were asked separately. The steps included assessing smoking behaviour, registration and counselling smoking partner, assessing motivation to quit, enhancing motivation to quit, discussing barriers to quit, reducing barriers to quit, setting a quit date, counselling of quit attempts and aftercare quit attempts. Definition of each step included questions on which actions were performed and to which proportion of eligible pregnant smokers the step was applied.

Additionally, we assessed preconditions including education of midwives in smoking-cessation counselling, policy for smoking-cessation counselling in the practice, support possibilities from other care providers and registration of smoking behaviour of pregnant smoker. A total score was derived from the average of all measures per practice. For the education time series, we determined the date on which the first midwife of each practice followed the smoking-cessation education. This education was a half day course, designed for midwives and organized by a national expertise centre.

Practice characteristics were collected for adjustment in the analyses, including number of midwives, number of practice assistants, number of pregnant women (per year), percentage of pregnant women with low socio-economic status (SES) and SES indicator of the practice. Percentage of low SES was defined by the Dutch payment construction for midwifery practices, in which practices receive more compensation for pregnant women who live in deprived neighbourhoods. The SES indicator of the practice was based on the neighbourhood of the practice. This score was obtained from a government agency (Netherlands Institute for Social Research [Sociaal en Cultureel Planbureau])[15] and ranges from −10 to +10, where higher scores represent higher SES. All practice characteristics, except

SES indicator of the practice, were self-reported.

Statistical Analysis

Practice characteristics and quality of the counselling were described on practice level. The effect of supervision was evaluated with linear and logistic regression models.
Dependent variable was the step of smoking-cessation counselling under study. Independent variable for evaluation of Programme A (the deadline) was pre- or post-intervention. Independent variable for evaluation of Programme B (the questionnaire) was programme (yes/no). Independent variables for evaluation of Programme C (the site visit) were programme (yes/no), and measurement (post intervention measurement 1 or post intervention measurement 2). All practices were measured twice and therefore analysed with hierarchical models with practice as random intercept.
In all analyses we adjusted for practice characteristics, to take possible changes in practices and differences between groups into account.

Table 1 *Practice Characteristics Stratified by Time of Measurement or Intervention and Control Group, Separated for the Three Programme Elements*

Evaluation A (Deadline) (measurement I and II)	**2010** (response rate: 94%)				**2012** (response rate: 75%)			
	n	**median**	**IQR**	**mean**	**n**	**median**	**IQR**	**mean**
Number of midwives	113	4	2-4.5	4	71	4	3-5	4
Number of practice assistants	NA				71	1	0-2	1
Number of pregnancies (per year)	113	303	200-500	361	71	322	214-497	269
Smokers	NA				52	12%	6-21%	15%
Pregnancies with low SES	108	0%	0-0%	4%	71	0%	0-2%	4%
SES indicator of practice	113	Low: 35% Middle: 30% High: 35%			71	Low: 39% Middle: 30% High: 31%		
Evaluation B (Questionnaire) (measurement III)	**Intervention** (response rate: 75%)				**Control** (response rate: 76%)			
	n	**median**	**IQR**	**mean**	**n**	**median**	**IQR**	**mean**
Number of midwives	71	4	3-5	4	71	3	2-4	4
Number of practice assistants	71	1	0-2	1	71	1	0-2	1
Number of pregnancies (per year)	71	322	214-497	269	70	316	189-435	353
Smokers	52	12%	6-21%	15%	56	13%	7-19%	14%
Pregnancies with low SES	60	0%	0-2%	4%	64	0%	0-1%	5%
SES indicator of practice	71	Low: 39% Middle: 30% High: 31%			71	Low: 42% Middle: 28% High: 30%		
Evaluation C (Site visit (measurement IV)	**Intervention** (response rate: 79%)				**Control** (response rate: 92%)			
	n	**median**	**IQR**	**mean**	**n**	**median**	**IQR**	**mean**
Number of midwives	11	4	3-6	5	35	4	3-6	5
Number of practice assistants	11	1	0-2	1	35	1	0-2	1
Number of pregnancies (per year)	11	350	269-402	409	35	302	230-600	402
Smokers	9	20%	8-28%	20%	29	12%	8-18%	13%
Pregnancies with low SES	10	0%	0-5%	4%	30	0%	0-2%	4%
SES indicator of practice	11	Low: 36% Middle: 36% High: 27%			35	Low: 43% Middle: 37% High: 20%		

Note. IQR = interquartile range; SES = socio-economic status

For each study outcome, a separate model was fitted and cases with missing dependent variables were excluded in the analysis. The only covariate with many missing values, number of practice assistants, was imputed with linear regression analysis based on three covariates (number of midwives, number of pregnant women, and percentage of pregnant women with low SES).

From the regression models beta's (in case of a continuous outcome) or odds ratios (OR, in case of a binary or ordinal outcome) and 95% confidence intervals were derived.

Education of the first midwife of each practice was visualized with Kaplan-Meier curves and analysed with Log Rank test. Trends were studied with analysis of rates.

For regression analyses we used statistical software package SAS version 9.3 (PROC LOGISTIC, PROC GLM, PROC MIXED, PROC GLIMMIX) (SAS Institute Inc.,) and for other analyses SPSS version 21.0 (IBM Inc.,).

RESULTS

Population

In the total study population the interquartile range (IQR) of the number of midwives per practice was 2–5 and the number of practice assistants was 0–2. The IQR of the number of pregnancies per practice per year was 204–460 and the IQR of the percentage of smokers per practice was 1%–17%. The IQR of percentage of pregnant women in the practices that had a low SES was 0%–1% (data not shown). Practices were equally distributed in low, middle, and high SES neighbourhoods. In general, practice characteristics were comparable between the groups in all evaluations (Table 1).

The Inspectorate's supervision programme consisted of a set of elements, which were evaluated separately (Supplementary Table S1). The response rates for evaluation of Programme A (the deadline) was 84% and 75%, Programme B (the questionnaire) was 78% and 76%, and Programme C (the site visit) was 79% and 92% (for details see Supplementary Figures S1 and S2).

Programme A: Announcement of Deadline

Announcement of the deadline (Programme A) was evaluated in a before and after study. Before the programme started, 28% of the practices reported using V-MIS in measurement I and after the programme, this increased substantially to 80% in measurement II. Total score of all steps was 73% before the programme and 85% after the programme. In the adjusted regression models, almost all steps of counselling improved after the announcement of the deadline (Table 2). Use of V-MIS (OR = 12.6, 95% confidence interval [95% CI] = 5.9–26.8), smoking policy (OR = 8.4, 95% CI = 4.4–16.0) and setting a quit date (β = 0.42, 95% CI = 0.32–0.53) improved the most.

Programme B: Questionnaire and Personal Report

We evaluated the effect of assessments with questionnaire and personal report (Programme B) on

counselling in an RCT without pre-intervention measurement. In measurement II, 80% of the intervention group reported to use V-MIS compared to 71% of the control group. Total score of all steps was 85% in the intervention group and 81% in the control group. Some outcomes improved statistically significantly with 4% in the intervention group compared to the control group in the adjusted regression analyses: enhancing motivation to quit (95% CI = 0.00–0.08), reducing barriers to quit (95% CI = 0.01–0.08) and the total score (95% CI = 0.00–0.08). Setting a quit date increased with 17%, compared to the control group (95% CI = 0.04–0.30).

Programme C: Site Visits and Personal Report

Assessments with site visit and a personal report (Programme C) was evaluated in an RCT without pre-intervention measurement. In measurement III, 80% of the intervention practices reported to use V-MIS compared to 71% of the control practices. The intervention group scored 86% on the total score and the control group 84%. There were no differences in reported provision of counselling between the groups

Table 2 *Estimated Adjusted Effects of Supervision Programme on Smoking-Cessation Counseling, Separated for the Three Programme Elements*

	A Deadline (95% CI) (measurement I and II)		**B Questionnaire (95% CI)** (measurement III)		**C Site visit (95% CI)** (measurement IV and V)	
Use V-MIS intervention (OR)	12.6	*(5.9-26.8)*	1.7	*(0.7-4.0)*	1.3	*(0.1-21.0)*
Education of midwives (OR)	5.2	*(2.8-9.5)*	1.2	*(0.6-2.4)*	1.6	*(0.3-1.8)*
Smoking policy (OR)	8.4	*(4.4-16.0)*	1.6	*(0.8-3.1)*	0.2	*(0.0-1.2)*
Support from other care providers (OR)	2.0	*(1.1-3.9)*	1.6	*(0.7-3.6)*	1.6	*(0.4-5.8)*
Registration smoking behaviour (OR)	2.4	*(0.9-5.9)*	2.9	*(0.9-8.9)*	0.9	*(0.2-4.1)*
Registration and counselling smoking partner (OR)	4.9	*(2.2-11.2)*	1.2	*(0.4-3.9)*	2.2	*(0.3-18.0)*
Aftercare quit attempt (OR)	0.9	*(0.37-2.1)*	0.6	*(0.2-1.6)*	1.2	*(0.1-12.1)*
Assessing smoking behaviour (*β*)	0.02	*(-0.01;0.06)*	0	*(-0.01;0.03)*	0.03	*(-0.06;0.11)*
Assessing motivation to quit (*β*)	0.06	*(0.01-0.12)*	0.02	*(-0.07;0.03)*	0.05	*(-0.03;0.14)*
Enhancing motivation to quit (*β*)	0.04	*(0.00-0.07)*	0.04	*(0.00-0.08)*	0.02	*(-0.05;0.08)*
Discussing barriers to quit (*β*)	0.16	*(0.08-0.23)*	0.05	*(-0.03;0.13)*	0.03	*(-0.07;0.13)*
Reducing barriers to quit (*β*)	0.03	*(0.01-0.06)*	0.04	*(0.01-0.08)*	0.03	*(-0.04;0.10)*
Setting a quit date (*β*)	0.42	*(0.32-0.53)*	0.17	*(0.04-0.30)*	0.05	*(-0.19;0.29)*
Counselling of quit attempt (*β*)	0.03	*(-0.03;0.08)*	0.01	*(-0.06;0.04)*	0.01	*(-0.05;0.07)*
Total score (*β*)	0.12	*(0.08-0.16)*	0.04	*(0.00-0.08)*	0.02	*(-0.05;0.09)*

Note. All models control for the number of midwives, number of practice assistants, number of pregnant women, percentage of pregnant women with low SES and SES indicator of the practice. Evaluation of programme A compares the performance of practices after announcement of the deadline with the performance before the deadline was announced. Evaluation of programme B compares the group of practices that was assessed with questionnaire and personal report with the group that was not assessed with questionnaire and personal report. Evaluation of programme C compares the group of practices that was assessed with a site visit and personal report with the group that was not assessed with a site visit and personal report. CI = confidence interval; OR = Odds Ratio.

in the adjusted regression analyses. Visited practices mostly scored somewhat higher on steps of counselling, but none of the differences was statistically significant.

V-MIS Training

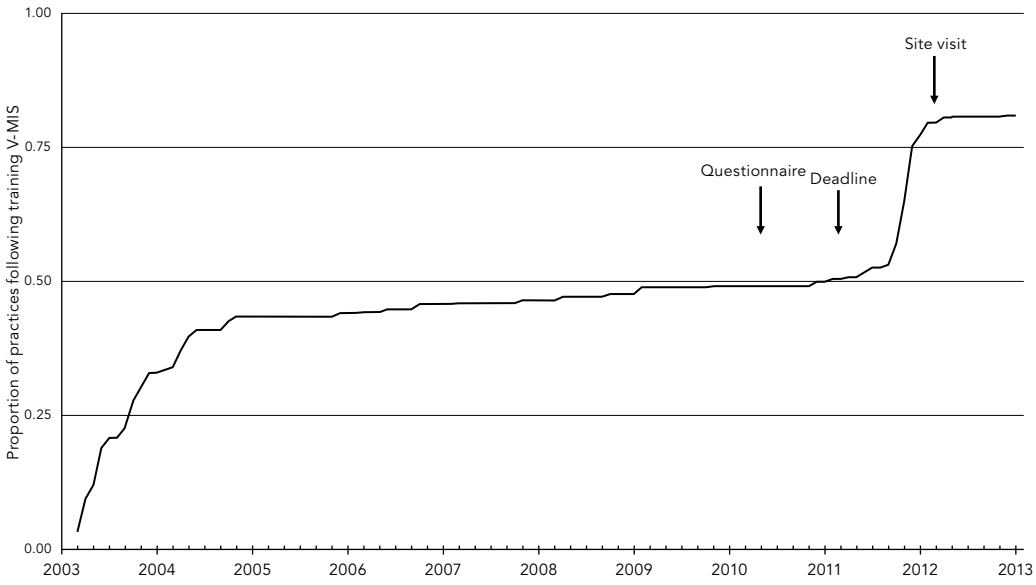
For the V-MIS training, we separated different time periods and groups (Figure 1). Regarding the programme period from 2010 to 2012, the intervention practices followed almost the same pattern as control group practices (p = .62). In general, we distinguish two time periods of rapid increase in practices that received V-MIS training. The first time period was from start of the V-MIS training from 2003 to 2005 and the second period was from 2011 onwards, during the Inspectorate's programme. The rate from 2006 to 2011 was significant lower than the rate in 2011 and 2012 (p = .00).

DISCUSSION

Summary of Main Findings

In the period under study, Dutch midwifery practices reported significantly more provision of smoking-cessation counselling. The training rate of midwives was also higher in this period than in the previous years. However, there was only minimal improvement in midwives' provision of counselling after individual site visits and after questionnaires by the Inspectorate, which

Figure 1 *Training V-MIS of midwifery practices per intervention group*



Note. From start of training in 2003 to end of supervision period in 2012, 3 parts in time period: 2003–2005, high training rate during start of V-MIS training; 2005–2011, low training rate; 2011–2012, high training rate during supervision period. Control group: only deadline announced.

were evaluated in an RCT. Also the training participation rate in counselling by midwifery practices did not differ between the intervention and control groups.

Strengths and Limitations

The design of this study has several strengths and limitations. The first strength is that we could separately study the effect of various elements of the programme. This enabled us to draw conclusions on the programme effects. A range of measures provided in-depth information on changes in counselling during the programme period. A second strength is the longitudinal data collection on training in counselling, which resulted in a clear trend over the last 10 years. A third strength is the high response rate to our study questionnaires and the high proportion of participating practices in the Netherlands. These high rates and proportions resulted in the high reliability and generalizability of the study.

A limitation of the study is that research on national policy lacks a control group, as was the case in the announcement of the deadline (Programme A). As the lack of a control group reduces the validity of the results, it is difficult to directly attribute the improvements in counselling to the Inspectorate's programme.

A second limitation is that the Inspectorate selected only 21 practices for the site visits (Programme C). This small intervention group reduces the power of the analysis. Although we had selected a larger control group, it is difficult to base strong conclusions on this small intervention group.

Due to the risk that the study questionnaire might influence midwives' behaviour before the supervision programme, we waived a pre-intervention measurement for Programmes B and C. Due to randomization of the intervention groups in Programmes B and C, we expect the results of the pre-intervention measurement to be similar.

Except for training in V-MIS, all the outcomes of our study were self-reported. Since the programme was closely related to the study questionnaire, practices may have tended to answer too positively on their provision of counselling, which may have led to socially desirable answers. We had external data on the training in V-MIS and compared the self-reported data on training in V-MIS with real training in V-MIS. The external data did not differ significantly from the self-reported data, and the results varied widely between several outcomes and within groups of our study. This variation indicates that practices were not reluctant to report negative aspects of quality of care. We therefore expect social desirability to have had little influence on the conclusion of our study. However, previous research showed that the majority of general practitioners, for example, highlighted the importance of smoking-cessation counselling, but only a part of them provided counselling systematically.[16]

Interpretation

This supervision programme consisted of a set of elements that

were evaluated separately. Provision of counselling improved significantly after the announcement of a deadline (Programme A). As assessments of smoking behaviour and counselling on quit attempts had almost reached the maximum score in the pre-intervention measurement, these steps did not further improve (ceiling effect). The lack of improvements in registration of smoking behaviour and aftercare of quit attempts remain unexplained.

After assessments with a questionnaire and personal report (Programme B), provision of counselling in practices improved slightly. As the evaluation took place two years after the start of the programme, we interpret these improvements as small but sustained. Due to the quantitative study design, we cannot explain why certain outcomes improved and other outcomes did not.

Due to low power, the effect of assessments with site visit and personal report (Programme C) was not determinable. Effect size and direction of the changes were comparable with those of Programme B. Where site visits may have led to small improvements in counselling, no hard evidence can be derived from this evaluation.

In our evaluation of the announcement of the deadline (Programme A), it is not possible to isolate the effect of Inspectorate's programme from that of the stakeholders' other quality improvement initiatives. The Inspectorate approached several stakeholders to collaborate. To indicate the effect of the Inspectorate's programme, we will describe these collaborations. From the beginning, the expertise centre had been keen to collaborate with the Inspectorate to reduce smoking rates (personal communication). However, the professional midwifery organization did not start new activities on smoking-cessation counselling: other health risks for pregnant women were seen as being as important as smoking (personal communication). After the assessments with questionnaires, it became clear that improvements in counselling were needed. From then on, the professional midwifery organization collaborated with the Inspectorate on improving counselling in practices (personal communication). To help practices with the improvements needed, the expertise centre and professional midwifery organization formulated a handbook on smoking-cessation counselling in midwifery practices.[17]

The Inspectorate was involved only indirectly in the following quality improvements. In 2011, the expertise centre discovered that, although it was very important for improving counselling, training of midwives was decreasing. By redistributing funds, it then arranged a discount on such training. Through various channels facilitated by the professional midwifery organization, it announced the training and the discount. The discount made the training very attractive to midwives, who need to complete a certain number of training hours each year for certification by the professional midwifery organization. In its communication, the expertise centre also mentioned the enforcement by the Inspectorate. These activities may

have led to the increase in training of midwives.

It remains questionable whether other stakeholders would have initiated activities to improve smoking-cessation counselling if the Inspectorate had done nothing. For example, even though a national report related high perinatal mortality rates to high smoking prevalence, the debate following its publication focused on a range of improvements but not specifically on smoking-cessation counselling.[18] We have no indications that organizations other than the Inspectorate, the expertise centre and professional midwifery organization undertook initiatives on improving smoking-cessation counselling.

The results of this study are consistent with the national decline in the tolerability of smoking, which is reinforced by smoking bans in public areas and workplaces.[19] This negative trend may have contributed to the uptake of the Inspectorate's programme.

Smoking-cessation counselling provision improved spectacularly in all midwifery practices in the Netherlands, which is a major achievement. As the Inspectorate's programme represents a substantial part of all the activities described above, we conclude in the absence of alternative explanations that these improvements in provision of counselling can be attributed partly to its supervision programme.

Implications and General Conclusion

Due to assessments with questionnaires and site visits, and also due to the announcement of a deadline, the quality of care improved substantially during the supervision programme. Assessments with questionnaires and site visits are time consuming and costly. The improvements that result from them are limited, but may be necessary for the credibility of supervision. In this case, the shortcomings revealed by the assessments with questionnaires justified the Inspectorate's programme.

V-MIS is an evidence-based method for reducing smoking. Although health outcomes were not measured in this study, its use increased and in theory improvements in health outcomes can therefore be expected. However, pregnant smokers are a small group and only 12% of them quit smoking after use of the V-MIS.[7] Further research may provide more evidence on the use of V-MIS and its effects on smoking cessation by pregnant smokers.

During the development of this study, we experienced the challenges of designing an optimal randomized controlled trial for evaluating government supervision. We concluded that creative study designs are needed for such work. When evaluating this supervision program, qualitative research might provide answers to the question of how counselling changed and why.

In summary, the (self-reported) provision of smoking-cessation counselling improved spectacularly in Dutch midwifery practices. Many quality of care improvement activities were performed during the study period, those of the Inspectorate representing a substantial

part. Despite some limitations of our study, the Inspectorate's supervision programme is likely to have contributed to the improvements in provision of counselling.

REFERENCES

1. Lanting CI, Van Wouwe JP, Van Den Burg I, et al. Smoking during pregnancy (Roken tijdens de zwangerschap) *Ned Tijdschr Geneeskd* 2012;156(50):2123.
2. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* 2004;6(Suppl 2):S125-S40.
3. Mohangoo AD, Buitendijk SE, Hukkelhoven CWPM, et al. High perinatal mortality in the Netherlands compared to other European countries: Peristat II study (Hoge perinatale sterfte in Nederland vergeleken met andere Europese landen: de Peristat-II-studie). *Ned Tijdschr Geneeskd* 2008;152(50):2718-27.
4. Segaar D. Adoption and implementation of smoking cessation support in healthcare [PhD Thesis]. Maastricht University, 2006.
5. Shaw C. External assessment of health care. BMJ 2001;322(7290):851.
6. Boyne G, Day P, Walker R. The evaluation of public service inspection: A theoretical framework. *Urban Stud* 2002;39(7):1197-212.
7. de Vries H, Bakker M, Mullen PD, et al. The effects of smoking cessation counseling by midwives on Dutch pregnant women and their partners. *Patient Educ Couns* 2006;63(1):177-87.
8. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983;51(3):390-95.
9. van Dishoeck AM, Oude Wesselink SF, Lingsma HF, et al. Transparency: can the effect of governmental surveillance be quantified? (Transparantie: is het effect van toezicht te meten?). *Ned Tijdschr Geneeskd* 2013;157(16).
10. Health Council of the Netherlands (Gezondheidsraad.). Towards evidence-based supervision. Research on effects of supervision by the Dutch Health Care Inspectorate (Op weg naar evidence based toezicht, Het onderzoek naar effecten van toezicht door de Inspectie voor de Gezondheidszorg). Den Haag, 2011.
11. Robben P. *Supervision in a glass house: The effectiviness of supervision on the quality of health care* (*Toezicht in een glazen huis, Effectiviteit van het toezicht op de kwaliteit van de gezondheidszorg*), 2010.
12. Stand J. The" Hawthorne effect"-what did the original Hawthorne studies actually show? *Scand J Work Environ Health* 2000;26(4):363-67.
13. Verdonk-Kleinjan WMI, Rijswijk PCP, Candel MJJM, et al. Agreement between self-reports and on-site inspections of compliance with a workplace smoking ban. *Nicotine Tob Res* 2012;14(9):1121-25.
14. Reulings PGJ, van der Lans SMGA, Wentzel HR, et al. Grab your chance, smoking cessation couseling by midwives (Pak je kans: stoppen met roken begeleiding door verloskundigen). *Tijdschrift voor Verloskundigen* 2011;april.
15. Netherlands Institute for Social Research (Sociaal en Cultureel Planbureau). Status scores.
16. Coleman T, Wynn A, Barrett S, et al. Discussion of NRT and other antismoking interventions in UK general practitioners' routine consultations. *Nicotine Tob Res* 2003;5(2):163-68.
17. Dutch professional midwifery organisation, Dutch expertise centre on tobacco control, (Koninklijke Nederlandse Organisatie van Verloskundigen, et al. Guide for policy smoking cessation counselling in midwife practices (Handreiking Ontwikkeling Beleid voor Begeleiding Stoppen met Roken in Verloskundigenpraktijken), 2012.
18. Steering committee pregnancy and birth (Advies Stuurgroep zwangerschap en geboorte). A good start; save care during pregnancy and birth (Een goed begin, Veilige zorg rond zwangerschap en geboorte). Utrecht (NL), 2009.
19. Nagelhout GE, Levy DT, Blackman K, et al. The effect of tobacco control policies on smoking prevalence and smoking-attributable deaths. Findings from the Netherlands SimSmoke Tobacco Control Policy Simulation Model. *Addiction* 2012;107(2):407-16.

APPENDIX BOX S1

Dutch Healthcare Inspectorate

The Dutch Healthcare Inspectorate (Netherlands, EU) is an independent agency of the Ministry of Health, Welfare, and Sport, aims to improve of public health and is expected to efficiently supervise a sector where 1.3 million people work for 40.000 institutions and companies. Its primary instruments are advice and encouragement. If these do not achieve the desired result, it can implement corrective action by, for example, increasing the supervision or through limiting the ability to practice a profession. The Inspectorate enforces 25 laws, for example the Care Institutions Quality Act. The supervision is performed by a combination of three methods:

1. theme based regulation, directed at specific issues in care, sometimes asked for by the minister or parliament
2. regulation in response to calamities in the event of emergencies that indicate structural shortcomings in care provision
3. risk-based supervision to assess the quality of healthcare by means of indicators

The intervention of this study is an example of theme based regulation.

APPENDIX BOX S2

Dutch Midwifery care system

In the Netherlands midwives are medical practitioners with competencies restricted to independently provide care during 'normal' pregnancy, childbirth and the postpartum period. Midwifery education is a 4-year vocational training (at a university for applied science) with a degree at bachelor level. Women with an uncomplicated pregnancy are expected to receive care from a primary care midwife and are not required to consult a gynaecologist/obstetrician; the midwife is fully responsible for the care provided. But when complications arise, or threaten, or the woman requests a form of pain relief that can only be given in secondary care, the midwife will have to transfer responsibility by referring the woman to a gynaecologist. She can choose to stay with her client or leave her in the care of a hospital midwife or nurse, but from the moment of referral onwards the gynaecologist is responsible for the care provided.

Primary care midwives work in private practice. They can work as self-employed practitioner or as employee in someone else's practice. Self-employed practitioners work alone or in partnership with one or more other midwives. Also many midwives work as locum and fill in vacancies in midwifery practices on a temporary basis. Primary care midwives are paid per care unit, separately for pre-natal, natal and post-natal care. Consequently, if a client leaves the practice, that practice suffers from financial loss.

Adapted from T.A. *Wiegers – Work and workload of Dutch primary care midwives in* 2010 (2013)

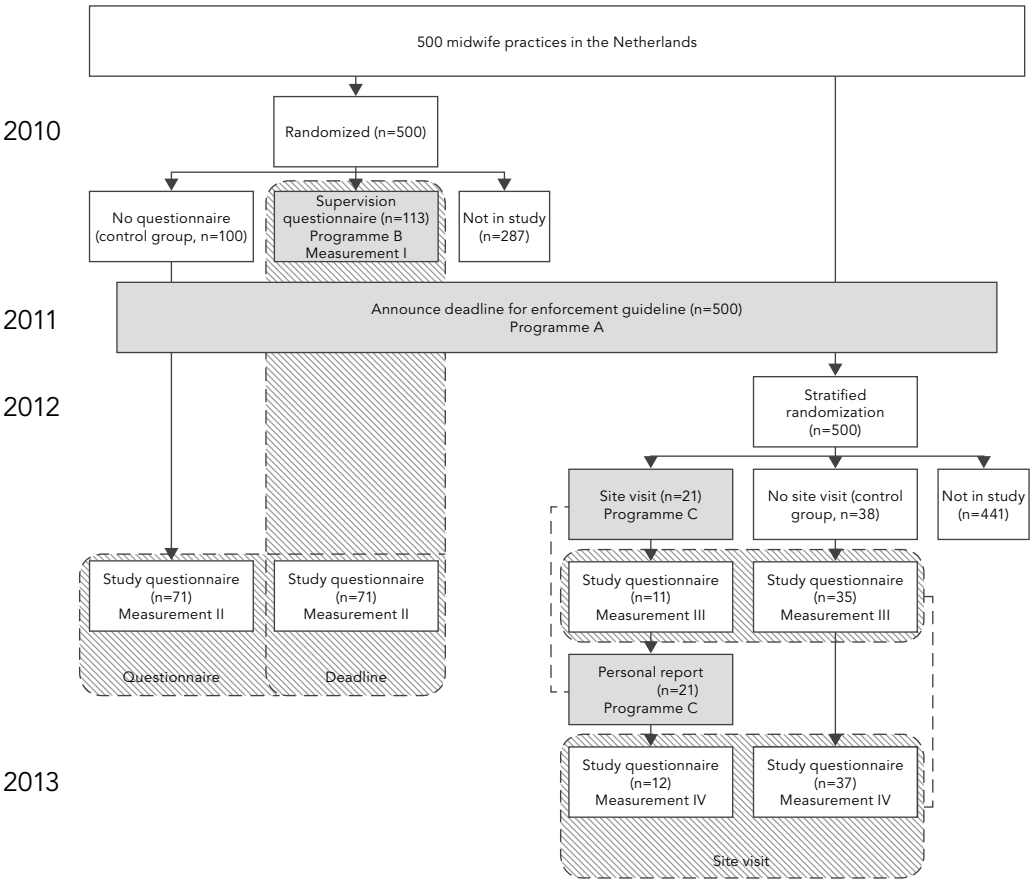
APPENDIX TABLE S1

Smoking-cessation counselling (scale 0 to 1) stratified by time of measurement or intervention and control group, separated for the three programme elements

Evaluation A (Deadline)	**2010**				**2012**			
(measurement I and II)	**n**	**mean**	**median**	**IQR**	**n**	**mean**	**median**	**IQR**
Use V-MIS intervention	113	0.28	0	0-1	69	0.80	1	1-1
Education of midwives	113	0.45	0.67	0-0.83	71	0.78	1	0.67-1
Smoking policy	113	0.31	0.33	0-0.33	71	0.65	0.67	0.33-1
Support from other care providers	113	0.43	0.33	0.33-0.33	71	0.53	0.33	0.33-1
Registration smoking behaviour	113	0.73	0.75	0.75-0.75	71	0.76	0.75	0.75-0.75
Assessing smoking behaviour	108	0.98	1	1-1	70	1	1	1-1
Registration and counselling smoking partner	113	0.84	1	0.67-1	71	0.96	1	1-1
Assessing motivation to quit	106	0.90	1	0.88-1	68	0.96	1	1-1
Enhancing motivation to quit	107	0.93	1	0.90-1	68	0.97	1	0.96-1
Discussing barriers to quit	102	0.76	0.8	0.5-1	67	0.93	1	0.9-1
Reducing barriers to quit	108	0.94	1	0.9-1	67	0.97	1	0.95-1
Setting a quit date	112	0.21	0.00	0.0-0.48	69	0.61	0.75	0.29-0.95
Counselling of quit attempt	60	0.92	1	0.88-1	62	0.94	1	0.95-1
Aftercare quit attempt	113	0.84	1	1-1	69	0.86	1	1-1
Total score	52	0.73	0.74	0.65-0.82	58	0.85	0.86	0.77-0.92
Evaluation B (Questionnaire)	**Intervention**				**Control**			
(measurement III)	**n**	**mean**	**median**	**IQR**	**n**	**mean**	**median**	**IQR**
Use V-MIS intervention	69	0.80	1	1-1	70	0.71	1	0-1
Education of midwives	71	0.78	1	0.67-1	71	0.77	0.67	0.66-1
Smoking policy	71	0.65	0.67	0.33-1	71	0.57	0.67	0.33-1
Support from other care providers	71	0.53	0.33	0.33-1	71	0.46	0.33	0.33-0.33
Registration smoking behaviour	71	0.76	0.75	0.75-0.75	71	0.74	0.75	0.75-0.75
Assessing smoking behaviour	70	1	1	1-1	71	0.99	1	1-1
Registration and counselling smoking partner	71	0.96	1	1-1	71	0.95	1	1-1
Assessing motivation to quit	68	0.96	1	1-1	69	0.97	1	1-1
Enhancing motivation to quit	68	0.97	1	0.96-1	64	0.93	1	0.88-1
Discussing barriers to quit	67	0.93	1	0.9-1	66	0.87	1	0.8-1
Reducing barriers to quit	67	0.97	1	0.95-1	62	0.93	1	0.9-1
Setting a quit date	69	0.61	0.75	0.29-0.95	66	0.49	0.5	0.18-0.8
Counselling of quit attempt	62	0.94	1	0.95-1	52	0.93	1	0.9-1
Aftercare quit attempt	69	0.86	1	1-1	70	0.81	1	1-1
Total score	58	0.85	0.86	0.77-0.92	50	0.81	0.85	0.75-0.89
Evaluation C (Site visit)	**Intervention**				**Control**			
(measurement IV)	**n**	**mean**	**median**	**IQR**	**n**	**mean**	**median**	**IQR**
Use V-MIS intervention	10	0.80	1	0.75-1	35	0.71	1	0-1
Education of midwives	11	0.82	0.67	0.67-1	35	0.72	1	0.67-1
Smoking policy	11	0.79	1	0.67-1	35	0.60	0.67	0.33-1
Support from other care providers	11	0.52	0.33	0.33-0.67	35	0.46	0.33	0.33-0.33
Registration smoking behaviour	10	0.78	0.75	0.75-0.75	35	0.76	0.75	0.75-0.75
Assessing smoking behaviour	10	0.99	1	1-1	35	0.97	1	1-1
Registration and counselling smoking partner	10	0.93	1	0.92-1	35	0.97	1	1-1
Assessing motivation to quit	10	0.99	1	1-1	32	0.95	1	1-1
Enhancing motivation to quit	9	0.94	1	0.95-1	34	0.95	1	0.95-1
Discussing barriers to quit	10	0.97	1	1-1	34	0.92	1	0.88-1
Reducing barriers to quit	9	0.96	1	0.94-1	33	0.96	1	0.95-1
Setting a quit date	10	0.71	0.88	0.40-1	31	0.58	0.5	0.2-0.99
Counselling of quit attempt	9	0.98	1	0.95-1	28	0.96	1	0.95-1
Aftercare quit attempt	10	0.90	1	1-1	35	0.80	1	1-1
Total score	9	0.86	0.88	0.79-0.91	26	0.84	0.86	0.77-0.90

APPENDIX FIGURE S1

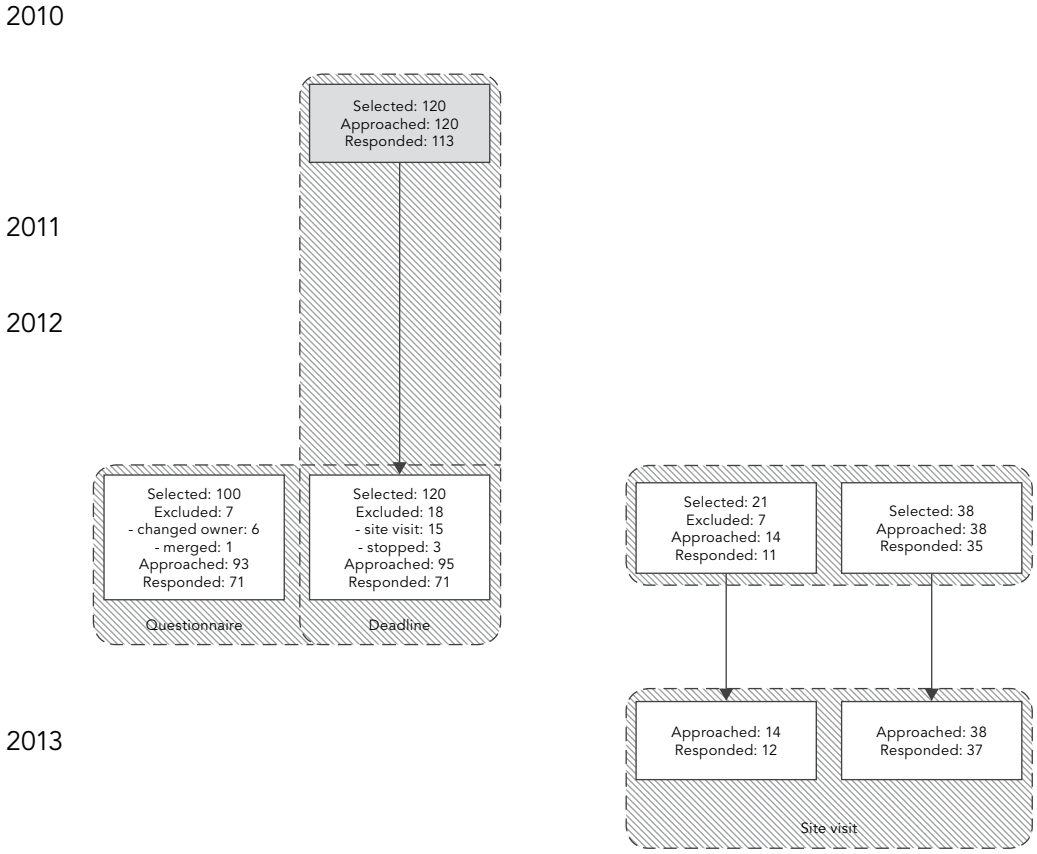
Flow chart and timeline of Inspectorate's programme and study participants



Note. Grey: Inspectorate activities. Shaded: Study evaluations. White: Study activities

APPENDIX FIGURE S2

Flow chart and timeline of Inspectorate's programme and study participants: response rate



Note. Grey: Inspectorate activities. Shaded: Study evaluations. White: Study activities

CHAPTER 4

Qualitative exploration of government supervision on quality of quit-smoking counselling in midwifery practices

Sandra F. Oude Wesselink
Annemiek Stoopendaal
Vicki Erasmus
Déan Smits
Johan P. Mackenbach
Hester F. Lingsma
Paul B.M. Robben

ABSTRACT

Introduction

The Dutch Healthcare Inspectorate supervises care providers in order to improve quality of care. Recently the inspectorate assessed and promoted the use of a guideline on smoking-cessation counselling in midwifery practices. The supervision programme consisted of an announcement of the enforcement deadline for the guideline and site visits. The purpose of our qualitative study was to identify factors related to guideline adherence after the supervision programme, and investigate whether the programme had helped improve adherence.

Methods

We conducted semi-structured interviews with inspected and non-inspected midwives. Additionally, we studied documents and observed the inspection process. The sampled midwives all work in primary care midwifery practices providing care to pregnant smokers. The questions included the current provision of smoking-cessation counselling, support to the midwife in counselling, recent changes in provision of counselling, reasons for recent changes, knowledge about the supervision programme, and experiences with supervision by the inspectorate.

Results

Our results show that guideline adherence depends on several factors. Awareness and familiarity with the guideline are important, as is outcome expectancy. Additionally, motivation, guideline factors and environment factors were mentioned. Besides these previously documented factors, we found that professional collaboration also determined guideline adherence. Increased collaboration in counselling is associated with greater adherence to the guideline, such as provision of counselling and taking required training. The supervision programme helped improve stop-smoking counselling, by making midwives aware of the counselling and giving them an extrinsic motivation to provide counselling.

Conclusions

In conclusion, of the factors related to guideline adherence, motivation and environmental aspects were the most important and professional environment was added as significant factor for guideline adherence. The improved adherence is partly attributable to the supervision programme.

INTRODUCTION

Six percent of women in the Netherlands smoke during pregnancy.[1] Among lower educated women, the prevalence of smoking during pregnancy is around 14%. Maternal smoking is associated with a higher risk of foetal mortality and of adverse birth outcomes such as stillbirth, preterm birth, small for gestational age, intrauterine growth restriction, and congenital heart defects.[2]

Improvement of quality of care is an ongoing multidimensional process with various approaches. One approach is external assessment, based on models of peer review, accreditation, and inspection.[3] This study focuses on the inspections enforced under national or regional statutes, whose standards are derived from regulation and existing guidelines.[3] Inspectorates can use various instruments, such as site visits and performance indicators.[4] The main focus lies on the competence of professional staff, compliance with professional standards, and outcomes for service users.[4] In the Netherlands, healthcare supervision is delegated to the national Dutch Healthcare Inspectorate (later: inspectorate) (Appendix 1).

In 2010, the inspectorate began a supervision programme on primary care midwives providing care to pregnant smokers. It focused on the evidence-based Minimal Invention Strategy for Smoking-Cessation Counselling for Midwifery Practices (Minimale Interventiestrategie Stoppen met Roken voor de Verloskundigenpraktijk, V-MIS[5]) (Box 1). The professional guideline recommends providing smoking-cessation counselling to pregnant smokers.[6] Apart from V-MIS, almost no other methods to provide counselling are used. In the period 2010–2012, the inspectorate promoted the use of V-MIS in a supervision programme intended to improve the quality of counselling and reduce smoking rates during pregnancy. The inspectorate collaborated with the Royal Dutch Organisation of Midwives (Koninklijke Nederlandse Organisatie van Verloskundigen, KNOV) and the Netherlands Expertise centre for Tobacco Control (Stichting Volksgezondheid en Roken, STIVORO). In a previous study we found that use of V-MIS increased substantially from 28% in 2010 to 80% in 2012.[7] This spectacular improvement in adherence to the guideline on smoking-cessation counselling is not fully attributable to the su-

***Box 1** V-MIS*

V-MIS comprises seven steps. In step 1, the midwife identifies the smoking behaviour of the woman and partner. In step 2, the midwife attempts to enhance the motivation to quit. In step 3, the midwife and woman discuss barriers for successful quitting and how to mobilise social support for quitting. In step 4, the midwife and woman agree on a quit date. In step 5, the midwives discuss and provide additional self-help materials. In step 6, the midwife provides aftercare if necessary. In step 7, the midwife supports the woman to prevent relapse after delivery. These steps can be provided in one or more consultations. When V-MIS is applied, 12% of the pregnant smokers quit, whereas 3% in the control group quit.[5]

pervision programme, because other organisations were also involved in promoting quit-smoking counselling. Therefore, we wanted to understand how this improvement was achieved. The purpose of our study was to identify factors related to guideline adherence after the supervision programme, and to investigate whether the supervision programme had helped improve adherence.

CONTEXT

Perinatal care in the Netherlands and smoking-cessation counselling

In the Netherlands, pregnant women have a free choice for place of birth, including at home.[8] For low-risk pregnancies and deliveries midwives may provide care on their own during gestation, childbirth and the postpartum period. All midwives have obtained a bachelor's degree from a university of applied sciences.[9] This education involves four years of theory and internships combined. After graduation, midwives have to take at least 200 hours of training courses and further education every five years to stay listed in the quality register for professional midwives.

Primary care midwives work in private practices, either as self-employed practitioners or as employees in someone else's practice. Self-employed practitioners work alone or in partnership with one or more other midwives. Many midwives work as locums, filling in temporary vacancies in midwifery practices.

Most practices work with teams of 3 to 5 midwives caring for one pregnant woman, with one team member assisting at delivery and the team sharing information on the woman through the electronic patient file. Each practice has a midwife on call 24/7. In the past few years, most hospitals work with multidisciplinary obstetric partnerships, involving all birth care providers in their hospital, including primary and secondary midwives, gynaecologists, paediatricians, maternity nurses and obstetric general practitioners.

Primary care midwives are paid per care unit, separately for pre-natal, natal and post-natal care.[10] Consequently, the practice suffers a financial loss if a client leaves the practice, especially during pre-natal care. To be paid, practices are required to have contracts with healthcare insurers.[8] The healthcare insurers may ask for improvements in quality of care when negotiating contracts.

According to the guideline, midwives must provide smoking-cessation counselling to pregnant smokers. A minimal intervention strategy, V-MIS increases the quit-smoking rate in pregnant smokers.[5] It targets midwifery practices and is based on the stages of change theory.[11] Midwives use V-MIS during their normal consultations with pregnant smokers or plan a separate consultation to provide smoking-cessation counselling.

Perinatal mortality in the Netherlands used to be higher than in other European countries so there was room for improvement in

perinatal care, specifically in smoking behaviour.[12]

Case: Programme of the Dutch Healthcare Inspectorate

The Dutch Healthcare Inspectorate programme aimed to improve the provision of smoking-cessation counselling to pregnant women by all primary care midwives in the Netherlands.

In 2010, inspectors visited a small sample (10 of 500) of midwifery practices to discuss counselling based on V-MIS with the midwives, first mailing an announcement of the impending visit and the supervision topics. In this exploratory phase, the inspectorate did not enforce compliance to the guideline. Two inspectors visited each site for 2 hours, with smoking-cessation counselling as the only topic of discussion. Despite the availability of V-MIS and the guideline, only a minority of Dutch midwives provided smoking-cessation counselling in 2010.[13] As the inspectorate is supposed to promote public health, part of their job is to monitor and encourage guideline adherence. Therefore, after these preliminary site visits, the inspectorate decided in consultation with the professional organisation to oblige midwives to use V-MIS, because this method is used most frequently and is most suitable for midwives. Then they announced the enforcement deadline of the guideline to all midwifery practices and all 10 inspected practices received a personal report with feedback on their counselling.

In 2012, the inspectorate again visited a sample (21 of 500) of midwifery practices to check whether midwives were complying with the guideline. They inspected policy documents, training certificates and registration forms, and evaluated the use of V-MIS and the midwives' knowledge of places they could refer women to for support on stopping smoking. Again, mails announced the site visits and supervision topics. Two inspectors took a half day to inspect each practice, spending 10% of the site visit on smoking-cessation counselling and using the rest of the time to address other topics relevant to the quality of midwifery care. Following the site visits, all the inspected practices received a personal report with feedback on their counselling and a time frame for implementing the required improvements. All reports, including the personal reports are available to the public.

The inspectorate's ultimate measure is to shut down a midwifery practice, in which case that practice cannot accept new clients and must hand over current clients to other midwifery practices. The inspectorate has never applied this ultimate measure to any Dutch midwifery practices, but does so occasionally in nursing homes, home care organisations and hospital departments.[14]

Alongside the inspectorate, healthcare insurance companies may audit guideline adherence. Insurers may ask practices for improvements to the quality of specific aspects of care. The insurers' ultimate measure is to cancel their contract with a midwifery practice so that

the midwives receive no payment for clients insured by that insurance company.

Support for midwifery practices on smoking-cessation counselling

The aim of the Netherlands Expertise centre for Tobacco Control (STIVORO) is to promote a cigarette smoke-free future. The professional midwifery organisation strives at the best care for pregnant women and their partner. STIVORO and professional midwifery organisation collaborated in the provision of support to midwifery practices to improve smoking-cessation counselling. STIVORO has a programme to reduce second-hand smoke, which also includes the reduction of pregnant smokers. They developed the V-MIS in cooperation with a scientific institute and offer training to midwives about how to use V-MIS in their practice. Self-help materials are also provided by STIVORO.

The Dutch professional midwifery organisation represents the interests of midwives in the Netherlands in a powerful way. Besides their national office, in each region of the Netherlands they have a local network of midwives, called a circle group. In these circle groups, they discuss all aspects of midwifery care in the Netherlands and the national office can provide input to these meetings. All chairs of these circle groups meet regularly with the national office. At these chair meetings other stakeholders may also introduce relevant topics. The chairs then pass the information to the other midwives in their region.

During the supervision period, the facilitation of smoking-cessation counselling improved. Both STIVORO and the professional organisation committed to helping midwifery practices improve counselling, after a consultation with the inspectorate. In 2011, STIVORO discovered that fewer midwives were taking training courses, although this was very important for improving counselling. Redistributing its funds, STIVORO then arranged a discount for the training course and announced this through various channels facilitated by the professional midwifery organisation. The discount made the training very attractive to midwives. In its communication, STIVORO mentioned the enforcement by the inspectorate. The course also paid attention to other referral options that would support pregnant smokers. Besides collaborating on the training course, STIVORO and the professional midwifery organisation jointly published a handbook on smoking-cessation counselling[15]. The midwifery practices could use this handbook to formulate their policy on smoking-cessation counselling in their practice. Lastly, STIVORO requested the software companies who provide software for patient record systems to include items on smoking-cessation counselling in the electronic patient record. Based on V-MIS, the items include the preferences of the pregnant smoker and the actual care provided by the midwife. Such enhancements improved the continuity of smoking-cessation counselling.

THEORETICAL FRAMEWORK

To identify factors related to guideline adherence and investigate the contribution of the inspectorate, we applied two different theories. The behaviour of midwives we describe with Cabana's guideline adherence theory[16] and the behaviour of the inspectorate according to the responsive regulation theory.[17]

Guideline adherence is determined by various factors. The sequence of behaviour change ranges from knowledge through attitudes to behaviour.[16] For knowledge, it is important to be aware of and familiar with the guideline. This includes, for example, the amount of information, the time needed to stay informed, and guideline accessibility. Attitude is determined by several factors including agreement with specific guideline characteristics, agreement with guidelines in general, outcome expectancy, self-efficacy, and motivation. Outcome expectancy refers to whether the midwives believe that following the guideline recommendations will lead to the desired outcome, in our case that pregnant women quit smoking. Self-efficacy means that the midwife believes that they can follow the guideline recommendations. Lastly, behaviour is influenced by external barriers, guideline factors, and environmental factors, which include time, resources, organisational opportunities, and reimbursement.

The inspectorates stimulates guideline adherence through responsive regulation. This method of supervision uses the reactions of the regulated entities to determine the degree of supervision, applying an enforcement pyramid, which ranges from persuasion at the bottom to license revocation at the top.[17] The idea behind the pyramid is that it will be easier to persuade regulated entities to follow the guidelines if they know about the 'big guns' (deterrents). In this case the deterrent is the power of the inspectorate to close the midwifery practice. The pyramid also shows that for small violations that care providers are willing to improve, the inspectorate has to start with the lowest step of the pyramid and not with the big guns.

METHODS

Data collection

Practices were first approached by e-mail and later by phone. Interviews took place at the midwife's workplace and were conducted preferably with the midwife responsible for smoking-cessation counselling in the practice. The interviewer and midwife had no pre-existing relation. All interviews lasted between 30 and 60 minutes and took place between March and June 2013.

The participants, professional midwives, gave written informed consent for participation and following their interview received compensation in the form of a voucher for 10 euros. One researcher (DS), an MSc student trained and experienced in conducting interviews, did all the interviews.

Besides the interviews, we collected additional data from the supervision

programme. We observed meetings of inspectors and inspections and, to be as well informed as possible, collected minutes and other documents by the inspectorate.

Study population

The study population was three groups of midwives working in primary care midwifery practices. The first group contained midwifery practices that were inspected in 2010. We approached 8 practices of which 5 took part in this study. The second group consisted of midwifery practices that were inspected in 2012. We approached 7 practices of which 4 participated. The

Box 2 *Interview guide*

Smoking-cessation counselling

1. How do you provide smoking-cessation counselling to pregnant smokers? What did you change in care to pregnant smokers last years? Which support can you turn to?
2. What did your colleagues change in care provided to pregnant smokers?
3. Which support did you receive in the care for pregnant smokers? Is this changed last years?

Inspection

4. What did you hear about the supervision programme on midwife practices with respect to smoking-cessation counselling? How have you obtained this information?
5. Why conducted the inspectorate this supervision programme according to you?
6. Have you read the supervision report or the publication in the journal of your professional organisation?
7. To what extent was this publication recognisable to you?

What changed as a result of inspection?

8. To what extent did the inspectorate contribute to this change?
   - Were these changes affected by other actors?
9. To what extent did the inspectorate contribute to the change of your colleagues?

Why did you change the way you work?

10. Which aspects contributed to compliance to instructions from the inspectorate?
    Conceptual model: autonomy, workload, way of inspection, motivation, field standards, transparency, trust, (in)dependence of the inspectorate, expectations and relationship with the inspectorate

Change in inspection

11. If you should perform the supervision, how would do it? How would be the impact of the supervision on your work be the largest?

third group held a random selection of practices that had not been inspected. Here we approached 11 practices, of which 5 agreed to participate. All practices were selected randomly and practices with any previous involvement in our research were excluded. This ensured that all practices included in the current study were not included in other studies performed by the Department of Public Health, Erasmus Medical Centre.

Questions

The interviews were based on a questionnaire guide (Box 2). The questions were about the current provision of smoking-cessation counselling, support to the midwife in counselling, recent changes in provision of counselling, reasons for recent changes, knowledge of the supervision programme, and experiences with supervision. The interview questions were first tested on colleagues.

Data analyses

All interviews were audio recorded and fully transcribed. Names and privacy-related information were removed. Interview transcripts underwent systematic content analysis based on grounded theory[18], using NVivo software, version 10 (QSR international, Doncaster, Australia). Phrases were combined to generate categories. This process continued until all transcripts were analysed and no new categories emerged. Subsequently, the content of the categories was analysed for overlapping or linking content. The categories were then compressed and clustered into themes. The themes were evaluated across the different groups and respondents to search for similarities and differences. Finally, we analysed the data using Cabana's model[16] and clustered the information into the guideline adherence factors.

Information from observations and document analyses were used as background information to understand the outcomes of the interviews. Therefore, they were not transcribed or analysed. A previous quantitative study on supervision on smoking-cessation counselling also functioned as background information.[7]

RESULTS

Population

In total 14 midwives participated in our study (Table 1). The average age of midwives in the inspected group was 44 years and in the non-inspected group 45 years. Almost all midwives were female, except one in the non-inspected group. In both groups, one midwife smoked and in the inspected group, two midwives were former smokers. Four midwives were physically present during the inspections of the inspectorate. On average, participants in the inspected group worked 16 years as a midwife and 11 years in this practice. Participants in the non-inspected group worked 18 years as a midwife and 15 years in this practice. The inspected group treated on average 28 smokers in their practice per year and the non-inspected group 48 smokers per year. The interviews lasted about

44 minutes in the inspected group and 40 minutes in the non-inspected group.

The following sections present the relevant factors related to guideline adherence, described in our theoretical framework.

Awareness of the guideline and the supervision programme

All midwives were informed about the supervision on smoking-cessation counselling, including both inspected and non-inspected midwifery practices. This information reached them through various channels. Some practices indicated that they received the information directly from the inspectorate by letter or report. Others stated that the professional organisation told them about the supervision. The last possibility was that practices were informed through neighbouring practices that had received an inspection.

"I *can remember it*. I *got a letter.* It *was a few years ago*. (…) So, *the letter did get here*." Midwife, female, not inspected by inspectorate

All midwives knew about the existence of the supervision on smoking-cessation counselling and were thus informed about the guideline. The midwives said that the supervision programme showed them how to improve their counselling. The midwives felt that counselling was effective: pregnant smokers smoked less or counselling did not take much extra time. Midwives are against smoking and are strongly motivated to strive for the good health of mother and child. Midwives are committed to healthy pregnancy and delivering a healthy baby. Perinatal audits raised awareness of the harm of smoking. In these audits, midwives in the multidisciplinary obstetric partnership discuss all the babies who had died during pregnancy or delivery. In almost all cases, the baby's mother was a smoker.

Some inspected practices were rated as 'inadequate' by the inspectorate. The midwives indicated that this rating caused embarrassment and disappointment in the practice. The public availability of the report enhanced these feelings as

Table 1 *Characteristics of participants*

	n	Inspected	n	Non-inspected
Age in years (mean)	9	44	5	45
Female (%)	9	9 / 9	5	4 / 5
Smoking behaviour midwife (%): *non smoker*	9	6 / 9	4	3 / 5
smoker		1 / 9		1 / 5
past smoker		2 / 9		0 / 5
Present during inspection (% yes)	9	4 / 9	5	0 / 5
Years working as midwife (mean)	9	16	5	18
Years working in this practice (mean)	9	11	5	15
Pregnant smokers in practice per year (mean)	9	28	1	48
Duration interview in minutes (mean)	9	44	5	40

for practices and clients that could see the report. However, in these practices it also worked as an extra motivation to provide better counselling. Public reporting created extra awareness in other midwives who were not inspected by the inspectorate, which spread knowledge about the guideline.

Familiarity with guideline

Care providers are deemed familiar with a guideline when they can correctly answer questions on the guideline and when they self-report familiarity. Midwives wanted to improve their familiarity with the guideline and smoking-cessation counselling:

"I *felt that my counselling to pregnant smokers was not good enough*. I *wanted to learn more about how to provide neutral and effective counselling*." Midwife, female, inspected by inspectorate

The midwives gained familiarity during courses in stop-smoking counselling which provided information on external referral possibilities:

"D*uring the course we learned how we should or can refer people, to let them quit smoking*. I*t's the task of the* GP, *but we need to refer them*." Midwife, female, inspected by inspectorate

They also learned how to refer from colleagues or through information they collected themselves:

"W*e have a map in this tray that shows where we send them to* (*referrals*), *the outpatient clinics and so on*. S*o it's become easier and clearer*." Midwife, female, inspected by inspectorate

These improvements made it easier to refer pregnant smokers to organisations outside the midwifery practice. This familiarity improved guideline adherence. However, despite a range of improvements, in some regions referrals are still not optimal. The guideline provides no clear guidance on how to inform midwives about external referral options or how often midwives require training in counselling.

Outcome expectancy

Guideline adherence should lead to the expected outcome. However, this is not always the case as adherence may have other, unintended outcomes. Some practices indicated to be less strict in providing smoking-cessation counselling. Clients left the practice after the midwife tried to persuade them to quit smoking:

"I *work on a small scale*. W*e always want to keep our clients*. W*hen* I *come down hard on a pregnant smoker's behaviour and the next practice doesn't do that, and my client hears about it, she can easily switch practices*. N*o midwife wants that*. (…) I*t has happened*. O*ne left my practice*." Midwife, female, inspected by inspectorate, stopped smoking 18 months ago

Many practices reinforced their counselling after inspection. An unintended side-effect was that some clients left the practice, which had a huge impact on the midwives. Midwives said that they do not want to be known as more rigorous than other practices. Reputation is very important for them given that a bad reputation can lead to fewer clients

registering at their practice and consequently less work and lower income. If a client left the practice, the midwife will decide to ease up on the counselling a little. However, if she followed the minimal intervention strategy carefully, clients should feel supported, not offended, by the midwives. Midwives who struggle with the methods might need more training to provide counselling without upsetting their therapeutic relationship with clients. The training in counselling provision might be too short for midwives who find it harder to provide counselling. Regularly repeating the course might be an option for them. One practice found another solution to improve counselling, without burdening the midwives. Here they referred pregnant smokers to other care providers in the practice. A specialist addiction nurse provides counselling to pregnant smokers, which might be more effective.

If the guideline adherence leads directly to the expected outcome, the situation is totally different:

"*When it's an improvement*, I *feel good about it*. I *believe everything can go to work towards a better outcome*. *That's very important for me, a good outcome*." Midwife, female, not inspected by inspectorate

This midwife indicates that counselling works as an improvement of care. Other midwives report counselling as standard provided care. Some midwives said that they do not feel as if it affects their autonomy, because the care improves. So, they can accept the inspections easily:

"*It's for a good cause, what they do. They want to improve the quality of care.*" Midwife, female, inspected by inspectorate

There is a common interest in improving quality of care. The midwives said that they understand and respect the inspectorate, although they think the inspections are inconvenient.

Self-efficacy

The belief that one can actually perform certain behaviour is called self-efficacy. The practices that did not improve their smoking-cessation counselling gave various reasons for thus. The midwife's own smoking behaviour might play a role:

"*My locum also smokes and she almost never talks about it with clients*. (…) *The point, of course, is since you smoke too, you don't ask about it at every consultation*." Midwife, female, not inspected by inspectorate, smoker

It seems hard to advise pregnant smokers to quit, if you yourself are also unable to quit. Both smokers and non-smoking midwives suggested this. On the one hand, one can argue that smoking midwives are connected more to pregnant smokers, but our study showed that it is mostly the other way around; midwives who smoke give almost no counselling. This is a lack of self-efficacy. The midwife does not believe that she can follow the guideline, because she smokes.

Self-efficacy is also needed to complete the training:

"*Interviewer: Did you take any training courses in smoking-cessation counselling?*

Respondent: N*o, none*. N*othing* (…) I*t just didn't happen*." Midwife, male, not inspected by inspectorate, non-smoker

This example shows that not all practices have one midwife trained in smoking-cessation counselling. These practices are thus not following the guideline.

Motivation

Lack of motivation can hinder guideline adherence in many ways. In one practice the responsible midwife, who is probably also the most motivated, left the practice:

"S*he was the specialist, but she left the practice last year and now we the same problem again*." Midwife, female, inspected by inspectorate

The practice has lost knowledge, because the midwife left. There is also a risk of deterioration in smoking-cessation counselling, because a new, motivated person has not been made responsible. The lack of motivation now hinders guideline adherence.
Motivation can be both intrinsic and extrinsic. This midwife is extrinsically motivated, by the obligation of the guideline:

"I *assume we had to*. (…) B*ecause somebody had to do it*. A*nd it does motivate, when things are mandatory, than somebody does it*." Midwife, female, not inspected by inspectorate, non-smoker

The midwives said they felt a sense of duty to the inspectorate and professional organisation. If these organisations told them to do something, they reported, then they would want to follow the instruction. They wanted to adhere to rules and protocols. However, some midwives were intrinsically motivated. They said that wanting to take training courses belongs to their professional attitude.
Lack of motivation can also lead to specific non-adherence. For example, the midwives had to buy (and pay for) the self-help materials themselves. Some midwives said that they ordered these materials together with other practices in the circle, because large orders received more discount. Previously the materials were free, but in recent years the price of leaflets has gone up:

"I *distribute stuff from the outpatient-smoking clinic at the hospital*. B*efore, it came from* STIVORO*, but this material is no longer free and* I *don't understand why* I *should have to pay for self-help materials for clients*." Midwife, female, inspected by inspectorate, non-smoker

Less self-help material is distributed among pregnant smokers, because of increasing costs. However, cost is not the only important aspect to consider:

"W*e don't have the* V-MIS *self-help material, because you have to pay to get it*. W*e requested it a while ago, but we had to pay a significant amount for it*. (…) I*'m willing to inform people for the good of the cause, but why should we care providers have to pay for it*?" Midwife, female, inspected by inspectorate, smoker

Although the provision of counselling is more expensive than the supportive self-help materials, the midwives decided to stop buying them.

Motivation is also important for following the preferred training courses. Some midwives did not bother going on a course in counselling:

"*The range of education available is so broad you have to make choices at a certain moment in time. So you go on courses for just the urgent problems in your practice. It has to do with incidence, and urgency, yes, in your practice.*" Midwife, female, inspected by inspectorate, smoker, practice with 20-25 smokers a year

This midwife did not take the training course, because the staff of the practice had agreed that a colleague would go instead. Others practices reported that their staff did not go to the course because other courses were more important to them. However, there was also extrinsic motivation for those who did go to the course. For example, midwives must follow several hours of training a year to stay in the quality register or because they thought that the course would look good on their CV. In general, most midwives said that they took the counselling course because it was easy to attend as it was put on in their region or during their midwifery education.

Motivation also relates to the perception of the midwife's task. Midwives differed in their opinion of whether they are responsible for their clients' addictions. Some midwives stated that they were not responsible and therefore they did not follow the guideline.

In summary, the reasons for following the prescribed training were both intrinsic and extrinsic. Midwives indicated that they wanted to learn more about effective counselling or felt forced by the inspectorate to take the course. The reasons for not taking the training relate to the attitude and motivation of the midwife. Non-attendees see the training as less important, compared to other courses and activities. In some cases this also relates to the number of pregnant smokers in the practice. Non-attendees have relatively few pregnant smokers in their practice.

Guideline factors

The organisation that prescribes the guideline affects adherence. However, we found that this is not the same for all midwives. Midwives differ in their opinion about the relation between midwives, the inspectorate and the professional organisation. According to the midwives, the professional organisation is closer to the midwives than the inspectorate is. The inspectorate is independent and therefore at a greater distance. However, this distance is interpreted differently by different midwives. Some find the advice of the professional organisation more important:

"I *think that the professional organisation is more credible for me.* (…) B*ecause they are there for the midwives.* (…) I*f the professional organisation says, 'You shouldn't use this programme' and the inspectorate says* 'N*o, you must use it', then* I*'d say, let's use the standard programme and follow the advice of the professional organisation.*" Midwife, female, inspected by inspectorate

The professional organisation defends the interests of the midwives

and is therefore experienced as a leader. For other midwives this close collaboration leads to less pressure:

"I *think a letter from the inspectorate comes across stronger than a letter from the professional organisation, because we have more correspondence with the professional organisation, and less from the inspectorate*." Midwife, female, inspected by inspectorate

Here the midwife says that a letter from the inspectorate has more influence in the practice, than one from the professional organisation, merely because the inspectorate is at more of a distance. Trust in the professional organisation and inspectorate also differs between practices. Some practices have more trust in the inspectorate, as an independent organisation. Others have more trust in the professional organisation, because they represent the interests of the midwives:

"I *do trust the inspectorate, but the professional organisation is for our profession, so* I *trust them more*." Midwife, female, not inspected by inspectorate

As the professional organisation is the advocate of the midwives, the midwives feel more connected to them. The inspectorate is seen as an organisation higher up in hierarchy, which underlines their independence:

"*The inspectorate was decisive for me, because they are another agency. The professional organisation is an association and* I *can follow what they say or not. The inspectorate is the highest agency in the hierarchy and they have to check whether people in healthcare are providing good care*." Midwife, female, inspected by inspectorate

The midwife recognises the legal status of the inspectorate. However, many midwives believe the professional organisation is more credible. Some employees of the professional organisation also work in midwifery practices, and not only in desk jobs. This helps them to have a good view on the practices of providing care.
The collaboration of the professional organisation and inspectorate is seen as important:

"I *would prefer that the inspectorate works with the professional organisation. No inspections without the professional organisation, because they are the representatives of all midwives*." Midwife, female, not inspected by inspectorate

Different midwives think differently about their position in relation to the professional organisation and the inspectorate. As the influence of these institutions on midwives is different, both can benefit from the differences in opinions by collaborating where possible. Close collaboration makes their message stronger and that leads to a coherent message to the midwives.

Environmental factors

Some factors that inhibit or foster guideline adherence are beyond the control of midwives. During the supervision programme, digital registration of smoking-cessation counselling became available and many practices switched over. They felt it was an improvement as it works as a reminder and check-

list, and it facilitates collaboration and transfer of information within the practice:

"*It's easy to get at; you don't have to open other programmes.*" Midwife, female, inspected by inspectorate

Many midwives started digital registration when it became available. One midwife said that digital registration was not possible because the computer crashed when she tried using it.
For external referrals, midwives depend on other organisations. Midwives reported that they had created a 'social map', which displays the external organisations to which they can send referrals. An example of referral options is individual coaching by care providers from mental healthcare practices, the GP or a practice nurse. They felt this was an improvement. The referred more clients once they had made a social map. However, the availability of external referral options changes from time to time. Some programmes that provide counselling ended or the coverage of the insurance company changed, which made a programme unattractive for pregnant smokers. This led to an unclear situation whereas keeping a social map up-to-date requires continuous time and effort:

"*It's not so clear, here. Yes, we always do* (*refer to the*) GP, *if necessary, but more than that*? I *know some regions have special lung clinics for outpatients who do something, but we don't have that listed here* (*on our social map*)." Midwife, female, not inspected by inspectorate, non-smoker

This midwife suggests other improvements to help pregnant smokers, but she is unable to arrange them. An important barrier is restricted time. Although midwives who provided counselling stated that it did not cost them much time, the midwives who did not provide counselling indicated that it would cost a lot of time. They believed that other care providers have more time for it.

"I *imagine it will take far more time. We see* 400 *pregnant women a year. If you spent one hour… Although, not all smoke.* (…) It (*still*) *costs a lot of time.*" Midwife, male, not inspected by inspectorate, non-smoker

Despite the fact that the counselling is effective, midwives have to invest extra time in caring for pregnant smokers. That extra time costs money and leaves less time for other important problems. These barriers are related to external factors, such as money restrictions. In addition, the influence midwives have on pregnant smokers is limited:

"*We* (*the Dutch government*) *banned the cigarette from pubs and restaurants, and that's great, it's a huge improvement, but now you still see people smoke outside.* (…) *If you smoke and if health insurance companies have a say in it.* (…) A *bonus, simple as that: if you don't smoke, you get a bonus* (*discount on your premium*) *from your insurance company. That's a good idea, we should do that!*" Midwife, female, inspected by inspectorate

This midwife feels that national policy measures have more effect on smoking than her own efforts

and that influences her counselling. She provides counselling, because it is a good way of getting pregnant women to quit smoking. However, some midwives doubt the impact on the national scale.

Professional collaborations

In addition to the Cabana model[16], we found that professional collaboration is an important factor in guideline adherence. Midwives indicate that it was easy to change the smoking-cessation counselling if the practice changed their composition of midwives. Midwives said that regional collaboration between midwives and in multidisciplinary obstetric partnerships led to improvements in counselling. A colleague's recommendation was reason enough to take the training course. The agreements made in these collaborations increased the motivation to adhere to these agreements. All practices had to formulate a protocol and sometimes they worked together on the draft protocol. If one practice wrote a protocol, other practices used it as well:

"*We often try to work together and combine our efforts*. (…) A*nd if somebody writes a policy document, we can all use it*." Midwife, female, inspected by inspectorate

Although practices are partly competitors, they try to cooperate where possible. These cooperation practices can save them time and money and they can learn from each other. The practices that provided less counselling had no collaboration in-house or a multidisciplinary obstetric partnership. These midwives said they worked very much as individuals in the practice. Sometimes there was even no collaboration with the professional organisation. In some cases, the multidisciplinary obstetric partnership with the hospital paid no attention to counselling:

"*The gynaecologists just say: we don't have time for that*." Midwife, female, inspected by inspectorate

Midwives find it demotivating when the gynaecologists have no time for smoking-cessation counselling. The midwives think that gynaecologists do not see the importance of counselling.
In the previous quotes it is apparent that professional collaboration makes it easier to follow the guideline. Practices do not have to do everything by themselves; collaboration makes guideline adherence efficient and they can avoid delivering lower quality of care than surrounding practices. The professional collaborations also played a role in the decision to take counselling training. Because the professional midwifery organisation and colleagues recommended this course, some midwives actually took it. However, it also works the other way around. Lack of professional collaboration inhibits guideline adherence. For example, the gynaecologists' negative attitude, that they have no time for counselling.

The Cabana model[16] also mentions 'view on guidelines in general', 'view on this specific guideline' and 'patient factors' as important factors

for guideline adherence. However, our participants did not specifically mention these factors.

Responsive regulation

The theory of responsive regulation suggests that if the ultimate measure is known, the care provider will be more inclined to follow the guideline.[17] One midwife said she was afraid of the inspectorate inspecting her practice:

"I *was scared that*. (…) I *thought that you wouldn't get your money back again and it all was so compulsive*." Midwife, female, not inspected by inspectorate

Despite the fact that the midwives in our study knew perfectly well about the ultimate measure, not all adhered to the guideline. An explanation for this contrary result might be that in the Netherlands no midwifery practice was ever closed. Although the midwives knew about the power of the inspectorate to close a midwifery practice, none had experienced the ruling put into force.

The guideline was ambiguous about the obligation to follow the guideline. For example, the practices should have a 'social map' and be trained in counselling. But the guideline provides no information on the required timeframe for updating the social map or taking training courses. Zuiderent-Jerak[19] described the lack of clarity on the guideline obligations, which can lead to reduced guideline adherence. Care providers will follow the guideline more often if the obligation to do so is clear. This is also related to responsive regulation: supervision is complicated if it is not clear when a rule is violated and the inspectees do not know when a regulator can sue them.

DISCUSSION

Summary of main findings

Guideline adherence depends on several factors. In our case study, awareness and familiarity with the guideline and supervision programme were important, as was outcome expectancy. We discussed extensively motivation, guideline factors and environment factors. Besides these previously documented factors, professional collaboration also determined guideline adherence. More collaboration in counselling is associated with more guideline adherence, such as provision of counselling and taking the required training. The supervision programme contributed to improvements in stop-smoking counselling, making midwives aware of the counselling and giving an extrinsic motivation to provide counselling.

Strengths and limitations

The design of this study has several strengths and limitations. One strength is that our sample of midwives was a mixed group differing in age, work experience, practice situation and whether they smoke. This enabled us to record many opinions on smoking-cessation counselling and inspection. A second strength is that the midwives in our sample were exposed to various kinds of supervision. Some were inspected

and others were not, also in different rounds. Therefore we could also describe the potential effects of supervision on uninspected practices. The last strength is that our data collection included the whole supervision programme. This resulted in information on all aspects of the programme.

A limitation of the study is that it is based solely on interviews. We could not check the answers on social desirability because we did not observe the midwives at work or analyse documents. However, we obtained extensive information from the interviews on why the midwives did what they did. In observations and document analysis you can only see what they do and not retrieve any information on the 'why'. In addition, we interviewed midwives only, and no other health care professionals involved with pregnant women or inspectors. However, besides the interviews, we did collect additional data on the supervision programme.

A second limitation is the possibility of recall bias. Some midwives were inspected about three years before the interview took place. This intervening period might have been too long to provide enough insights in motivations of their actions at the time.

The final limitation is that our study contained only one case of supervision. Therefore it is difficult to generalise our results to other supervision programmes. When comparing midwives to general practitioners, we believe that the profession of midwife is closely related to the GP's. Both care providers are situated in the neighbourhood, close to their patients. Their types of practice are comparable; both private (individual) and with employees are possible. The referral options for smoking-cessation counselling are also comparable. Initial training for GPs is much longer than for midwives, but their post-initial requirement training is comparable (200 hours every five years). All this indicates that our case is generalisable to other primary care providers, for example GPs.

Interpretation

This study followed a previous quantitative study which found that V-MIS use increased substantially from 28% in 2010 to 80% in 2012.[7] After our current qualitative study, we can conclude that this improvement is related to the supervision programme. However, it is not fully attributable to the supervision programme, since other stakeholders also played roles in the improvement. Our combination of study methods provides additional knowledge on the effectiveness of the supervision programme and the factors that contribute to guideline adherence.

Our study found two groups of midwives who were intrinsically and extrinsically motivated to adhere to the quit-smoking guideline. The intrinsically motivated midwife acts at once when she hears of opportunities to improve counselling. For example, when she hears about training courses, she immediately signs up, whereas an extrinsically motivated midwife first needs an external motivation, such as advice

from a professional organisation or the force of the inspectorate. Most midwives want to follow guidelines and the advice from the inspectorate, if they aim at a common goal. Although an extrinsically motivated midwife also follows the guideline, the lack of intrinsic motivation carries a risk. If the external motivation is omitted, the quality of her counselling might deteriorate to the old level.

In our analysis we focused on whether improvements resulted from the supervision programme. Despite this focus, we also encountered deterioration in quality of care caused by, for example, a trained midwife leaving the practice or the counselling was provided less strictly to pregnant smokers. Since the inspectorate took a large part of the responsibility to improve counselling, it could be that midwifery practices feel less responsible for ongoing improvements to counselling. Following the supervision programme, responsibility should be returned to the practices to enhance the self-regulatory capacity of the midwives. Our data collection took place 18 months after the deadline for guideline adherence imposed by the inspectorate. Even in this short time, we noted some deterioration in guideline adherence. As this deterioration began quite soon, it is important to prevent further deterioration in the future.

Implications and general conclusion

Further research should investigate whether the conclusions of our study are valid in other supervision practices. Studies with additional participating observations or document analysis in care practices might be useful to obtain more insight and evidence on whether supervision programmes are effective and how they work. Furthermore, the focus can be extended towards deterioration in quality of care after the supervision programme has ended.

In this study we found that the combination of methods used to distribute the supervision programme was successful. Future programmes should also aim at using multiple elements in each supervision programme to reach all targeted care providers.

Most of the factors determining guideline adherence found in this study were in line with Cabana's model.[16] Additionally, we found that professional collaboration also has an impact on guideline adherence. Therefore, we recommend considering professional collaborations when attempting to improve or measure guideline adherence.

As we also found obstacles that inhibited improvements in counselling, we recommend giving attention to causes of deterioration in quality of care. This attention should not necessarily come from the inspectorate, because care providers are also responsible for improving quality of care. However, the inspectorate can monitor whether the profession is paying attention to this issue.

Both the professional organisation of midwives and the inspectorate are seen as policy makers.

Collaboration between them strengthens the message of the importance and requirements of quit-smoking counselling.
In conclusion, we explored factors related to guideline adherence. Motivation and environmental factors were the most important and we added professional environment as a significant factor for guideline adherence. Improved adherence is partly attributable to the supervision programme.

REFERENCES

1. Lanting CI, Van Wouwe JP, Van Den Burg I, et al. Smoking during pregnancy (Roken tijdens de zwangerschap) *Ned Tijdschr Geneeskd* 2012;156(50):2123
2. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* 2004;6(Suppl 2):S125-S40.
3. Shaw C. External assessment of health care. BMJ 2001;322(7290):851.
4. Boyne G, Day P, Walker R. The evaluation of public service inspection: A theoretical framework. *Urban Studies* 2002;39(7):1197-212.
5. De Vries H, Bakker M, Mullen PD, et al. The effects of smoking cessation counseling by midwives on Dutch pregnant women and their partners. *Patient Educ Couns* 2006;63(1):177-87.
6. Stop Smoking Partnership (Partnership Stop met Roken). Guideline treatment of tobacco addition revision 2009 (Richtlijn Behandeling van tabaksverslaving herziening 2009). Den Haag, 2009.
7. Oude Wesselink SF, Lingsma HF, Reulings PGJ, et al. Does Government Supervision Improve Stop-Smoking Counseling in Midwifery Practices? *Nicotine Tob Res* 2014:ntu190.
8. Royal Dutch Organisation of Midwives. Midwifery in the Netherlands 2012. In: M. de Geus, F. Cadée, editors. Utrecht, 2012.
9. Wiegers TA, Warmelink JC, Spelten ER, et al. Work and workload of Dutch primary care midwives in 2010. *Midwifery* 2013.
10. Dutch Healthcare Authority (Nederlandse Zorgautoriteit). Midwivery (Verloskunde).
11. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983;51(3):390-95.
12. Mohangoo AD, Buitendijk SE, Hukkelhoven CWPM, et al. High perinatal mortality in the Netherlands compared to other European countries: Peristat II study (Hoge perinatale sterfte in Nederland vergeleken met andere Europese landen: de Peristat-II-studie). *Ned Tijdschr Geneeskd* 2008;152(50):2718-27.
13. Reulings PGJ, van der Lans SMGA, Wentzel HR, et al. Grab your chance, smoking cessation couseling by midwives (Pak je kans: stoppen met roken begeleiding door verloskundigen). *Tijdschrift voor Verloskundigen* 2011;april.
14. Dutch Healthcare Inspectorate (Inspectie voor de Gezondheidszorg). Measures (Maatregelen), 2014.
15. Dutch professional midwifery organisation, Dutch expertise centre on tobacco control, (Koninklijke Nederlandse Organisatie van Verloskundigen, et al. Guide for policy smoking cessation counselling in midwife practices (Handreiking Ontwikkeling Beleid voor Begeleiding Stoppen met Roken in Verloskundigenpraktijken), 2012.
16. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;282(15):1458-65.
17. Braithwaite J, Makkai T, Braithwaite VA. *Regulating aged care: Ritualism and the new pyramid*: Edward Elgar Publishing, 2007.
18. Glaser BG, Strauss AL. *The discovery of grounded theory: Strategies for qualitative research*: Transaction Publishers, 1967.
19. Zuiderent-Jerak T, Jerak-Zuiderent S, Bovenkamp van de H, et al. Variation in Guidelines, what is the problem? (Variatie in richtlijnen wat is het probleem?). The Hague: Institute of Health Policy & Management, 2011.

APPENDIX 1

Dutch Healthcare Inspectorate

The Dutch Healthcare Inspectorate (Netherlands, EU) is an independent agency of the Ministry of Health, Welfare, and Sport. It aims to improve population health and is expected to efficiently supervise a sector where 1.3 million people work for 40,000 institutions and companies. Its primary instruments are advice and encouragement. If these do not achieve the desired result, it can implement corrective action by, for example, increasing the supervision or by limiting the ability to practice a profession. The inspectorate enforces 25 laws, including for example the Care Institutions Quality Act. The supervision is performed by using a combination of three methods:

1. theme-based supervision, directed at specific issues in care, which are sometimes requested by the minister or parliament
2. supervision in response to calamities or emergencies that indicate structural shortcomings in care provision
3. risk-based supervision to assess the quality of healthcare by means of indicators.

The programme described in this study is an example of theme-based regulation.

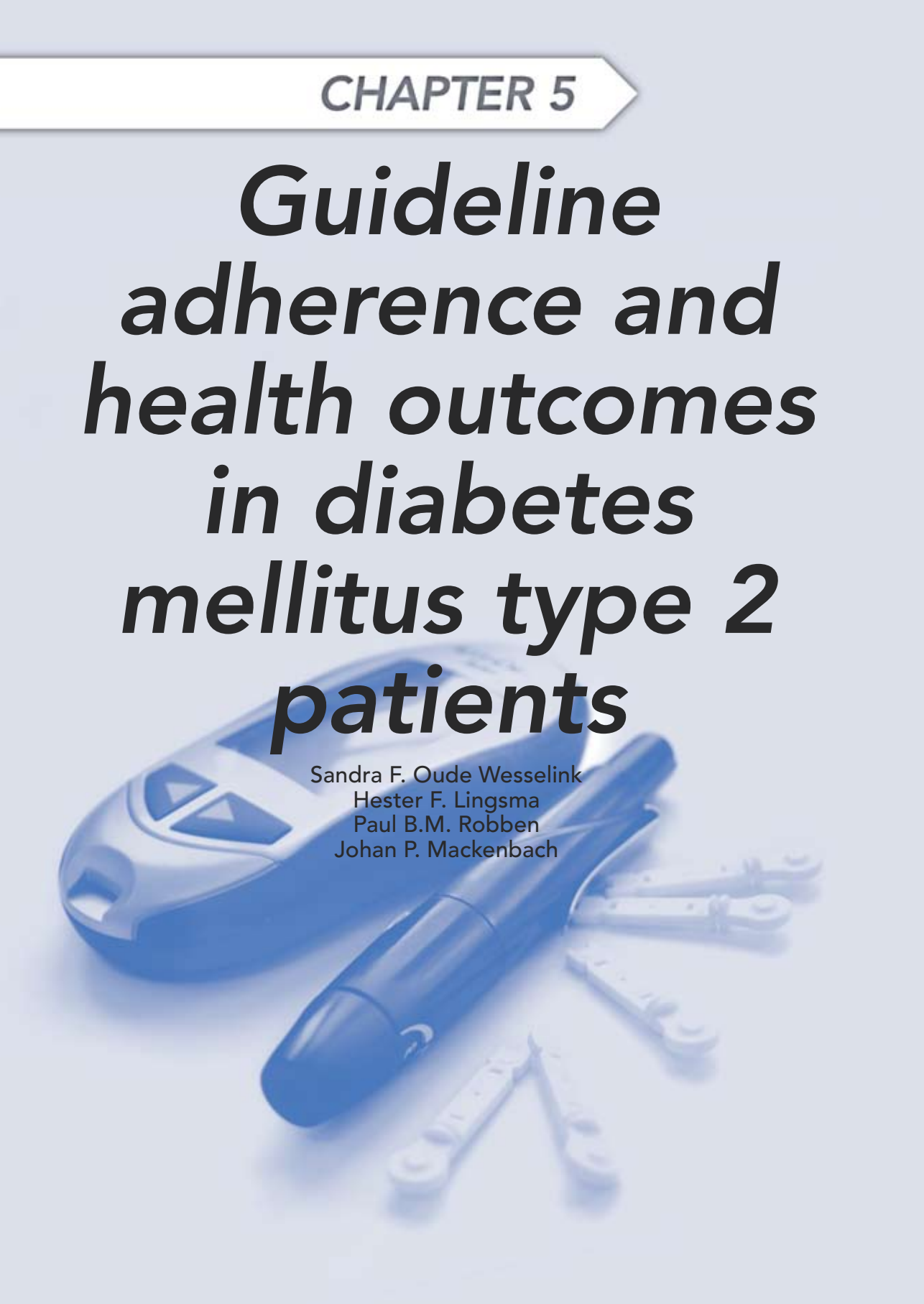
PART II

Relation between quality of care indicators in supervision programmes

CHAPTER 5

Guideline adherence and health outcomes in diabetes mellitus type 2 patients

Sandra F. Oude Wesselink
Hester F. Lingsma
Paul B.M. Robben
Johan P. Mackenbach



ABSTRACT

Introduction

The complex disease of diabetes mellitus type 2 (T2DM) requires a high standard of quality of care. Clinical practice guidelines define norms for diabetes care that ensure regular monitoring of T2DM patients, including annual diagnostic tests. This study aims to quantify guideline adherence in Dutch general practices providing care to T2DM patients and explores the association between guideline adherence and patients' health outcomes.

Methods

In this cross-sectional study, we studied 363 T2DM patients in 32 general practices in 2011 and 2012. Guideline adherence was measured by comparing structure and process indicators of care with recommendations in the national diabetes care guideline. Health outcomes included biomedical measures and health behaviours. Data was extracted from medical records. The association between guideline adherence and health outcomes was analysed using hierarchical linear and logistic regression models.

Results

Guideline adherence varied between different recommendations. For example 53% of the practices had a system for collecting patient experience feedback, while 97% had a policy for no-show patients. With regard to process indicators of care, guideline adherence was below 50% for foot, eye and urine albumin examination and high (>85%) for blood pressure, HbA1c and smoking behaviour assessment. Although guideline adherence varied considerably between practices, after adjusting for patient characteristics we found guideline adherence not to be associated with patients' health outcomes.

Conclusions

Guideline adherence in Dutch general practices offering diabetes care was not optimal. Despite considerable variations between general practices, we found no clear relationship between guideline adherence and health outcomes. More research is needed to better understand the relationship between guideline adherence and health outcomes, specifically for guidelines that are based on limited scientific evidence.

Adapted from: Oude Wesselink SF, Lingsma HF, Robben PBM, Mackenbach JP. Guideline adherence and health outcomes in diabetes mellitus type 2 patients: a cross-sectional study. BMC Health Services Research, 2015.
http://www.biomedcentral.com/1472-6963/15/22/

BACKGROUND

The complex disease of diabetes mellitus type 2 (T2DM) requires a high standard of quality of care. The objectives of treatment are controlling glycaemia, blood pressure and blood lipid levels, improving lifestyle behaviour and reducing tobacco use.[1] These objectives are expected to lead to a reduction in the burden of diabetes and its complications. However, optimal treatment is not consistently implemented in clinical practice[2], one possible reason being inadequate adherence to clinical practice guidelines. Such guidelines may well help to improve care, because they specify optimal care for care providers and allow adherence to be monitored. Additional benefits from following guidelines include improving health outcomes, empowering patients, improving the quality of clinical decisions, supporting quality improvement activities, increasing efficiency, and identifying areas where there is insufficient evidence to support optimal care.[3]

A national guideline on integrated diabetes care in the Netherlands was formulated in 2007.[4] It was developed by the Dutch Diabetes Federation and therefore includes contributions from patients, practitioners and scientists. As well as aiming to improve outcomes of care and to reduce the costs of managing T2DM, it also contains general information about diabetes and recommendations on the content, organisation and quality of diabetes care. With a focus on integrated care, it contains instructions for the structure of diabetes care and recommends specific assessments and examinations. Because the guideline is only based on consensus, most of its recommendations are not evidence based. The consensus is based on opinions, on other guidelines and on legislation.

The extent to which care professionals adhere to the broad scope of this guideline is unknown. Previous studies have shown that the proportion of patients in which practices conducted annual measurements of HbA1c, blood pressure, LDL-cholesterol and urine albumin ranges between 49% and 86%.[2] However, for other quality-of-care indicators adherence is unknown.

Although following guidelines should theoretically improve health outcomes, evidence from empirical studies is mixed.[5-7] Previous studies relating guideline adherence to health outcomes have only tested a few specific elements of diabetes guidelines and do not provide conclusions about the overall effect of guideline adherence in T2DM patients. We therefore studied a range of structure and process indicators of care that are mentioned in the guideline.

This study aims to quantify guideline adherence in general practices providing care to T2DM patients in the Netherlands and explores the association between guideline adherence and patients' health outcomes.

METHODS

Study population

In the Netherlands, care groups are organisations that provide integrated diabetes care to patients in primary care (Box S1). These groups consist of 3 to 250 general practitioners, which are funded under a bundled payment system.[8] Care groups are similar to accountable care organizations.[9] From the approximately 100 care groups in the Netherlands, we randomly selected 33. Of these, 18 care groups participated and 15 refused participation (response rate 55%) (Figure S1). The reasons given for refusal were as follows: too busy with providing care to patients (n = 4), no compensation for time loss due to research (n = 2), do not agree with purpose of research (n = 1), currently involved in other research (n = 1) and unknown (n = 7).

Each participating care group selected one or two practices to participate in this study, based on their availability to participate in research. In total 32 practices participated in our study, together employing 32 practice nurses. Practice nurses either have a registered nursing degree or are practice assistants who have followed a two-year practice nursing degree.

Patient and practice data were collected from the general practices cross-sectionally between June 2011 and July 2012. For each practice, based on the schedule of last month, we randomly selected between 7 and 18 patients who had had a check-up in the last month. For these patients, data from the medical records was extracted by the practice nurses, together with the research assistants. In addition, the practice nurses were also asked to complete a questionnaire about guideline adherence.

No-one (care groups, practices or patients) received financial compensation for participating in this study. The local ethics committee of Erasmus University Medical Centre waived ethical approval for this particular analysis. A written informed consent was obtained from all participating practice nurses.

Study variables and definitions

The main outcome parameters of the study were guideline adherence in the practices and health outcomes in the patients. We also collected data on patient characteristics for use in the statistical analysis.

All variables were obtained from patient files and the most recent measurements were used. Health outcomes were BMI, systolic blood pressure, HbA1c, LDL cholesterol, urine albumin, glomerular filtration rate (GFR) and smoking behaviour. BMI (kg/m2), blood pressure (mmHg) and smoking behaviour (yes/no) were assessed in general practices and documented in patient files. Glucose (mmol/L), HbA1c (mmol/mol), LDL cholesterol (mmol/L), urine albumin (mg/L) and GFR (ml/min) were assessed in cooperation with diagnostic centres and documented in patient files. Measurements from before 2009 were not used, because these health outcomes were regarded as potentially outdated.

Guideline adherence was assessed by considering both structure and process indicators of care. Structure indicators, defined at practice level, were assessed by asking practice nurses whether they: have a system for collecting patient experience feedback, have regulations on access to patient files, have quarterly multidisciplinary meetings, have policies for checking medical equipment, have received training in self-management and have policies for no-show patients. Each question was given a numerical score of 1 (yes) or 0 (no) and these numbers were added to compose an aggregated score for structure indicators of care at each practice, ranging from 0 for the lowest to 6 for the highest quality of care in terms of structures (Table 1).

We defined process indicators at the patient level as being the previously mentioned annually measured health outcomes and obtained data from the patient records. A measurement was considered to be annual if the time period between two measurements was less than 366 days. For each practice we then calculated the proportion of patients in whom annual measurements had been done, as recommended by the guidelines. Once more an aggregated score for processes of care at each practice was composed, ranging from 0 for the lowest to 9 for the highest quality of care in terms of processes (Table 1).

For each practice, we analysed guideline adherence according to structure and process indicators of care. Each structure indicator was coded as yes/no per practice, resulting in an aggregated percentage of guideline adherence in all practices for each structure indicator. For process indicators we also calculated an aggregated percentage of patients per practice, resulting in an average guideline adherence across all practices for each process indicator.

The patient characteristics consisted of demographic factors and clinical factors. Demographic factors were age at data collection (years), sex and an indicator of socioeconomic status (SES). Clinical factors were years since diagnosis of T2DM (between diagnosis and the moment of data collection) and

Table 1 *Construction of guideline adherence scores, divided into structures and process indicators of care*

Structure score [1]	**Points**
System for collecting patient experience feedback	0 or 1
Regulations on access to patient files	0 or 1
Quarterly multidisciplinary meetings	0 or 1
Policies for checking medical equipment	0 or 1
Practice nurse trained in self-management	0 or 1
Policy for no-show patients	0 or 1
Score range	*0-6*
Process score [2]	
Annual assessment of BMI	0-1
Annual assessment of blood pressure	0-1
Annual assessment of HbA1c	0-1
Annual assessment of LDL cholesterol	0-1
Annual assessment of urine albumin	0-1
Annual assessment of GFR	0-1
Annual assessment of smoking behavior	0-1
Annual foot examination	0-1
Annual eye examination	0-1
Score range	*0-9*

[1] Structure score, composite score of different structure indicators of care, scoring: present = 1 point, absent = 0 points.

[2] Process score, composite score of different process indicators of care, the proportion of patients per practice that were tested annually for each indicator, range between 0 and 1.

comorbidities (defined using ICPC codes).[10] Comorbidities unrelated to T2DM were based on general national guidelines [11] while comorbidities related to T2DM were derived from the National Diabetes Guideline (Table S2).[4] All patient characteristics, except SES, were collected from patient files. The SES was based on the neighbourhood (postal code) of the general practice where the patient was treated. This score was obtained from a government agency (Netherlands Institute for Social Research)[12] and ranges from -10 to +10. A higher score represents a higher SES.

Statistical analysis

Guideline adherence was described at the practice level. Patient characteristics and health outcomes were described at both the patient and practice levels. The associations were evaluated using hierarchical linear and logistic regression models. In hierarchical models, the clustering of patients within practices is taken into account.[13]

Variation in guideline adherence between practices was described per indicator with interquartile ranges (IQR). The statistical significance of the variation between practices was tested in hierarchical regression models with the structure or process indicator of care as the dependent variable and a random intercept for the practices. Variation in health outcomes between practices was analysed with hierarchical regression models with the health outcome as the dependent variable and a random intercept for the practices.

Finally, to analyse the association between guideline adherence and health outcomes, adjusted for patient characteristics, we added to the model the aggregated score for structures or processes per practice as independent variable. The different health outcomes served as dependent variable. General practice was included in the model as a random intercept. Each health outcome was analysed separately and patients with missing outcomes (9% at most) were excluded from the analysis. Guideline adherence was analysed at practice level to avoid confounding by indication. Confounding by indication is a common problem in observational studies, where treatment is usually only given to patients who require it.[14] Without adjustment on practice level in the data analysis, appropriate treatment will always be related to poor health outcome, particularly in individual-level studies. By analysing adherence at the practice level, we can test whether practices with generally good guideline adherence have good outcomes.

The one covariate for which values were missing in about 50% of cases, namely year since diagnosis, was imputed with linear regression analysis based on seven covariates (sex, age, related and unrelated comorbidities, SES, structure and process quality scores).

From the regression models we derived beta scores (in the case of a continuous outcome) or odds ratios (in case of a binary outcome) and 95% confidence intervals. For regression analyses we used statistical software package SAS version

9.3 (PROC MIXED and PROC GLIMMIX) (SAS Institute Inc., Cary, NC) and for other analyses SPSS version 21.0 (IBM Inc., Somers, NY).

RESULTS

Thirty-two general practices participated in the study. We included between 7 and 18 patients per practice, resulting in 363 patients in total. Guideline adherence was assessed according to structure and process indicators of care (Table 2). Structure scores are expressed as the proportion of practices that answered positively to each of the questions about guideline adherence. For example, only 53% of practices reported that they systematically collect patient experience feedback. Adherence was also below 70% for regulations on access to patient files (63%) and for conducting quarterly multidisciplinary meetings (66%). For other structural aspects, guideline adherence was better: almost all practices had a policy for no-show patients (97%), a practice nurse trained in self-management (94%) and policies for checking medical equipment (91%). Overall, there was limited variation between practices for guideline

Table 2 *Description of mean scores for guideline adherence indicators in general practices (n = 32)*

	Guideline adherence (proportion of practices)	**IQR**	**p-value***
Structure score			
System for collecting patient experience feedback	0.53		
Regulations on access to patient files	0.63		
Quarterly multidisciplinary meetings	0.66		
Policies for checking medical equipment	0.91		
Practice nurse trained in self-management	0.94		
Policy for no-show patients	0.97		
Total structures of care score (points)	mean: 4.7 (scale 0-6)	4 - 5	1.00
	Guideline adherence (proportion of patients per practice)	**IQR**	**p-value***
Process score			
Annual assessment of BMI	0.7	0.41 - 1.00	0.00
Annual assessment of blood pressure	0.97	0.92 - 1.00	1.00
Annual assessment of HbA1c	0.91	0.85 - 1.00	0.00
Annual assessment of LDL cholesterol	0.59	0.42 - 0.76	0.00
Annual assessment of urine albumin	0.49	0.35 - 0.66	0.01
Annual assessment of GFR	0.65	0.48 - 0.78	0.00
Annual assessment of smoking behavior	0.89	0.86 - 0.98	0.00
Annual foot examination	0.33	0.12 - 0.49	0.00
Annual eye examination	0.28	0.10 - 0.43	0.03
Total processes of care score (points)	mean: 5.9 (scale 0-9)	5.3 - 6.6	1.00

Note. IQR, interquartile range; BMI, body mass index; HbA1c, glycosylated haemoglobin; LDL cholesterol, low-density lipoprotein cholesterol; GFR, glomerular filtration rate.
* Differences between practices, tested in hierarchical regression model, without other independent factors.

adherences according to structure indicators of care; the IQR of the total structure score was 4 to 5 (p = 1.00).
Guideline adherence measured according to process indicators of care, also shown in Table 2, is expressed as the proportion of patients receiving treatment according to the guideline. For example, 70% of patients were assessed annually for BMI and the IQR for the percentages per practice was 41% to 100%. Guideline adherence was below 50% for the assessment of urine albumin (49%) and for the examination of the feet (33%) and eyes (28%). However, guideline adherence was above 85% for blood pressure (97%), HbA1c (90%) and smoking behaviour (89%). The scores for the process indicators varied largely between practices: the IQR was often more than 30%. The IQR of the total process score was 5.3-6.6 (p = 1.00).

Table 3 shows the patient characteristics and health outcomes. Median age of the 363 patients was 65 years

Table 3 *Description of patient characteristics and unadjusted health outcomes for all patients (n = 363) and at practice level (n = 32)*

		Patients			Practices		
		n	median	IQR	median	IQR	p-value*
Patient characteristics							
Age (years)		363	65	58 - 74	66	63 - 70	0.02
Sex (% males)		363	49%		52%	43 - 58%	1.00
Years since diagnosis		175	6	3 - 9	7	5 - 9	0.04
Number of related comorbidities [1]		363	1	0 - 1	0.6	0.4 - 0.9	0.00
Number of unrelated comorbidities [2]		363	0	0 - 1	0.4	0.2 - 0.5	0.02
SES indicator [3]	*Low*	363	46%		44%		
	Middle		34%		34%		
	High		20%		22%		
Health outcomes							
BMI (kg/m²) [4]		343	29	26 - 33	29	29 - 31	0.10
Systolic blood pressure (mmHg) [5]		363	135	123 - 145	134	131 - 140	0.02
HbA1C (mmol/mol) [6]		361	50	45 - 56	51	48 - 55	0.01
LDL cholesterol (mmol/L) [7]		358	2.3	1.9 - 2.9	2.5	2.2 - 2.6	1.00
Urine albumin (mg/L) [8]		341	6	3 - 13	11	6 - 32	0.06
GFR (ml/min) [9]		356	68	60 - 89	74	63 - 86	0.00
Smoking (% smokers)		353	18%		17%	11 - 25%	1.00

Note. IQR, interquartile range.
* Differences between practices, tested in hierarchical regression model, without other independent factors.
ICC of patient characteristics, range: 0.07-0.16.
ICC of outcomes, range: 0.04-0.25.
[1] in total 12 ICPC codes.
[2] in total 63 ICPC codes.
[3] Score calculated nationally based on the postal code of the general practice where the patient was treated, low = lowest tertile, middle = middle tertile, high = highest tertile.
[4] Body mass index (range in this study: 19-53).
[5] Systolic blood pressure (range in this study: 95-197).
[6] Glycosylated haemoglobin (range in this study: 32-99).
[7] Low-density lipoprotein cholesterol (range in this study: 0.5-5.9).
[8] Urine albumin (range in this study: 0-1023).
[9] Glomerular filtration rate, higher is better (range in this study: 13-178).

and about half were males (49%). Median HbA1c was 50 mmol/mol and median GFR was 68 ml/min. The percentage of missing health outcomes was low; less than 7%. The study population was comparable to those of previous studies.[7,15] Several patient characteristics varied significantly between practices. For example mean age per practice had an IQR of 63 to 70 years (p = 0.02). Health outcomes also differed across practices. For example, the IQR of HbA1c was 48 to 55 mmol/mol (p = 0.01) and for GFR IQR was 63 to 86 ml/min (p = 0.00) (Table 3). The patient characteristics were not associated with guideline adherence, indicating that adherence was not better or worse in specific patient groups (data not shown).
When we explored our patient population further using regression models, we found several associations between patient characteristics and health outcomes (Table S1). For example, HbA1c was higher in patients who had a longer duration of diabetes (95% CI 0.0-0.4) or a higher number of related comorbidities (95% CI 0.2-3.2). Low GFR was associated with a higher age, meaning that kidney function worsens with older age (95% CI -0.9;-0.6).

Finally, we explored the associations between guideline adherence and health outcomes (Table 4). We found no clear relationships between guideline adherence and the health outcomes under study. The estimated effects point to both positive and negative associations between guideline adherence and health outcomes. Only systolic blood pressure was positively related to both structure and process indicators of care, but this relationship was not statistically significant.

DISCUSSION

Summary of main findings

Guideline adherence varied between different recommendations.

Table 4 *Associations between guideline adherence and health outcomes, analysed with hierarchical linear and logistic regression models (n = 363)*

	β	(95% CI)
BMI (n=343)		
Structure	-0.03	(-0.88;0.82)
Process	0.51	(-0.24;1.26)
Systolic blood pressure (n=363)		
Structure	1.70	(-1.08;4.47)
Process	1.97	(-0.46;4.39)
HbA1C (n=361)		
Structure	0.75	(-1.01;2.50)
Process	-0.64	(-2.19;0.90)
LDL cholesterol (n=358)		
Structure	-0.017	(-0.120;0.086)
Process	-0.013	(-0.107;0.080)
Urine albumin (n=341)		
Structure	-1.58	(-16.44;13.29)
Process	3.32	(-10.20;16.84)
GFR (n=356)		
Structure	4.54	(-0.39;9.47)
Process	-2.53	(-6.99;1.94)
Smoking* (n=353)		
Structure (OR)	1.13	(0.80;1.60)
Process (OR)	0.93	(0.68;1.28)

Note. CI, confidence interval; BMI, Body Mass Index (kg/m2); Systolic blood pressure (mmHg); HbA1c, Glycosylated haemoglobin (mmol/mol); LDL cholesterol, Low-density lipoprotein cholesterol (mmol/L); Urine in albumin (mg/L); GFR, Glomerular Filtration Rate (higher is better) (ml/min).
*Analysed with logistic regression: estimated odds ratio (OR) (smoker = 1, non-smoker = 0).
Structure: structures of care score per practice, see Table 1.
Process: processes of care score per practice, see Table 1.
All models control for age, sex, years since diagnosis, number of related and unrelated comorbidities and social economic status.

For example 53% of the practices had a system for collecting patient experience feedback, while 97% had a policy for no-show patients. With regard to process indicators of care, guideline adherence was below 50% for foot, eye and urine albumin examination and high (>85%) for blood pressure, HbA1c and smoking behaviour assessment. Although guideline adherence varied considerably between practices, after adjusting for patient characteristics we found guideline adherence not to be associated with patients' health outcomes.

Strengths and limitations

The study has several strengths and limitations. One strength is that we retrieved actual guideline adherence with regard to several indicators at the patient level, which is more precise than information collected at the practice level. A second strength is the method of data collection, as the research assistant collected the data from the patient files together with the practice nurse, our data were reliable with only a low number of missing values.

A limitation of this study was the substantial number of care groups that refused to participate. Of the 33 care groups that were approached, 15 refused to participate. Since it is possible that refusal is associated with poor guideline adherence, actual guideline adherence may be worse than that found in this study.

Another limitation of this study was the cross-sectional design, with guideline adherence and health outcomes being measured at the same time. As in practice it will take some time before better guideline adherence results in improved health outcomes, we have to assume that differences in guideline adherence between practices are relatively constant over time. Violation of this assumption, i.e. rapid changes in guideline adherence over time in individual practices, may have led to an underestimation of associations between guideline adherence and health outcomes.

Two arbitrary choices were made during data analysis, that may also have affected our results: firstly, we constructed the guideline adherence scores based on a comparable scientific paper,[5] and secondly, for the process indicators we used the number of days between check-ups to determine whether or not there was adherence to the guideline. In order to assess the robustness of the different choices made regarding these two issues, we performed a series of sensitivity analyses in which the scoring was adjusted as follows: (1) putting double weight on multidisciplinary meetings, policy for no-show patients and training in self-management, as these are directly related to patient care; (2) putting double weight on all indicators with the exception of the foot and eye examination; (3) giving three points instead of one for annual assessment of the specific outcome under study; (4) increasing the number of days between assessments considered to be as adherent from 366 to 400, 450 or 500 days. All four sensitivity analyses yielded the same results as those

found in the main analyses (data not shown).

Finally, when comparing our results with those of others, it should be taken into account that hard morbidity and mortality outcomes are frequently considered to be more important and more relevant to patients than the health outcomes used in this study. However, these endpoints, that include cardiovascular events and death, occur too infrequently in a general practice population[16] to be used as an outcome in our study.

A further consideration is that while the general practices were selected by the care groups, we do not expect selection bias to have occurred as the care groups were not aware of the aim of this study in advance. Furthermore, patients were selected at random.

Interpretation

Guideline adherence was suboptimal for several structure and process indicators of care. With regard to structure, a system for collecting patient experience feedback had not yet been widely implemented, although some practices had started a first investigation. Our finding that regulations on access to patient files also scored below 70% was likely due to the fact that such agreements often do not cover the care group as a whole, as stated by the practices nurses. Practices nurses reported also that quarterly multidisciplinary meetings did not cover all care providers of the care groups and that these meetings sometimes did not include discussion of individual patients.

With regard to process, guideline adherence was limited for half of the process indicators of care, with three scoring below 50%: assessment of urine albumin and examination of the eyes and feet. Eye examinations were always conducted outside the general practice and the practice did not always receive the patient's report after these examinations, as stated by practice nurses. In contrast, foot examinations were performed in the general practice, but often less than once a year. The fact that assessment of urine albumin scored lower than other laboratory assessments may well be due to patients not always supplying urine samples to the laboratory when necessary.

Since many quality indicators scored below 70% adherence, we conclude that guideline adherence is suboptimal in general practices offering diabetes care. Not only are our results in line with those of previous research,[2] but they also give insight into guideline adherence with respect to structure indicators of care. The fact that we found large differences in guideline adherence between practices is also consistent with previous studies in diabetes care.[17]

If we assume that guidelines describe optimal and evidence-based care, then variation in guideline adherence is undesirable. However, we found no relationship between guideline adherence and health outcomes. This finding is in line with those of other studies. For example, Ackerman et al found that in diabetes patients improvements in processes of diabetes care were not

associated with improvements in health outcomes.[18] Another study showed that improved processes of care were associated with an improved mental health score, but not with a physical health score.[5] A systematic review also concluded that structure and process indicators of diabetes care are largely unrelated to surrogate and hard outcomes.[19] While these previous studies tested just a few specific elements of diabetes guidelines, our results expand on these previous results and suggest that guideline adherence in general is not associated with health outcomes.

The guideline that we studied is not completely evidence based. While it is comparable to the NICE diabetes guidelines,[20] not every single element of the guideline has been underpinned with evidence.[21] This absence of evidence underlying some aspects of the guideline might be the explanation for the findings in our study – and in most previous studies – that adherence to elements of diabetes guidelines appears to have no effect on health outcomes. This is supported by the fact that, in other disease fields where the guidelines are more evidence-based, an association between adherence and health outcomes has been found.[22-24] In the case of T2DM, the lack of evidence for single elements of the diabetes guideline is at least partly due to the heterogeneity of the patient population, which implies that not all guideline recommendations apply to all patients. Additionally, while most recommendations concern the frequency with which a measurement should be done, e.g. measuring albumin once a year, they do not specify the actions that should subsequently be taken to improve outcome. However, the lack of a relationship between adherence and health outcomes might also partly be explained by the degree of self-management in diabetes: because it is a lifestyle-related disease, the role played by the patient is central. If the patient does not adapt his or her lifestyle to the disease, then his or her health will not improve.[25] Nevertheless, the lack of evidence underpinning clinical practice guidelines is a common problem in many disease fields[26] and cannot be ruled out as an explanation for our results.

The weak scientific evidence underpinning the guideline might also explain the relatively poor adherence: previous research has shown that lack of familiarity and lack of outcome expectancy can sometimes be barriers to guideline adherence.[27] Another possible explanation for our results is that Dutch general practices might use alternative guidelines.[28] Nevertheless, as the guideline we studies is the only guideline for integrated diabetes care in the Netherlands, it was this guideline that the Dutch Healthcare Inspectorate used to evaluate integrated diabetes care in 2011.[29] The Dutch Diabetes Federation, who developed the guideline, represents a broad spectrum of stakeholders and the guideline therefore is acknowledged by all professions involved in diabetes treatment.

Conclusion

In conclusion, guideline adherence in Dutch general practices offering diabetes care was not optimal. Despite considerable variations between general practices, we found no clear relationship between guideline adherence and health outcomes. For clinical practice, policy making and supervision it is important to consider the large variation in guideline adherence. While quality improvement initiatives might reduce the observed variation, our study suggests that better guideline adherence will not automatically lead to better health outcomes.

More research is needed to better understand the relationship between guideline adherence and health outcomes, specifically for guidelines that are based on limited scientific evidence.

REFERENCES

1. American Diabetes Association. Standards of medical care in diabetes - 2013. *Diabetes Care*;36 (suppl. 1):S11-S66.
2. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, et al. A longitudinal study examining adherence to guidelines in diabetes care according to different definitions of adequacy and timeliness. *PloS One* 2011;6(9): e24278.
3. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ 1999;318(7182):527.
4. Dutch Diabetes Federation (Nederlandse Diabetes Federatie). NDF Care Standard (NDF Zorgstandaard). Amersfoort 2007:44.
5. Harman JS, Scholle SH, Ng JH, et al. Association of Health Plans' Healthcare Effectiveness Data and Information Set (HEDIS) performance with outcomes of enrollees with diabetes. *Med Care* 2010;48(3):217-23.
6. Selby JV, Uratsu CS, Fireman B, et al. Treatment intensification and risk factor control: toward more clinically relevant quality measures. *Med Care* 2009;47(4):395-402.
7. Sidorenkov G, Voorham J, de Zeeuw D, et al. Treatment quality indicators predict short-term outcomes in patients with diabetes: a prospective cohort study using the GIANTT database. BMJ *Qual Saf* 2013;22(4):339-47.
8. de Jong-van Til JT, Lemmens LC, Baan CA, et al. The organization of care groups in 2011 (De organisatie van zorggroepen anno 2011). Bilthoven (NL): National Institute for Public Health and the Environment (RIVM), 2012.
9. Struijs JN, Baan CA. Integrating care through bundled payments—lessons from The Netherlands. N *Engl* J *Med* 2011;364(11):990-91.
10. Wood M, Lamberts H. International Classification of primary care: prepared for the World Organisation of national colleges, academies and academic associations of general practitioners/family physicians (WONCA) by the ICPC Working Party: Oxford: Oxford University Press, 1987.
11. Hoeymans N, Schellevis FC. Selection of chronic diseases (Selectie van chronische ziekten), 2008.
12. Netherlands Institute for Social Research (Sociaal en Cultureel Planbureau). Status scores.
13. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*: Springer Verlag, 2009.
14. de Koning JS, Klazinga NS, Koudstaal PJ, et al. The role of 'confounding by indication' in assessing the effect of quality of care on disease outcomes in general practice: results of a case-control study. BMC *Health Serv Res* 2005;5(1):10.
15. van Hateren KJJ, Drion I, Kleefstra N, et al. A prospective observational study of quality of diabetes care in a shared care setting: trends and age differences (ZODIAC-19). BMJ *Open* 2012;2(4).
16. De Grauw WJC, Van de Lisdonk EH, Hoogen HJM, et al. Cardiovascular Morbidity and Mortality in Type 2 Diabetic Patients: a 22-year Historical Cohort Study in Dutch General Practice. *Diabet Med* 1995;12(2):117-22.
17. Stone MA, Charpentier G, Doggen K, et al. Quality of Care of People With Type 2 Diabetes in Eight European Countries Findings from the Guideline Adherence to Enhance Care (GUIDANCE) study. *Diabetes Care* 2013.
18. Ackermann RT, Thompson TJ, Selby JV, et al. Is the number of documented diabetes process-of-care indicators associated with cardiometabolic risk factor levels, patient satisfaction, or self-rated quality of diabetes care? The Translating Research into Action for Diabetes (TRIAD) study. *Diabetes Care* 2006;29(9):2108-13.

19. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, et al. Review: relation between quality-of-care indicators for diabetes and patient outcomes: a systematic literature review. *Med Care Res Rev* 2011;68(3):263-89.
20. National Institute for Health and Care Excellence. Diabetes in adults quality standard. London, 2011.
21. Nagy E, Watine J, Bunting PS, et al. Do guidelines for the diagnosis and monitoring of diabetes mellitus fulfill the criteria of evidence-based guideline development? *Clin Chem* 2008;54(11):1872-82.
22. Dirks M, Niessen LW, Wijngaarden J, et al. The effectiveness of thrombolysis with intravenous alteplase for acute ischemic stroke in daily practice. *Int J Stroke* 2012;7(4):289-92.
23. Hoeks SE, Scholte op Reimer WJM, Lingsma HF, et al. Process of care partly explains the variation in mortality between hospitals after peripheral vascular surgery. *Eur J Vasc Endovasc Surg* 2010;40(2):147-54.
24. Kolfschoten NE, Gooiker GA, Bastiaannet E, et al. Combining process indicators to evaluate quality of care for surgical patients with colorectal cancer: are scores consistent with short-term outcome? BMJ *Qual Saf* 2012;21(6):481-89.
25. Steinsbekk A, Rygg L, Lisulo M, et al. Group based diabetes self-management education compared to routine treatment for people with type 2 diabetes mellitus. A systematic review with meta-analysis. BMC *Health Serv Res* 2012;12(1):213.
26. Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009;301(8):831-41.
27. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;282(15):1458-65.
28. Rutten GEHM, De Grauw WJC, Nijpels G, et al. Diabetes guideline of the Dutch College of General Practitioners (NHG-Standaard Diabetes mellitus type 2 (derde herziening)). *Huisarts Wet* 2013;56(10):512-25.
29. Dutch Health Care Inspectorate (Inspectie voor de Gezondheidszorg). Implementation care standard diabetes after four years of insufficient progress (Implementatie zorgstandaard diabetes na vier jaar onvoldoende gevorderd). Utrecht: Ministry of Health, Welfare and Sport, 2012.
30. de Bakker DH, Struijs JN, Baan CA, et al. Early results from adoption of bundled payment for diabetes care in the Netherlands show improvement in care coordination. *Health Aff (Millwood)* 2012;31(2):426-33.

5

APPENDIX BOX S1

Care groups

In the Netherlands, care groups are organisations that provide integrated diabetes care to patients in primary care. Care groups consist of 3 to 250 general practitioners, which are funded under a bundled payment system.[8] Bundled payment means that health insurance companies pay a single fee for all medical services involved in an episode of care;[30] in this case, paying a yearly fee for each diabetes patient in the care group. Care groups are the main contractor of a diabetes care program, and are responsible for the organisation, coordination and delivery of diabetes care.

Care groups consist of multiple health care providers and are often owned by general practitioners. Both general practitioners and practice nurses provide diabetes care within the care group, with practice nurses mainly performing check-ups for diabetes patients. Other care providers are contracted by the care group.

While care groups are similar to accountable care organisations,[9] accountable care organisations in the United States have a much broader scope, which includes hospital care. Legal requirements for care groups are far more extensive in the U.S. than in the Netherlands.

APPENDIX TABLE S1

Associations between patient characteristics and health outcomes, analysed with hierarchical linear and logistic regression models (n=363)

	β	(95% CI)
BMI		
Age (years)	-0.10	(-0.16;-0.05)
Sex (0=female, 1=male)	-0.61	(-1.65;0.43)
Years since diagnosis	-0.05	(-0.16;0.06)
Number of related comorbidities	0.84	(0.07;1.62)
Number of unrelated comorbidities	0.86	(0.02;1.69)
SES indicator (per practice)	0.41	(-0.21;1.03)
Systolic blood pressure		
Age (years)	0.20	(0.04;0.36)
Sex (0=female, 1=male)	1.47	(-1.78;4.72)
Years since diagnosis	-0.16	(-0.51;0.18)
Number of related comorbidities	5.86	(3.40;8.32)
Number of unrelated comorbidities	-0.71	(-3.29;1.87)
SES indicator (per practice)	1.10	(-0.95;3.14)
HbA1c		
Age (years)	-0.03	(-0.13;0.06)
Sex (0=female, 1=male)	-0.05	(-1.97;1.86)
Years since diagnosis	0.22	(0.02;0.42)
Number of related comorbidities	1.70	(0.24;3.16)
Number of unrelated comorbidities	0.37	(-1.16;1.91)
SES indicator (per practice)	-0.16	(-1.42;1.11)
LDL cholesterol		
Age (years)	-0.012	(-0.020;-0.004)
Sex (0=female, 1=male)	-0.270	(-0.438;-0.102)
Years since diagnosis	0.001	(-0.016;0.019)
Number of related comorbidities	-0.051	(-0.172;0.071)
Number of unrelated comorbidities	0.106	(-0.024;0.236)
SES indicator (per practice)	-0.033	(-0.110;0.043)
Urine albumin		
Age (years)	-0.60	(-1.53;0.33)
Sex (0=female, 1=male)	8.64	(-9.86;27.15)
Years since diagnosis	0.51	(-1.44;2.46)
Number of related comorbidities	17.27	(2.90;31.65)
Number of unrelated comorbidities	7.25	(-7.39;21.88)
SES indicator (per practice)	-1.62	(-12.40;9.16)
GFR		
Age (years)	-0.75	(-0.94;-0.55)
Sex (0=female, 1=male)	0.97	(-2.79;4.73)
Years since diagnosis	0.19	(-0.21;0.59)
Number of related comorbidities	-0.45	(-3.41;2.51)
Number of unrelated comorbidities	0.03	(-3.04;3.10)
SES indicator (per practice)	0.22	(-3.45;3.89)
Smoking*		
Age (years) (OR)	0.99	(0.97;1.02)
Sex (0=female, 1=male) (OR)	2.09	(1.18;3.70)
Years since diagnosis (OR)	0.93	(0.87;1.00)
Number of related comorbidities (OR)	1.24	(0.85;1.81)
Number of unrelated comorbidities (OR)	1.57	(1.05;2.35)
SES indicator (per practice) (OR)	0.95	(0.74;1.22)

Note. CI, confidence interval; BMI, body mass index; HbA1c, glycosylated haemoglobin; LDL cholesterol, low-density lipoprotein cholesterol; GFR, glomerular filtration rate (higher is better);
* Analysed with logistic regression: estimated odds ratio (OR) (smoker=1, non-smoker=0)

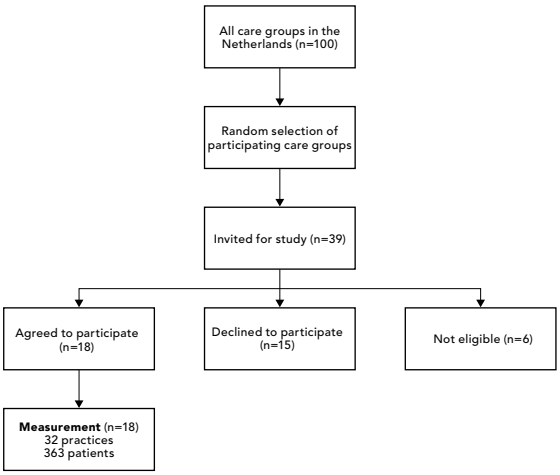
APPENDIX TABLE S2

List of comorbidities related and unrelated to diabetes

	ICPC code
Related comorbidities	
Angina pectoris	K74
Acute myocardial infarction	K75
Hypertension	K86 and K87
Transient ischemic attack	K89
Stroke	K90
Intermittent claudication	K92
Aneurysm aortae	K99
Diabetic neuropathy	N92
Depression	P03 and P76
Unrelated comorbidities	
Tuberculosis	A70
HIV/AIDS	B90
Cancer	A79, B72, B73, D74, D75, D76, D77, L71, N74, R84, R85, S77, T71, U75, U76, U77, W72, X75, X76, X77, Y77, Y78
Peptic ulcer	D85, D86
Ulcerative colitis	D94
Visual disturbance	F83, F84, F92, F93, F94
Hearing impairment	H84, H85, H85
Congenital heart defect	K73
Heart failure	K77
Chronic neck and back problems	L83, L84, L85, L86
Rheumatoid arthritis	L88
Osteoarthritis	L89, L90, L91
Osteoporosis	L95
Congenital neurological disorder	N85
Multiple sclerosis	N86
Parkinson's disease	N87
Epilepsy	N88
Chronic alcohol abuse	P15
Dementia	P70
Schizophrenia	P72
Anxiety disorder, other neurosis, PTSS	P74, P79
Anorexia nervosa	T06
Mental retardation	P85
COPD	R91, R95
Asthma	R96
Eczema	S87, S88

APPENDIX FIGURE S1

Flow chart of participating care groups



CHAPTER 6

Provision and effect of quit-smoking counselling by primary care midwives

Sandra F. Oude Wesselink
Hester F. Lingsma
Paul B.M. Robben
Johan P. Mackenbach

ABSTRACT

Objective

We aimed to evaluate the provision of quit-smoking counselling by midwives in the Netherlands and its effect on smoking behaviour and birth weight.

Design

Quasi-experimental study in which we collected information from pregnant women who smoke throughout their pregnancy by extracting data from electronic patient files.

Setting: Primary care midwifery practices

Participants: 851 pregnant women who smoke, treated between 2011 and 2014

Intervention: Quit-smoking counselling

Measurements and findings:

The midwives decided to provide quit-smoking counselling to the participant or not. Non-counselled women were used as the control group. The primary outcome parameter was quit smoking, defined as 'quit smoking by end of pregnancy'.
At intake, 67% of the women smoked 1–9 cigarettes a day, 23% smoked 10–20 cigarettes a day and 4% more than 20 cigarettes a day. The midwives began counselling with 42% of the participants, but seldom completed all the counselling steps. The average quit rate was 10% and average birth weight of the babies was 3200 grams. We found no difference in quit rate or birth weight between counselled women and those who were not. However, the data suggested that counselling is more effective when more steps of counselling are completed.

Key conclusions

No effect was found of quit-smoking counselling on quit-smoking rate or birth weight. Possibly, counselling is effective when provided extensively throughout pregnancy.

Implications for practice

Our study shows that provision of counselling can be improved.

Adapted from: Oude Wesselink SF, Lingsma HF, Robben PBM, Mackenbach JP. Provision and effect of quit-smoking counselling by primary care midwives. Midwifery, 2015. http://www.midwiferyjournal.com/article/S0266-6138(15)00187-4/abstract

INTRODUCTION

It is known that smoking during pregnancy is associated with a higher risk of foetal mortality and of adverse birth outcomes such as stillbirth, preterm birth, small for gestational age, intrauterine growth restriction, and congenital heart defects.[1-5] Nevertheless, 6% of women in the Netherlands smoke during pregnancy.[6] Among lower educated women, the prevalence of smoking during pregnancy is around 14%.[6] In the United States the smoking rate during pregnancy is around 10%[7] and in the United Kingdom 12%.[8] Further reduction of smoking during pregnancy is of major importance for reducing perinatal mortality.[9]

Quit-smoking counselling during pregnancy has been shown to reduce numbers of smokers in late pregnancy, low birth weight and preterm birth.[10] Nevertheless, the provision of counselling to pregnant women is often suboptimal.[11-13]

In the Netherlands, midwives play a central role in providing quit-smoking counselling for pregnant women who smoke, as they see 80% of all pregnant women at an early stage of pregnancy in their midwifery practices (Appendix 1). The counselling applies a minimal intervention strategy (V-MIS), based on the stages of change theory[14] and adapted for the midwifery care setting. V-MIS is effective in reducing smoking among pregnant women.[15] V-MIS comprises seven steps. In step 1, the midwife categorises the smoking behaviour of the woman and her partner. In step 2, the midwife tries to enhance the motivation to quit. In step 3, the midwife and woman discuss the barriers to successful quitting and how to mobilise social support. If the client is motivated to quit, they agree a quit date in step 4. In step 5, the midwife discusses additional self-help materials and gives them to the woman. In step 6, the midwife provides aftercare if required. In step 7, the midwife provides support to prevent relapse after delivery.[15]

A previous study on the implementation of V-MIS in midwifery practices found several differences between adherent users and non-adherent users.[11] Adherent users had significantly more knowledge about V-MIS, were more convinced of the importance of providing counselling, and were more likely to agree that counselling is a task of midwives. Adherent users had a more positive attitude towards the V-MIS and perceived less 'cons' than non-adherent users. Adherent users perceived more support from their social environment than non-adherent users. Finally, adherent users were more convinced that they had mastered the skills required to implement the V-MIS than non-adherent users.[11]

Multiple studies have addressed quit-smoking counselling of pregnant women who smoke.[10 16-19] As the method of counselling is not the same across studies, also the effectiveness of the studied interventions is different. Furthermore, the reduction in smoking during pregnancy during the last decade might also have changed the population of pregnant women who

smoke. The last time the effectiveness of V-MIS was studied, in a cluster randomised controlled trial (RCT) was more than 15 years ago.[15] Moreover, nothing is known about V-MIS implementation in clinical practice. As a result, it is uncertain whether this strategy is still effective.

Evidence for the provision and effectiveness of counselling would provide the opportunity to decrease numbers of pregnant women who smoke. Therefore we aimed to evaluate the provision of quit-smoking counselling by midwives and its effect on smoking behaviour and birth weight.

METHODS

Study design and population

Our quasi-experimental study included pregnant women who smoke, registered in primary care midwifery practices who had smoked at least one cigarette after entering the practice between February 2011 and November 2013. The midwives decided to provide quit-smoking counselling to the participant or not. Non-counselled women were used as the control group.

The participants were under treatment between 12 and 30 weeks of gestation at least, for at least 8 weeks in total. Women were excluded if they had no intake date or consultations registered. Those without health insurance were also excluded, since coverage is obligatory in the Netherlands. Finally, those with HIV, hepatitis B or syphilis were excluded from the analysis.

The midwifery practices in our study all worked with the same patient registration system and used electronic patient records (EPR) to register quit-smoking counselling. Since all the practices did not start using EPR at the same time for registration of counselling, the inclusion of participants began on different dates for each practice.

Data collection

Throughout the pregnancy we collected data retrospectively from the EPR. All data was registered by midwives as part of normal care. We retrieved participant information only from the EPR, meaning we gathered no additional information about individuals through other channels. However, we did collect additional information on the practice through phone interviews with the midwife responsible for quit-smoking counselling.

Participating midwives received a small compensation in the form of a gift voucher for 25 Euros on completion of the entire study procedure. The ethics committee of Erasmus University Medical Centre waived ethical approval for this analysis. We obtained informed consent from all participating midwives.

Study variables and definitions

The primary outcome parameter was quit smoking, defined as 'quit smoking by the end of pregnancy', which was self-reported in the last consultation. Answer categories were 'non-smoking', 'smoking less than 10 cigarettes a day', 'smoking between 10 and 20 cigarettes a day' and 'smoking more than 20

cigarettes a day'. The secondary oucome parameter was birth weight of the child, taken directly after birth. The studied intervention was quit-smoking counselling, based on the various steps of V-MIS. To assess the effect of applying V-MIS more or less extensively, we used different cut-offs to define counselling as yes/no.

Most study variables were retrieved directly from the EPR and therefore definitions depended on the midwives who filled in the patient records. Other variables were defined according to information extracted from the EPR. Growth retardation was defined as a score below the 5th percentile in an ultrasound of head circumference, abdominal circumference or femur length. Gestational hypertension was defined as before 20 weeks normotensive and after 20 weeks diastolic blood pressure equal or greater than 90 mmHg, or systolic blood pressure equal or greater than 140 mmHg. Pre-eclampsia was defined as hypertension and presence of albumin in laboratory urine test. Infectious diseases were defined as present when the EPR recorded a positive lab test. Diabetes was defined as present when a diabetes lab test above the cut off level of that laboratory was recorded. The socio-economic status (SES) of the participants was based on the status of the participants' home neighbourhood. This score was obtained from a government agency[20] and ranged from -10 to +10, where higher scores represent higher SES. Ethnicity was defined as Western when the woman was Dutch, European or other Western, and defined as non-Western when the woman was North African, other African, Turkish, South-Asian, East and South-East Asian, other Asian or other non-Western.

Statistical analysis

Characteristics of the clients and their pregnancy were described on the patient level. The provision of counselling was described on the practice level, to assess differences between practices. The effect of counselling on quit-smoking rate and birth weight was evaluated using hierarchical linear and logistic regression models with a random intercept for practices. Clustering of clients within practices is taken into account in hierarchical models.[21]

The dependent variables were quit smoking and birth weight. For the analysis of birth weight, we excluded women without partum. The independent variable was quit-smoking counselling (yes/no), defined using different cut-offs.

As this was a quasi-experimental study, the assignment of intervention was not random, but depended on client and practice characteristics. Therefore, we constructed a propensity score, which represents the probability of a woman receiving quit-smoking counselling. The propensity score was constructed in a regression model with all relevant client and practice characteristics as predictors. Subsequently, backward selection with a p-value of 0.10 was used to reduce the number of predictors in the propensity score. In the analysis, the effect of counselling was adjusted for the

individual propensity score, to correct for the likelihood of receiving counselling.

Missing values in baseline characteristics (body mass index, living situation, ethnicity, SES, drug use, alcohol consumption, psychosocial problems or violence, sexually transmitted diseases, and gestation) were imputed with linear regression analysis based on all relevant covariates.

We fitted a separate model for each study outcome and excluded cases with missing outcomes from the analysis. From the regression models we derived betas (for birth weight) or odds ratios (OR, for quit smoking) and 95% confidence interval. For regression analyses we used the statistical software package SAS version 9.3 (PROC MIXED and PROC GLIMMIX) (SAS Institute Inc, Cary, NC) and for other analyses we used SPSS version 21.0 (IBM Inc., Somers, NY).

Table 1 *Pregnant women who smoke in primary midwifery care (n=851)*

	Without counselling (n=645) mean (sd) / n (%)	With counselling (n=206) mean (sd) / n (%)
Anamneses		
Age (years)	27.3 (5.0)	27.6 (5.2)
Previous pregnancies	2.2 (1.4)	2.1 (1.3)
BMI	24.4 (5.0)	24.8 (4,5)
Living situation (married/cohabiting)	502 (83%)	173 (89%)
Ethnicity (Western)	512 (93%)	168 (93%)
SES		
Low	365 (57%)	102 (50%)
Middle	197 (31%)	74 (36%)
High	76 (12%)	28 (14%)
Drugs user	530 (9%)	18 (9%)
Alcohol consumption	627 (2%)	2 (1%)
Psychosocial problems or violence	205 (38%)	70 (34%)
Suffered from STD in the past	76 (13%)	25 (12%)
Pregnancy and post-delivery		
Week of first consultation	9 (3,9)	9 (3,2)
Week of last consultation	40 (4,6)	40 (4,2)
Weeks in consultation	30 (5,8)	30 (5,1)
Cigarettes at intake (per day)		
1-9 cigarettes	408 (63%)	158 (77%)
10 - 20 cigarettes	165 (26%)	34 (17%)
more than 20 cigarettes	27 (4%)	9 (4%)
Cigarettes during pregnancy (per day) [1]	4.7 (4.4)	4.5 (3.5)
Gestational hypertension	90 (14%)	38 (18%)
Pre-eclampsia	17 (3%)	1 (1%)
Gestational diabetes	30 (5%)	5 (2%)
Foetal growth retardation	48 (7%)	13 (6%)
Gestation (weeks) [2]	39.2 (2.5)	38.6 (4.1)
APGAR (5 minutes) [2]	9.6 (1.0)	9.2 (2.0)
Birth weight (gram) [2]	3210 (570)	3153 (702)

Note: Counselling defined as V-MIS step 1 till step 4

[1] Cigarettes during pregnancy: excluding clients who stopped smoking during pregnancy

[2] Gestation, APGAR, birth weight: excluding clients without delivery

BMI indicates Body Mass Index, SES = socio-economic status,

STD = sexually transmitted diseases

RESULTS

Population

In total, we included 851 participants (Table 1). Of them 206 (24%) were counselled up to step 4 of V-MIS at least. At intake, 67% smoked 1–9 cigarettes a day, 23% smoked 10–20 cigarettes a day and 4% more than 20 cigarettes a day. The mean age was 27 years and they had had 2.2 previous pregnancies. 9% were drug users, 2% drank alcohol and 37% has psychosocial problems or violence, at present or previously. The average quit-smoking rate was 10% and birth weight was on average 3200 grams. Clients entered the practice in week 9 of their pregnancy, on average. Appendix 2 gives detailed self-reported information about the practices in our study.
The response rate of the practices was 57% (Appendix 3). Reasons for refusal were too busy (n=6), unwilling to share EPR data (n=1) or unknown (n=5).

Provision of counselling

The midwives began quit-smoking counselling with 42% of the women (Appendix 4), and provided each consecutive step to fewer women. Although 10% of the women stopped smoking, only 5% of the women went through the last step of aftercare. Provision of counselling varied substantially between

Figure 1 *Provision of quit-smoking counselling per step of V-MIS (%), median (box) and interquartile range (lines), aggregated per practice*

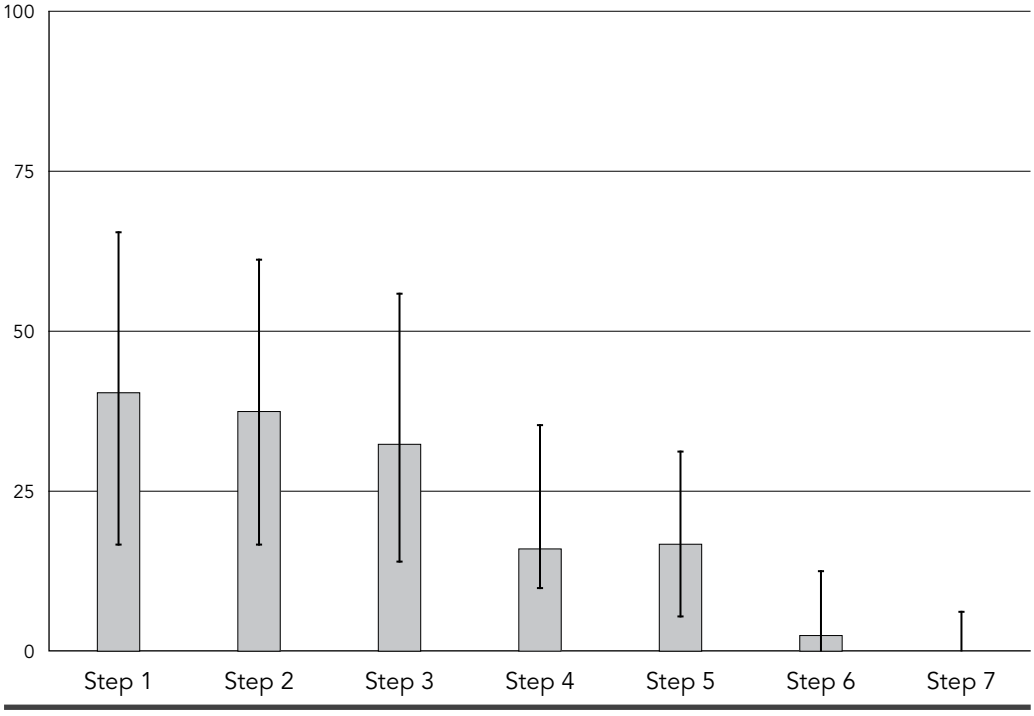
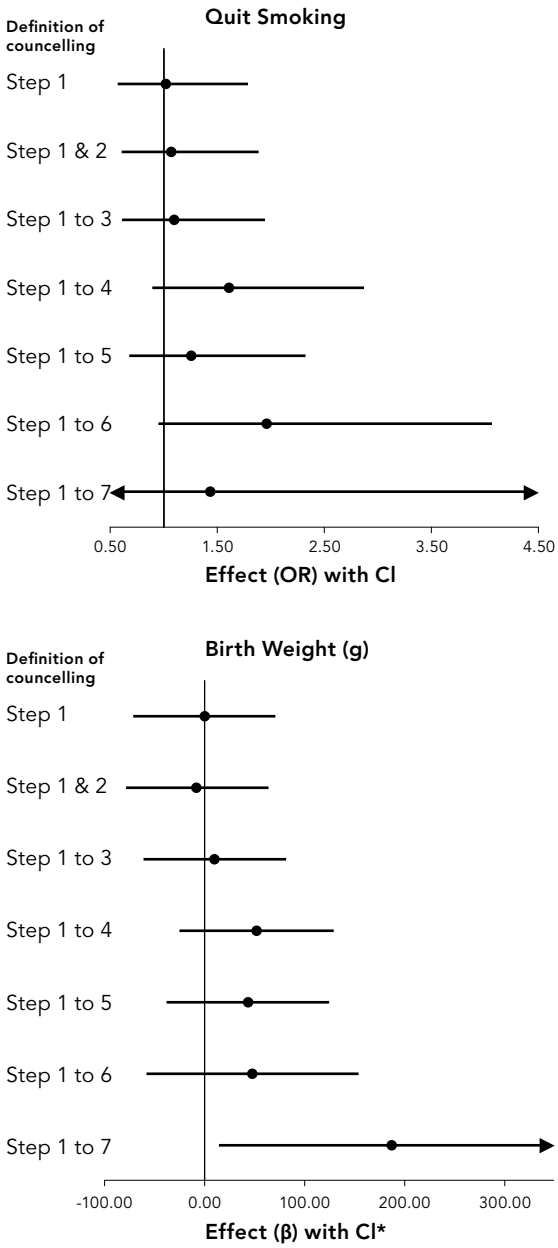


Figure 2 *Estimated adjusted effect of quit-smoking counselling on smoking and birth weight*



Note: Counselling defined on individual level
Corrected for propensity score on the provision of counselling (step 1 of V-MIS)
* Also corrected for current health status of the woman (twin pregnancy, fertility treatment, diabetes, hypertension, preeclampsia) APGAR and gestation

practices (Figure 1). The IQR (interquartile range) of initiation of counselling was 24–66% and decreased in the following steps. For step 7 the median was 0%, that is, more than half of the practices did not provide step 7 to any of their clients.

Association of client and practice characteristics and provision of counselling

We calculated a propensity score to estimate the likelihood that a woman would receive quit-smoking counselling (Appendix 5). Women were more likely to receive counselling if they were in a practice that treated more clients with psychosocial problems, more low-educated clients, more non-workers and more pregnancies per year. Women were less likely to receive counselling if they were in a practice that treated more non-Western clients, had more communication problems, older midwives, longer first consultations and longer quit-smoking consultations. No client characteristic was significantly associated with the likelihood of receiving counselling. This propensity score model had an R^2 of 0.42.

Effects of counselling

Independent of the definition, provision of counselling had no statistically significant effect on quit rate or birth weight (Figure 2). However, the magnitude of the effect increased when more steps of V-MIS were executed. For example, the quit rate in women who received only step 1 of V-MIS was 10.1%, compared to 10.3% in the control group (OR=1.02; 95% confidence interval

(CI)=0.58-1.78). Of the women who received the complete V-MIS, the quit rate was 16.0%, compared to 10.0% in the control group (OR=1.43; 95% CI=0.45-4.55) (Table 2). The same trend could be observed for birth weight: the effect of counselling increased when more steps of counselling were completed.

DISCUSSION

Summary of main findings

At intake, 67% of the women smoked 1–9 cigarettes a day, 23% smoked 10–20 cigarettes a day and 4% more than 20 cigarettes a day. The midwives began counselling with 42% of the participants, but seldom completed all the counselling steps. The average quit rate was 10% and average birth weight of the babies was 3200 grams. We found no difference in quit rate or birth weight between counselled women and those who were not. However, the data suggested that counselling is more effective when more steps of the minimal intervention strategy are completed.

Strengths and limitations

The design of this study had several strengths and limitations. One strength was the information on quit-smoking counselling per participant. To our knowledge, this was the first study with such extensive information about counselling in primary care midwifery practices. A second strength was the broad inclusion criteria, including all women who smoked at least one cigarette during pregnancy after entering the midwifery practice, regardless of other client characteristics.

A limitation of our study was the quasi-experimental study design. Midwives decided which women received counselling and which women not. However, we had a wealth of information on each participant

Table 2 *Quit-smoking counselling and corresponding percentage of smoking and birth weight (crude)*

	Definition of counselling	n	**Without** counselling (%)	n	**With** counselling (%)
1. Continuous abstinence (OR)	Step 1	494	10.1%	357	10.3%
	Step 1 and 2	510	10.0%	341	10.6%
	Step 1 to 3	533	9.9%	318	10.7%
	Step 1 to 4	645	9.3%	206	13.1%
	Step 1 to 5	687	9.8%	164	12.2%
	Step 1 to 6	761	9.5%	90	16.7%
	Step 1 to 7	826	10.0%	25	16.0%
2. Birth weight in grams (β) *	Step 1	475	3234	341	3144
	Step 1 and 2	491	3236	325	3137
	Step 1 to 3	512	3226	304	3147
	Step 1 to 4	618	3210	198	3153
	Step 1 to 5	658	3206	158	3157
	Step 1 to 6	730	3195	86	3213
	Step 1 to 7	791	3186	25	3514

Note: Counselling defined on individual level

and each midwifery practice to include in the propensity score to adjust for confounders. Propensity score adjustment is a robust method to remove bias due to all observed confounders.[22]

One possibly important determinant of the provision of counselling, intention to quit, was not present in the data and therefore, we were not able to correct for it in the propensity score. However, as the quit rate was 10% for both non-counselled and women in whom counselling was initiated, it appears that midwives did not select strongly on intention to quit.

A second limitation was that we used only registration data from the EPR. Previous studies have shown that registrations do not always contain all information about patients.[23] Loss of information could come from midwives not registering their findings, but also from lack of medical examinations and tests. For example, if a woman is not tested for an infectious disease, then that outcome will not appear in the EPR. The practices in our study employed at least 3 midwives. Therefore, the EPR was a very important tool for communication that enabled midwives to provide appropriate care. The practices reported using the EPR to register counselling. That is why all relevant information is likely to be recorded properly.

A third limitation was that we had to exclude many participants whose treatment duration in the midwifery practice was too short to provide counselling. Based on a priori power calculation, the analyses from step 4 onwards were underpowered. Another limitation was that we selected practices with many pregnant women who smoke, thereby excluding practices with few smokers. Practices excluded from our study were generally smaller, with fewer than 300 pregnant women entering the practice each year. They worked on a small scale and have fewer than 50 pregnant women who smoke entering the practice each year. This limits the generalizability of our findings to only larger practices.

Three arbitrary choices were made during data analysis that may also have affected our results: firstly we defined quit smoking as not smoking at the last consultation. Secondly, we assumed that women who were in treatment for at least 8 weeks had had enough time to complete quit-smoking counselling. And thirdly, we defined counselling on the individual level. To assess the robustness of the choices with regard to these three issues, we performed a series of sensitivity analyses in which the definition was adjusted as follows: (1) analyse the effect of counselling on number of cigarettes smoked during pregnancy; (2) defined 16 weeks of treatment as sufficient for counselling; (3) defined counselling on the practice level, using the proportion of participants counselled. All three sensitivity analyses yielded the same results as those in the main analyses (data not shown).

Interpretation

This study revealed that quit-smoking counselling by midwives was

not conducted as well as it could have been. Step 1 of the counselling was given to 42% of the women but this percentage decreased rapidly for the following steps. Although the V-MIS guideline prescribed that step 5 should be provided to every pregnant woman who smokes, only 23% actually completed this step. Therefore, we conclude that the steps of V-MIS were not strictly followed, a finding comparable to previous studies.[12 13 24] The initiation percentage was higher in these studies, but the last counselling steps were also not completed.

Due to our study design, we have no insight in why quit-smoking counselling was suboptimal. Previous studies reported on barriers to provide counselling and what predicted more extensive counselling. Barriers, reported by a qualitative study, were lack of time and communication skills, and fear of provoking resistance.[25] These results were also confirmed by other research.[26] Besides the implementation study[11], another study shows that strong control beliefs were predictive of a higher likelihood of engaging in more extensive interventions.[24] These control beliefs involve efficacy expectations (belief in capacity to conduct V-MIS) and outcome expectations (expectation that V-MIS will reduce smoking). Opportunities to increase counselling therefore might be to increase the control believes and communication skills, for example through regular training.

We found major differences between midwifery practices. Some initiated counselling in almost all pregnant women who smoke and others in only a very few smoking clients. Practice variation is undesirable, but does provide a window of opportunity to learn from better-performing peers.

We selected only practices that use the novel method of digital registration (an EPR) to record counselling. Practices that used a novel method could be considered as pioneers and might therefore have a higher quality of care than practices without digital registration. The poor quality of counselling we found may be an overestimation of the quality of counselling in midwifery practices in general.

In our analysis, we tried to predict counselling initiation. However, none of the characteristics in our analysis was statistical significantly related to the likelihood of receiving counselling. The predictive value of practice characteristics was also limited. Therefore, we conclude that initiation of counselling only partly depended on the pregnant woman who smokes and practice.

Our study found that counselling had no effect on quit-smoking rate or birth weight. The results suggested, however, that the more steps of counselling are completed, the more smokers will quit.

Most other studies promoting quit smoking during pregnancy showed positive effects. A study in South Africa on disadvantaged smokers showed more quitting and less smoking in the intervention group.[16] A systematic review of interventions for promoting quit smoking during pregnancy showed positive effects

on smoking in late pregnancy, and a reduction of low birth weight and preterm births.[10] In contrast, a meta-analysis on the effect of counselling solely, showed only small effects of counselling.[17] However, the type of counselling studied in this meta-analysis was less intensive than the type of counselling in our study.

V-MIS specifically was evaluated in a cluster RCT[15] undertaken in 1996. In that study, half the practices implemented V-MIS. They did not measure the actual provision of counselling, however, 12% of the intervention group and 3% of the control group quit smoking after pregnancy. In our study, we found that 12% of the counselled and 10% of the non-counselled women quit smoking. We can thus conclude that although the effectiveness of counselling remained the same, quit rates rose in non-counselled women. This may have reduced the additional effect of counselling compared to non-counselled women.

The major difference between the previous cluster RCT and our study was the design.[15] As V-MIS is used nationally[27], the contrast between counselled and non-counselled women was smaller than in the setting of a cluster RCT. Almost all midwives knew about V-MIS and most practices had one midwife trained in counselling with V-MIS. Another small difference was in the study population. Our population entered the midwifery practices earlier, drank less alcohol and had more previous pregnancies than those in the cluster RCT. In recent years, midwives encouraged women to enter the practice as early as possible, because studies have shown that the first months of pregnancy are very important for the development of the unborn child. The lifestyle of the women in these first months is crucial. However, the differences in study population are unlikely to have influenced our study results.

The same quit rate in the intervention and control group might also be explained by national attention to quit smoking, because perinatal mortality was higher in the Netherlands than in other European countries.[9] The Dutch Healthcare Inspectorate started a programme on pregnant women who smoke in 2010, which is continuing. In addition, the Netherlands Expertise Centre on Tobacco Control and the Netherlands Institute of Mental Health and Addiction took various opportunities to increase awareness and improve counselling methods, including a handbook, guideline and training courses.[27] These activities probably increased knowledge about counselling and the attention paid to pregnant women who smoke. These initiatives might have improved care to pregnant women who smoke in general, which could have contributed to the increased quit rate in the control group.

We did find a large significant effect of counselling on birth weight, when counselling was provided fully from step 1–7. The large difference in birth weight was not proportional to our results found in quit-smoking rate or number of cigarettes smoked a day after counselling. Therefore, we attributed this

finding to chance and not to the effect of smoking or counselling.

In our study, the participants who did not quit, smoked around 5 cigarettes a day. On average, people in the Netherlands who smoke, smoke 13 cigarettes a day.[28] Our study population smoked less compared to other pregnant women populations.[17]

Smoking during pregnancy has reduced in recent years.[6] Therefore, it was suggested that the pregnant women who smoke currently treated in midwifery practices are hard-core smokers[29] who never can quit smoking. As the percentage of quitters was the same as 15 years ago, we conclude that there are still possibilities for a reduction of pregnant women who smoke. In addition, the fact that quit rates did not differ between counselled and non-counselled women, indicated that midwives were not able to select for counselling those pregnant women who smoke and were able to quit with counselling. Therefore, counselling should be initiated for all pregnant women who smoke.

Several changes are possible to improve the effectiveness of quit-smoking counselling. One example from South Africa shows that social support from peer counsellors can be valuable for counselling.[30] Next, after text messages to support counselling[18], current progression of information and communication technology offers a window of opportunity to improve effectiveness of counselling. The increase in digital applications for mobile phones connected to the internet makes it possible to develop more personalised counselling methods.[31]

Besides counselling, also other strategies might help to reduce smoking during pregnancy. Recently, an RCT was performed on the effect of financial incentives for quit smoking.[19] Although the results are very promising, this intervention is still politically controversial in many countries.[32 33]

Implications and general conclusion

The provision of quit-smoking counselling was poor: counselling started with only 42% of the participants, and it varied largely between practices. The consecutive steps of V-MIS were not always executed, while our data suggest that V-MIS becomes effective when more strictly applied. Thus, our study showed that the number of pregnant women who smoke would decrease further if midwives initiated counselling for all pregnant women who smoke and applied the counselling strictly according to the V-MIS guidelines. Regular training for midwives might help to improve quit-smoking counselling for pregnant women who smoke.

Policy makers should draw more attention to quit-smoking counselling and provide more incentives for midwives to provide counselling to every pregnant woman who smokes. Policy should also focus on promoting more extensively provision of counselling.

In conclusion, no effect was found of quit-smoking counselling on quit-smoking rate or birth weight. Possibly, counselling is effective when provided extensively throughout pregnancy. Therefore, our results

do not suggest discouraging the use of V-MIS. In addition, our study shows that provision of counselling can be improved.

REFERENCES

1. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* 2004;6(Suppl 2):S125-S40.
2. U.S. Department of Health and Human Services. The Health Consequences of Smoking: 50 Years of Progress. A *Report of the Surgeon General*. A Report of the Surgeon General. Atlanta, 2014.
3. Leonardi-Bee J, Smyth A, Britton J, et al. Environmental tobacco smoke and fetal health: systematic review and meta-analysis. *Arch Dis Child Fetal Neonatal Ed* 2008;93(5):F351-F61.
4. Roelands J, Jamison MG, Lyerly AD, et al. Consequences of smoking during pregnancy on maternal health. *J Womens Health* 2009;18(6):867-72.
5. Yerushalmy J. The relationship of parents' cigarette smoking to outcome of pregnancy—implications as to the problem of inferring causation from observed associations. *Int J Epidemiol* 2014;43(5):1355-66.
6. Lanting CI, Van Wouwe JP, Van Den Burg I, et al. Smoking during pregnancy (Roken tijdens de zwangerschap). *Ned Tijdschr Geneeskd* 2012;156(50):2123
7. Centers for Disease Control and Prevention. Tobacco Use and Pregnancy, 2014.
8. Health and Social Care Information Centre. Statistics on Women's Smoking Status at Time of Delivery - England, 2014.
9. Mohangoo AD, Buitendijk SE, Hukkelhoven CWPM, et al. High perinatal mortality in the Netherlands compared to other European countries: Peristat II study (Hoge perinatale sterfte in Nederland vergeleken met andere Europese landen: de Peristat-II-studie). *Ned Tijdschr Geneeskd* 2008;152(50):2718-27.
10. Lumley J, Chamberlain C, Dowswell T, et al. Interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst Rev* 2009;3(3).
11. Segaar D. Adoption and Implementation of Smoking Cessation Support in Healthcare [PhD Thesis]. Maastricht University, 2006.
12. Price JH, Jordan TR, Dake JA. Perceptions and Use of Smoking Cessation in Nurse-Midwives' Practice. *J Midwifery Womens Health* 2006;51(3):208-15.
13. Sarna L, Bialous SA, Wells M, et al. Frequency of nurses' smoking cessation interventions: report from a national survey. *J Clin Nurs* 2009;18(14):2066-77.
14. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983;51(3):390-95.
15. De Vries H, Bakker M, Mullen PD, et al. The effects of smoking cessation counseling by midwives on Dutch pregnant women and their partners. *Patient Educ Couns* 2006;63(1):177-87.
16. Everett-Murphy K, Steyn K, Mathews C, et al. The effectiveness of adapted, best practice guidelines for smoking cessation counseling with disadvantaged, pregnant smokers attending public sector antenatal clinics in Cape Town, South Africa. *Acta Obstet Gynecol Scand* 2010;89(4):478-89.
17. Filion KB, Abenhaim HA, Mottillo S, et al. The effect of smoking cessation counselling in pregnant women: a meta-analysis of randomised controlled trials. BJOG 2011;118(12):1422-28.
18. Naughton F, Prevost AT, Gilbert H, et al. Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). *Nicotine Tob Res* 2012:ntr254.
19. Tappin D. Financial incentives for smoking cessation in pregnancy (CPIT): a randomised controlled trial. BMJ 2015.
20. Netherlands Institute for Social Research (Sociaal en Cultureel Planbureau). Status scores.
21. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*: Springer Verlag, 2009.
22. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70(1):41-55.
23. Arts DGT, De Keizer NF, Scheffer G-J. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc* 2002;9(6):600-11.
24. Lemola S, Meyer-Leu Y, Samochowiec J, et al. Control beliefs are related to smoking prevention in prenatal care. *J Eval Clin Pract* 2013;19(5):948-52.
25. De Wilde K, Tency I, Steckel S, et al. Which role do midwives and gynecologists have in smoking cessation in pregnant women?–a study in flanders, belgium. *Sex Reprod Healthc* 2015.

26. Cantin C, Peterson WE, Davies B, et al. Opportunities to Improve the Role of Family Practice Nurses in Increasing the Uptake of Evidence-Based Smoking Cessation Interventions for Pregnant Women: An Exploratory Survey. *Primary Health Care* 2014;4(174):2167-1079.1000174.
27. Oude Wesselink SF, Lingsma HF, Reulings PGJ, et al. Does Government Supervision Improve Stop-Smoking Counseling in Midwifery Practices? *Nicotine Tob Res* 2014:ntu190.
28. National Expertise Centre Tobacco Control (Nationaal Expertisecentrum Tabaksontmoediging). Factsheet Continuous Survey of Smoking Habits 2013 (Factsheet Continu Onderzoek Rookgewoonten 2013). Utrecht: Netherlands Institute of Mental Health and Addiction, 2014.
29. Warner KE, Burns DM. Hardening and the hard-core smoker: concepts, evidence, and implications. *Nicotine Tob Res* 2003;5(1):37-48.
30. Petersen Z, Nilsson M, Steyn K, et al. Identifying with a process of change: A qualitative assessment of the components included in a smoking cessation intervention at antenatal clinics in South Africa. *Midwifery* 2013;29(7):751-58.
31. Abroms LC, Padmanabhan N, Thaweethai L, et al. iPhone apps for smoking cessation: a content analysis. *Am J Prev Med* 2011;40(3):279-85.
32. Promberger M, Brown RCH, Ashcroft RE, et al. Acceptability of financial incentives to improve health outcomes in UK and US samples. *J Med Ethics* 2011;37(11):682-87.
33. Park JD, Mitra N, Asch DA. Public opinion about financial incentives for smoking cessation. *Prev Med* 2012;55: S41-S45.
34. Wiegers TA, Warmelink JC, Spelten ER, et al. Work and workload of Dutch primary care midwives in 2010. *Midwifery* 2013.

APPENDIX 1

Dutch midwifery care system

In the Netherlands midwives are medical practitioners with competencies restricted to independently provide care during 'normal' pregnancy, childbirth and the postpartum period. Midwifery education is a 4-year vocational training (at a university for applied science) at Bachelor's degree level. Women with an uncomplicated pregnancy are expected to receive care from a primary care midwife and are not required to consult a gynaecologist/obstetrician; the midwife is fully responsible for the care provided. But when complications arise, or threaten to arise, or the woman requests a form of pain relief that can only be given in secondary care, the midwife has to transfer responsibility by referring the woman to a gynaecologist. The midwife can choose to stay with her client or leave her in the care of a hospital midwife or nurse, but from the moment of referral onwards the gynaecologist is responsible for the care provided.

Primary care midwives work in private practice. They can work as self-employed practitioners or as an employee in someone else's practice. Self-employed practitioners work alone or in partnership with one or more other midwives. Many midwives also work as a locum, temporarily filling in vacancies in midwifery practices. Primary care midwives are paid per care unit, separately for pre-natal, natal and post-natal care. Consequently, if a client leaves the practice, that practice has a financial loss. *Adapted from* T.A. *Wiegers – Work and workload of Dutch primary care midwives in* 2010.[34]

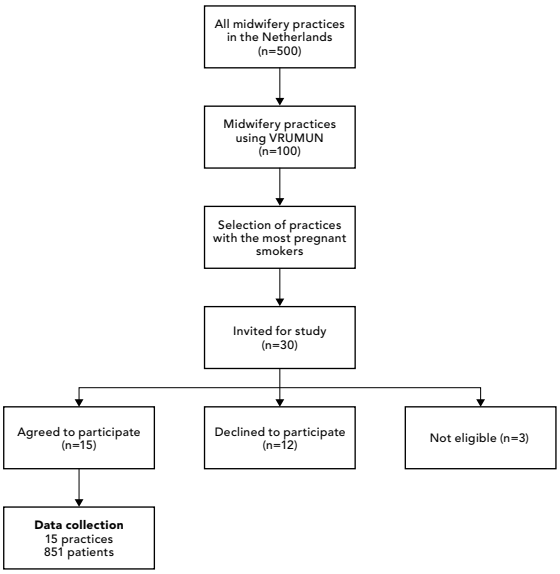
APPENDIX 2

Self-reported practice characteristics (n=15)

	mean / %	IQR
Clients		
Psychosocial problems	16%	10-20%
Low education	37%	20-55%
Non-workers	15%	10-17%
Non-Western	18%	5-25%
Language problems	5%	1-7%
Communication problems	8%	2-10%
Smokers[1]	23%	12-36%
Midwives and practice		
Midwives in practice	5	4-5
Sex midwives (female) [2]	99%	100-100%
Age midwives [2]	39	35-42
Working experience midwives (years) [2]	14	12-17
Smoking midwives [2]	10%	0-20%
Practice assistants in practice	1,7	1-2
Number of pregnancies (per year)	466	340-550
Practice type		
Solo	0%	
Duo	0%	
Group	93%	
Health centre	7%	
Urbanisation of clients [3]		
City	67%	
Village	67%	
Rural	47%	
Duration first consultation (minutes)	44	40-45
Duration regular consultation (min.)	14	15-15
Workload too high		
Always	13%	
Mostly	40%	
Sometimes	40%	
Never	7%	
Number of midwives per pregnant smoker	5	4-5
Duration smoking-cessation consultation (min.) [4]	14	8-10
Provide separate smoking-cessation consultation	27%	
Time for smoking-cessation counselling		
Sufficient	13%	
Mostly sufficient	33%	
Sometimes sufficient	47%	
Never sufficient	7%	
Importance of smoking-cessation counselling (0-10)	8,9	8-10
Educated in smoking-cessation counselling [2]	28%	0-67%
Policy in smoking-cessation counselling	60%	33-66%
Support in smoking-cessation counselling [5]	62%	33-100%

APPENDIX 3

Flow chart of participants



APPENDIX 4

Provision of quit-smoking counselling to participants per V-MIS step (n=851)

	mean	95% CI
Step 1	**42%**	**39-45%**
Registered smoking profile woman	41%	38-44%
Registered smoking profile man	34%	28-34%
Thought about quitting woman	38%	35-41%
Thought about quitting man	20%	17-22%
Pregnancy reason to quit woman	35%	32-38%
Pregnancy reason to quit man	13%	11-15%
Step 2	**40%**	**37-43%**
Motivation to quit	35%	32-38%
Discussed disadvantages of smoking and advantages of quitting	37%	34-41%
Giving advice to stop	37%	34-41%
Step 3	**38%**	**34-41%**
Investigate barriers to quit	33%	30-36%
Discuss specific barriers to quit	32%	29-35%
Step 4	**25%**	**22-28%**
Setting a quit date	13%	10-14%
Step 5	**23%**	**20-25%**
Provided information materials	22%	19-24%
Referral to other support possibilities	11%	9-13%
Step 6	**13%**	**11-15%**
Discussed smoking after agreed quit date	12%	9-14%
Discussed whether extra support is needed	3%	2-4%
Step 7	**5%**	**4-7%**
Discussed smoking after delivery	5%	3-6%
Provided information materials	5%	3-6%
Registered smoking behaviour woman	4%	2-5%
Registered smoking behaviour man	3%	2-4%

Note: CI = confidence interval
Each step scores as 'yes' if any action in that step was registered

APPENDIX 5

Propensity score: probability of a woman receiving quit-smoking counselling

	Association counselling (OR)	(95% CI)
Client		
Ethnicity (Western)	0.88	(0.77;1.02)
Week of first consultation	0.96	(0.92;1.01)
Week of last consultation	0.96	(0.92;1.00)
Practice		
Psychosocial problems* (practice)	1.07	(1.04;1.10)
Low education* (practice)	1.06	(1.04;1.08)
Non-workers* (practice)	1.17	(1.06;1.30)
Non-Western* (practice)	0.89	(0.85;0.94)
Communication problems* (practice)	0.95	(0.91;0.99)
Age midwives	0.82	(0.77;0.88)
Number of pregnancies per year	1.01	(1.01;1.01)
Duration first consultation in minutes	0.90	(0.86;0.93)
Duration smoking-cessation consultation in minutes	0.90	(0.84;0.97)

*Defined with percentages, OR indicates 1 percent point difference
Note: Counselling defined as completed step 1 of V-MIS
R^2=0.417

CHAPTER 7

Discussion government supervision in healthcare

The general aim of this thesis was to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes. Therefore, we evaluated two cases of government supervision: supervision on integrated diabetes care and supervision on quit-smoking counselling by midwives. We posed two research questions:

1. What are the effects of these two government supervision programmes on structures and processes of care as well as on health outcomes?
2. In these cases of government supervision, what is the relation between structures and processes of care, and health outcomes?

To answer these questions, we evaluated two supervision programmes of the Dutch Healthcare Inspectorate using different study designs: supervision on quality of diabetes mellitus type 2 care and supervision on quit-smoking counselling for pregnant women by midwives.

MAIN FINDINGS

Our first research question focussed on the effect of government supervision on quality of care and specifically health outcomes. In a randomised controlled trial (RCT) we evaluated the effect of supervision on the quality of integrated diabetes care, in terms of structures and processes of care and health outcomes. The intervention consisted of announcements of inspections, site visits and sending individualised reports with specific recommendations. We found no clear evidence that the supervision programme has an effect on quality. Although structures of care improved over time, these changes were similar in the intervention and the control group. Therefore the improvement could not be attributed to the supervision programme. Processes of care and health outcomes did not improve over time (chapter 2).

The supervision programme for quit-smoking counselling by primary care midwives consisted of announcing a deadline by which all practices were expected to comply with professional norms. In addition, this intervention also included supervision questionnaires to the midwives, site visits and sending individualised reports with specific recommendations. This supervision programme was evaluated in two studies. In an RCT we found that inspected midwives provided counselling more often compared to non-inspected ones. Both the control and intervention group largely improved over time (chapter 3). In a qualitative interview-based case study, we investigated whether the inspectorate may have contributed to this improvement. Midwives indeed indicated that the supervision programme influenced their provision of quit-smoking counselling. Therefore, we concluded that the supervision programme

of the inspectorate most likely led to at least some improvements in the provision of counselling (chapter 4).

Our second research question focussed on the relation between structures and processes of care and health outcomes. In integrated diabetes care, adherence to structures and processes of care described in the relevant guideline varied largely between the recommendations. The percentage of patients receiving adequate care ranged from 28% to 97% between specific guideline recommendations. Guideline adherence also varied considerably between practices. However, in this cross-sectional study we found no relation between adherence to structures and processes recommended in the guideline and health outcomes (chapter 5).

Guideline adherence in quit-smoking counselling was poor. Midwives started counselling with 42% of the pregnant smokers and only 5% of the women reached the last step of counselling. The results from our propensity score adjusted analysis suggested that better quit-smoking counselling, i.e. providing more steps of the counselling, led to more pregnant women quitting smoking (chapter 6).

METHODOLOGICAL CONSIDERATIONS

In this section we will first discuss the most important methodological aspects of the studies in this thesis and then how these aspects might have affected our results.

Design of studies evaluating supervision

To investigate the effect of the supervision programme on integrated diabetes care, we performed a cluster-RCT (chapter 2). An RCT is the gold standard in research to establish causal relationships. However, we found no effect of the supervision programme on quality of care. In retrospect, we believe that we have not succeeded in creating sufficient contrast between the intervention and control group. For example, some managers worked for both inspected and non-inspected care groups and there were national meetings of care groups that were attended by the inspectorate. As a result, the intervention and control group did not differ substantially in exposure to the intervention.

To test for the effect of the supervision programme on quit-smoking counselling (chapter 3), we combined different quantitative study designs. One element of the supervision programme, i.e. announcing a deadline by which all practices were expected to comply with professional norms, was implemented nationally. Therefore, a control group was not available and we used a before-after study design and an interrupted time-series design. In an RCT we evaluated the effect of site visits and questionnaires, because we could select a non-inspected control group randomly. Due to low statistical power, it was not possible to provide any evidence for the effectiveness of the site visits specifically. In our interview-based qualitative case study (chapter 4), we learned that midwives in the

control group were also exposed to the intervention. For example, the inspected midwives discussed their site visit extensively with other midwives in their area.
In addition, the qualitative study supported the findings on the effectiveness of supervision. However, the qualitative study design made it impossible to quantify the effect and limited the representativeness of the results for the whole population.[1]

Design of studies on the relation between structures and processes of care and health outcomes

In order to assess the relation between structures and processes of care and health outcomes for diabetes care, we performed a cross-sectional study, because no longitudinal data on the same patients was available (chapter 5). Such a cross-sectional design may provide insight into the relation between structures and processes of care and health outcomes. We adjusted in our analyses as much as possible in order to account for possible differences between the groups at baseline. However, the cross-sectional design cannot provide evidence of causal relations.
To assess the relation between structures and processes of care and health outcomes in quit-smoking counselling, we used propensity score adjustment, because the intervention was allocated by the treating midwife, which may introduce confounding by indication (chapter 6). With the propensity score we adjusted for possible differences at baseline between the groups. However, the differences between the groups in observed confounders appeared to be small, indicating that midwives did not systematically select specific patients for the intervention. Nevertheless, we have to be cautious in our interpretation of the results as the presence of unobserved confounders can never be excluded completely.

Timing of study measurements

In the RCT on supervision on integrated diabetes care, it was challenging to determine the timing of the study measurements. Looking back, a post-intervention period of on average five months was possibly too short to detect an effect of the supervision programme. Because of the complicated treatment of diabetes patients in care groups and the fact that improvements were initiated top-down, it may take much longer for the possible effects of supervision to be detectable. It was however not possible to extend the research period because other supervision activities were planned to start immediately after our study. This short follow-up time could explain why no effect of supervision was found in this case.

Complexity of intervention

The supervision programmes we evaluated are considered complex interventions. Complex interventions comprise a number of separate elements, which together form the intervention. However, it is often not possible to indicate the most important ingredient that

determines the effect of the intervention.[2] Therefore, we studied the supervision interventions in this thesis as a whole, without formulating separate conclusions about parts of the intervention.

Other characteristics of complex interventions are that the environment in which the intervention takes place is often not stable and different stakeholders are involved.[3] A complex intervention further may have a long time span. By the time the whole intervention is completed its effects may be diluted because there is also a general improvement in medical care over time. Often the perfect timing of study measurements does not exist and a balance should be found between the risk of measuring too early when there is yet no effect and the risk of measuring too late when the effect has already been diluted

Study measures

We selected our study measures based on statistical power, expected effect of the intervention and the relevance to population health.

From a statistical point of view, continuous outcome measures are more attractive than dichotomous measures, as the former have more statistical power. Hard endpoints, such as morbidity and mortality outcomes occurred infrequently in our study populations. Thus, using such endpoints would have resulted in analyses with low statistical power. Therefore we used mostly continuous study measures in the diabetes case, such as kidney functioning. These outcomes are relevant for population health, but in retrospect were perhaps not closely related to the intervention.[4 5] Although the supervision programme aimed at improving patients' health, it was directed at an organisational level instead of directly at patient care. For example, the inspectorate enforced the establishment of a patient complaints committee, which is important at an organisational level but the effect on patient care is expected to be small.

In the smoking study we used a dichotomous study measure, namely, whether or not the woman quit smoking. Importantly, this study measure was closely related to the intervention. Furthermore, if more pregnant smokers quit smoking, this will improve population health. However, statistical power of the analysis was reduced by choosing a dichotomous study measure. The choice for study measures thus is a trade-off between optimising the likelihood of detecting an effect of the intervention and the relevance for patients and population health. In supervision research it will often be unlikely to detect an effect on hard endpoints such as mortality that, although relevant for population health, are far down the causal chain started by the intervention of the inspectorate and occur infrequently in most diseases.

Response rate and sample size

In one of our studies problems with statistical power might have affected our conclusions, namely the study on the effect of counselling on quit smoking. Although we reached the required number of pregnant smokers, the group of counselled

women was small in the last steps of counselling. As a result, these analyses were underpowered and it was not possible to show a statistically significant difference between the intervention group and the control group. However, the data suggest that better counselling was associated with more of the women quitting smoking.

Summary of methodological considerations

In two of the studies presented in this thesis, the conclusions might have been affected by methodological shortcomings of the study. In our study on the effectiveness of the supervision programme for integrated diabetes care we encountered two methodological problems. First, the intervention and control group did not differ substantially in exposure to the intervention. Second, the post-intervention measurement might have taken place too soon after the inspectorate's visit and report, before the changes were implemented.
The study on the relation between structures and processes of care and health outcomes in quit-smoking counselling also had two methodological problems. First, some of the analyses were underpowered. Second, the quasi-experimental design with propensity score adjustment limited causal inference.

INTERPRETATION OF FINDINGS

The interpretation of the results will be discussed for each research question separately.

Research question 1 – Effect of government supervision on quality of care

In order to evaluate the effectiveness of supervision programmes, we studied two cases: supervision on integrated diabetes care and supervision on quit-smoking counselling by midwives.

Diabetes case

With regard to the effectiveness of supervision on integrated diabetes care, we did not find that supervision had an effect. Next to the difficulties we encountered in the study design and execution, the supervision programme itself also had some limitations. The enforced guideline was not evidence-based and the supervision programme was directed at an organisational level instead of directly at patient care. Besides that, the supervision programme was directed at risks instead of problems with a high potential for improvement, such as the adequate exchange of electronic patient files. Care groups have substantial problems regarding ICT, which the supervision programme was unlikely to solve. In summary, we found no evidence that the supervision programme improved quality of care or health outcomes. However, given the methodological shortcomings of our study design, we also cannot exclude the possibility that the supervision programme had an effect on structures and processes of care.
No previous studies were found which have generated empirical evidence of the effects of government supervision on quality of care

and specifically health outcomes. Therefore, we were unable to compare our results and conclusion with other studies.

Smoking case

To determine the effectiveness of supervision on primary care midwives, we combined evidence from quantitative and qualitative studies. In this case, we studied the effect of supervision on structures and processes of care. In the quantitative study we found that the group that was inspected improved quit-smoking counselling more than the control group. Furthermore, we found a significant improvement in midwifery practices in counselling during the supervision period, probably resulting from the national programme of the inspectorate. Through a qualitative study, we found that, according to the midwives, the supervision programme helped improve quit-smoking counselling, by making midwives aware of the counselling and giving them an extrinsic motivation to provide counselling. Thus, we conclude that the supervision programme probably contributed to the improved counselling. However, we were unable to quantify the exact contribution of the inspectorate and of the specific elements of the intervention.

Previous qualitative studies focussed on single elements of supervision, namely experiences of inspectees and inspectors with regard to site visits and reports.[6-10] Therefore, it was again impossible to compare our results and conclusions with other studies.

Research question 2 – Relation between quality of care indicators in supervision programmes

In two studies we investigated the relation between structures and processes of care and health outcomes: in integrated diabetes care by care groups and in quit-smoking counselling by primary care midwives.

Diabetes study

In the study on diabetes, we found no relation between structures and processes (i.e. guideline adherence) on the one hand, and health outcomes on the other hand. Despite our relatively weak design, our conclusions are in line with previous studies, which found that structure and process indicators of diabetes care showed mostly no effect on surrogate and hard outcomes.[11-13] Despite the fact that our study included more structure, process and outcome indicators than previous studies, the conclusion of our study is consistent with those of the previous studies; we conclude that the relation between structure and process indicators and health outcomes is unlikely to exist in this case.

Several mechanisms may explain the absence of a relation between structure and process indicators and health outcomes. One explanation might be that most of the processes in the guideline we studied were not evidence-based. This is the case for many processes and guidelines in diabetes as well as in other fields of healthcare.[14] Furthermore, the guidelines only rec-

ommend the measurement of risk factor levels, but the subsequent steps for risk factor control are not specified. However, the lack of correlation between structures and processes of care and health outcomes is also common for other diseases and more evidence-based processes. Another problem is that the patient population consists of heterogeneous diabetes patients and therefore it is difficult to formulate general guidelines that improve health outcomes in all patients. Finally, another possible explanation is dilution.[15] Guideline adherence varied in our study typically between 50 and 80%. The difference of 30% is the maximum percentage of patients who might have better health outcomes because of the better processes of care, while the effect is evaluated in the total population. Consequently, the effect is diluted and may go undetected in noisy observational data.

Our study further showed that guideline adherence was not optimal. Some aspects of the guideline were only followed for 50% of the patients or practices. This finding is in line with previous studies, which showed that processes of care were not strictly followed for diabetes patients.[16 17] The participating care groups indicated that ICT did not function properly and that care providers did not agree with the guideline, which might explanation this suboptimal adherence.

Smoking study

In the study on the effectiveness of quit-smoking counselling on the smoking behaviour of clients of midwifery practices, we found a trend towards more of the women quitting smoking after more extensive counselling. Propensity scores were used to adjust for possible confounders. We included more women in the study than was needed according to the power analysis, but still the analysis of the effect of 'perfect' counselling was underpowered, as the counselling was completed so infrequently. Most RCTs on quit-smoking counselling during pregnancy have shown a positive effect on quitting smoking, low birth weight and preterm birth.[18] Therefore, we conclude that quit-smoking counselling in pregnant smokers most likely positively affects the number of women who quit smoking.

We further found that quit-smoking counselling was initiated for 42% of the pregnant smokers (chapter 6), which was regarded as low. We did not identify specific practice or client characteristics that predicted the start of counselling. From our qualitative study we learned that midwives often regard smoking as a less important problem, compared to problematic living situations, financial problems and violence in which some pregnant smokers were situated.

Conclusions

Research question 1 addressed the effectiveness of two supervision programmes on quality of care. Based on our findings and the methodological considerations described above, we conclude that the theme-based supervision programme on quit-smoking counselling probably

was effective at improving quality of care. We found no evidence that the theme-based supervision programme on integrated diabetes care also improved quality of care, but given the methodological shortcomings we also cannot exclude the possibility that the supervision programme had an effect on structures and processes of care.

Research question 2 addressed the relation between structures and processes of care and health outcomes. We found conflicting results in our two cases. Based on our findings, the methodological considerations and the existing literature, we conclude that the relation between structures and processes of care and health outcomes, which (by definition) can be assumed to exist in the case of evidence-based structures and processes, is difficult to demonstrate using observational data.

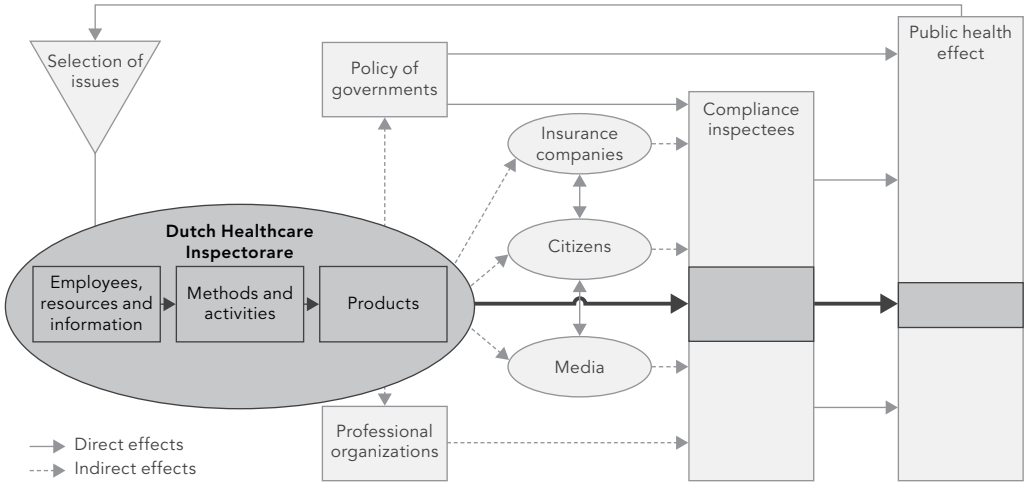
IMPLICATIONS AND RECOMMENDATIONS

Figure 1 shows the effect chain used by the Dutch Healthcare Inspectorate to show the expected effects and the context of its activities. The effect of supervision on public health can only be indirect, through inspectees.[19] Despite our ambition to quantify the effect of supervision on health outcomes, we only succeeded in providing evidence for the first arrow in the figure, the effect of the inspectorate on the compliance of inspectees, in terms of structures and processes of care. These somewhat disappointing results have important implications for the evaluation of supervision.

Research on supervision

Previous quantitative studies on supervision did not aim to evaluate its effects on quality of care.[20-23] This thesis presents one of the first

Figure 1 *Effect chain of Dutch Healthcare Inspectorate*



studies to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes, and we will therefore also reflect on the possibilities to evaluate supervision. First we discuss the possibilities and impossibilities of different study designs, then we will reflect on our choices in the selection of topics and finally we will present some considerations for further research on the effect of supervision.

Research designs in evaluating government supervision

In the studies presented in this thesis different study designs were used. To evaluate the first programme we used an RCT (Table 1). For the evaluation of the second programme we combined different research designs: an RCT, interrupted time-series design, before-after study and an interview-based qualitative case study. Therefore, it's possible to discuss the different options for research designs and their consequences.

Experimental and quasi-experimental designs

RCTs are the gold standard for establishing causal effects in clinical research.[24] However, it is unclear whether that is also the case for research on the effects of government supervision.[25] The term 'randomised' implies that care providers are at random allocated to the intervention condition. In both cases we studied, we learned that random allocation was possible. In the diabetes case, the inspectorate had no pre-existing information on quality of care groups and therefore it was decided to randomly allocate the inspections to care groups. In the smoking case, the inspectorate had pre-existing information about quality of care in midwifery practices. Therefore stratified randomisation was preferred over normal randomisation. In theme-based supervision, it will often be possible to randomly allocate the intervention. However, in other types of supervision randomisation is unethical or legally impossible. For example, supervision in response to calamities or emergencies is necessary for that particular care organisation and not for others.

The term 'controlled' implies that a group is allocated to 'no intervention'. In both cases we learned that this is difficult to achieve with care providers. In general, care providers

Table 1 *Study designs in this thesis*

Study	Design	Chapter
Effects of supervision on quality of integrated diabetes care	- Randomised controlled trial	2
Effect of supervision on midwives' quit-smoking counselling	- Combination of: RCT, interrupted time series design and before-after study	3
	- Interview-based case study	4

in the Netherlands are strongly organised at a local and national level, thus a spill-over effect may occur when care providers distribute their knowledge. The inspected care providers often share their experiences with the inspectorate among their colleagues from other practices. In such a situation, creating a strong contrast between inspected care providers and those not inspected is difficult and the controlled design fails. It is difficult to measure the spill-over effect because this would require all existing collaborations to be described in detail. Some collaborations however are only informal and therefore difficult to measure. It might be that under certain circumstances it is possible to create groups which are not connected to each other in any way, for example in different regions. However, we believe that it is unlikely researchers could do this for care providers in a small and condensed country such as the Netherlands. Furthermore, keeping the control group in the dark regarding the inspectorate's activities is not in line with the aim of the inspectorate, because it intends to reach as many care providers as possible and develop national programmes.

In summary, an RCT is a perfect design to generate empirical evidence of the effects of government supervision on quality of care, as long as sufficient contrast can be created between the intervention and control group. However, creating such contrast is rarely possible. Even when random allocation of the programme is possible, creating sufficient contrast between the intervention and control group is very challenging. Creating sufficient contrast between these groups is also difficult with quasi-experimental designs with a control group.

An alternative is a quasi-experimental study without a control group, for example a before-after study or interrupted time-series design.[26] These designs are more feasible to conduct and complete in studying supervision. However, the lack of a control group limits causal inference. This is a major limitation of these study designs. To increase the level of evidence, researchers should collect as much information as possible about the external factors that influence the time trend, increase the number of study measurements and optimise the timing of study measurements.

In conclusion, quasi-experimental designs without a control group are often more suitable for evaluating supervision than RCTs. However, in quasi-experimental designs there is more risk of bias and they should be performed and interpreted carefully.

The supervision programmes in this thesis have been designed by the Dutch Healthcare Inspectorate. The researchers had no influence on the process of designing and performing the supervision programmes. When researchers were present during meetings and inspections, they were not allowed to interact with the process and were only observers. The few changes to the process requested by the researchers were accepted, such as a request to keep information and letters confidential a few more weeks, before distribut-

ing them among care providers. This position as a silent observer was the price the researchers had to pay in order to be able to study supervision in a real-life setting.

In (quasi) experimental studies researchers need to decide which type of study measures will be used. In the case that previous studies support the causal effect of structures and process of care on health outcomes, it may be assumed that improved structures and processes of care will lead to better health outcomes. Although health outcomes might be considered most relevant, our research and many previous studies showed that it is often difficult to attribute differences in health outcomes between healthcare organisation to differences in structures and processes of care, due to numerous confounding factors.[4 27-29] Therefore health outcomes might not reflect structures and processes of care and should be used with caution in the evaluation of supervision.

Qualitative designs

The effects of supervision on quality of care can be well studied in qualitative designs. Especially in complex supervision programmes, qualitative designs are complementary to quantitative methods, in the so-called mixed methods approach.[30-32] We showed that such a mixed methods approach provides additional information on the interpretation of quantitative results, such as on what element of the supervision programme was effective and how it was effective. The limitations of qualitative designs in the study of supervision are the limited representativeness for the whole population and the impossibility of quantifying effects.

Selection of cases

The cases we used to address our research questions were selected based on methodological criteria[33] and feasibility for use in this study. Both cases that were selected scored high on almost all criteria for selection of cases for research (Table 2). Nevertheless, we experienced difficulties ruling out bias when interpreting improvements found over time. Next, in the case of supervision on integrated diabetes care, the effect mechanism was not straightforward because the supervision programme was directed at an organisational level instead of directly at patient care. Therefore, it is less likely that the supervision could have led to improvements in quality of care.

Selection of suitable cases was considered an important part of this research. We tried to find the best cases possible that would fulfil the methodological criteria. Four other supervision programmes were extensively analysed, but did not fulfil the requirements. Therefore, the results of this research are a realistic estimation of the possibilities of studying the effects of supervision and it is unlikely that better cases could have been selected. Other researchers should be prepared that selection of topics is not easy. The only way to select more suitable cases is to adapt a supervision programme in cooperation with scientific researchers.

The selected cases were both examples of theme-based supervision. This was not decided beforehand, but was the result of the selection procedure. It's likely that this is not a coincidence, because the other types of supervision do not fit to the selection criteria as presented in Table 2. For example, for supervision in response to calamities and risk-based supervision, a (comparable) control group does not exist under regular conditions. Moreover, such supervision programmes will usually be less specifically targeted to one topic, which makes the effect mechanism more difficult to specify and the selection of effect measures more complicated. Therefore these types of supervision are even more difficult to evaluate in terms of their effect on quality of care.

Considerations for future research

In this research, we encountered several characteristics of supervision that make a quantification of its effect on quality of care challenging. The Health Council published criteria that should be considered when studying the effect of supervision.[33] Based on our research, we propose to extend this list with several additional criteria (**bold**):

- Description of problem, aim, intervention and effects: Can the problem, aim, content of the supervision programme and the intended effects be clearly described?
- Effect mechanism: Can the effect of the supervision programme be described, is this effect plausible and **is the programme not too complex**?
- **Control group: Can a control group be created that is sufficiently separated from the intervention group and assigned to receive no or less supervision?**
- Randomisation of intervention and control group: Can care providers be allocated to the control and intervention groups randomly?
- **Statistical power: Can sufficient care providers be included in the study to perform the analysis with sufficient statistical power?**
- Effect measures: Can the effect measures reliably be measured, are the study measurements from an independent data source and are they measured by an independent researcher? **Are structures and processes of care measured? Are the ef-**

Table 2 *Criteria for selection of cases for research for evaluation of supervision*

Criteria	Diabetes	Smoking
Description of problem, aim, intervention and effects	+	+
Effect mechanism	+/-	+
Randomisation of intervention and control group	+	+
Effect measures	+	+
Data before/after comparison or trend data	+	+
Bias	+/-	+/-

fect measures closely related to the aim of the supervision programme? Are the effect measures studied at an appropriate moment in time (not too early to measure an effect, but also not too late to avoid dilution of the effect)?

- Data before/after comparison or trend data: Is information available on the effect measures to compare before-after the intervention?
- Bias: Is reliable information available on confounding variables that may explain the effect of a supervision programme?

Although we presented many challenges regarding the quantification of the effect of supervision above, some possibilities for evaluating supervision remain. Creating a control group of care providers without supervision is undesirable if they are not completely separated from the supervised care providers. If the control group is not isolated from the inspectorate and other inspected care providers, the study will underestimate the effect of supervision. When it is possible to create a completely separated control group, to provide an optimally valid result, the care providers should be randomly assigned to the groups. In addition, the size of the groups should be based on a power calculation. Although such an RCT is optimal for providing evidence of causality, this situation will rarely occur in practice.

This thesis is one of the first attempts to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes. Although we did not completely succeed in identifying successful or unsuccessful supervision activities, the effectiveness of supervision programmes remains a relevant question. To maximise the chance of identifying effective supervision activities, research should be designed to incorporate the complexity of the supervision programmes, but also provide strong evidence. This evidence of causality is especially important for the world outside the inspectorate. In addition, inspectorates might also be interested in which specific elements of supervision programmes work and how they work.[34] Scientists may evaluate supervision, but a close collaboration with inspectors is necessary to reach the best possible fit between the supervision programme and its evaluation.

Policy implications

This thesis shows that theme-based supervision has the potential to improve quality of care. With theme-based supervision, the inspectorate is capable of addressing problems in quality of care, problems that otherwise would be unaddressed. In this thesis we have not measured the cost of supervision or the burden to care providers. Therefore, the cost-effectiveness of theme-based supervision remains unknown.

There are opportunities for the inspectorate to increase the effectiveness of theme-based supervision programmes. A common approach in supervision programmes

is looking at risks in quality of care. However, reducing risks may not always be an effective way to improve the overall quality of care, for example if the prevalence of the risk is low. Instead, it may be more effective to look at problems that have a large potential for improving quality. Therefore we recommend including an analysis of the potential improvements in quality of care in every theme-based supervision project plan of the inspectorate.

The cases of quit-smoking counselling and diabetes care differed on one essential point. The guideline we studied in quit-smoking counselling is evidence-based and the guideline for diabetes care is not evidence-based. Lack of evidence-based guidelines is an essential problem for a supervision programme. The inspectorate has limited opportunities for sanctioning a care provider that ignores a non-evidence-based guideline. Furthermore, the inspectorate's ultimate aim to improve health outcomes will not be reached by enforcing guidelines that may have no effect on health. Therefore, the inspectorate should not enforce non-evidence-based guidelines.

In contrast, evidence-based guidelines can be enforced by the inspectorate by looking at compliance to these guidelines, in terms of structures and processes of care. As described in the recommendations for research, health outcomes might not reflect structures and processes of care and therefore should only be used with caution in supervision.

The inspectorate also has opportunities to increase the possibilities to evaluate supervision. As supervision in its current form is too challenging to evaluate, supervision programmes need to be adapted to make evaluation possible. These adaptations should focus on formulating concrete aims, conducting a proper problem analysis, executing well-planned programmes and facilitating data collection. Concrete aims and a proper problem analysis are important for constructing the effect mechanism, but also for executing a successful programme. Well-planned programmes are necessary for research, because only then can a high-quality research study be designed which fits the supervision programme. To facilitate data collection, supervision programmes should be planned over a longer time period. This allows care providers to change their behaviour after each supervision activity and gives researchers the opportunity to measure this change. These adaptations of the supervision programme are possible because they will have a limited impact on the supervision process. We propose minor changes that are simple to apply in practice. These differences will not only improve the researchability but also the supervision itself.

Inspectorates need to decide on the methodology and number of supervision programmes that should be studied. Not every supervision programme can be studied nor necessarily should be studied.[35] The Health Council of the Netherlands reported criteria for the relevance of the evaluation of supervision.[33]

These criteria are valid and useful. However, one of the criteria is that larger supervision programmes are more relevant to study than smaller supervision programmes. Extensive supervision programmes are more relevant to study, but larger programmes also provide more challenges with respect to researchability. Therefore, researchers should strike a balance between researchability and relevance of the supervision programme.

Besides the relevance criteria, the methodological criteria should also be considered. The effectiveness of supervision should only be studied if these criteria are satisfied, to increase the likelihood of success. Otherwise evaluation of supervision is a costly occupation, which will not result in evidence-based supervision.

This thesis focuses on the Dutch Healthcare Inspectorate. However, other Dutch inspectorates, as well as inspectorates from other countries, might also benefit and learn from the recommendations in this thesis. Our recommendations and tightened criteria may be used in the design of studies on the effectiveness of supervision, especially in healthcare, to contribute to evidence-based supervision.

To conclude, in this thesis we present one of the first attempts to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes. Although we encountered many challenges, we expect that this thesis may form the basis for further improvement of supervision and for generation of more empirical evidence of the effects of government supervision on quality of care.

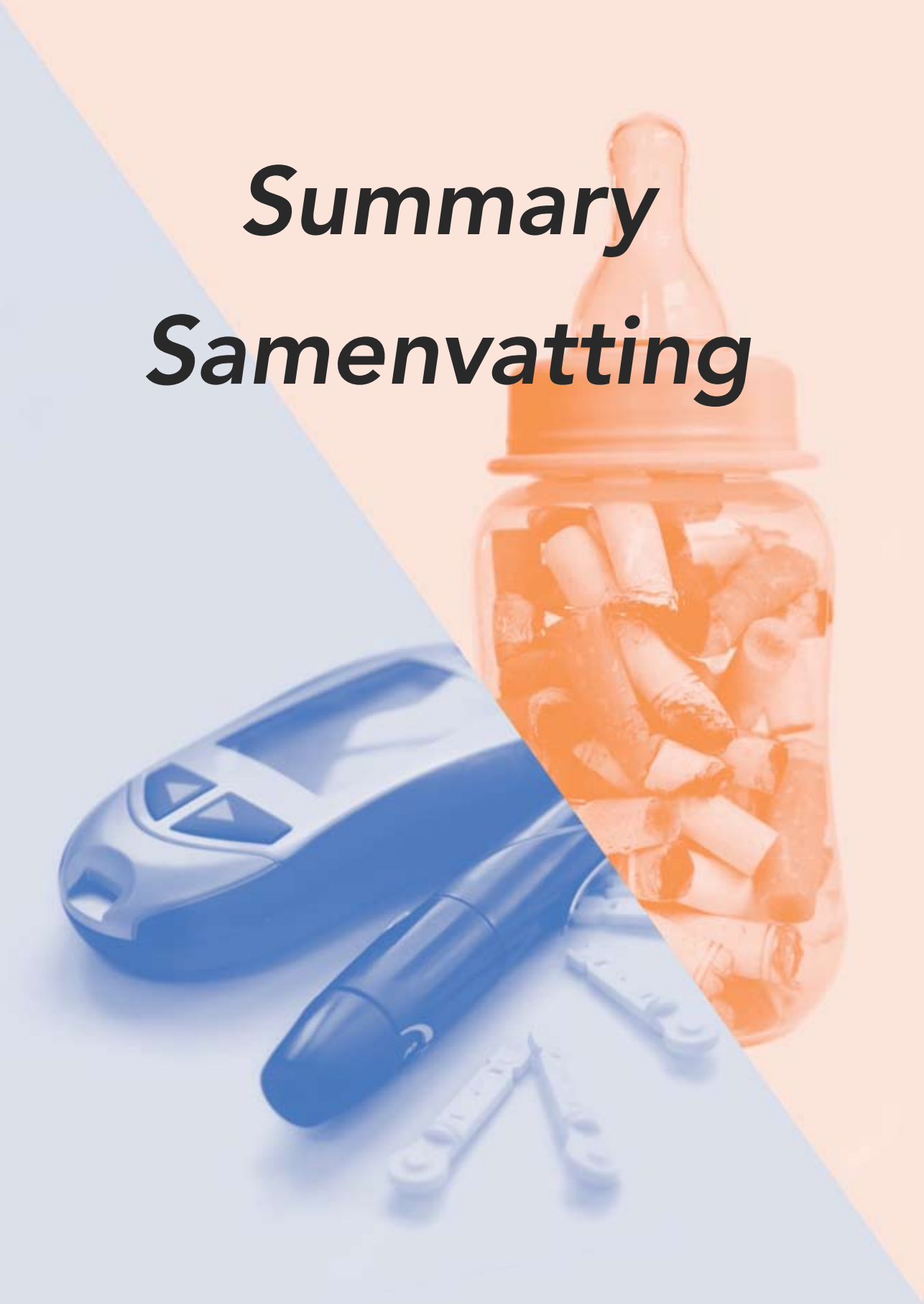
REFERENCES

1. Mortelmans D. H*andbook qualitative research methods* (H*andboek kwalitatieve onderzoekmethoden*). Leuven: Acco, 2009.
2. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337.
3. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000;321(7262):694.
4. Mant J. Process versus outcome indicators in the assessment of quality of health care. I*nt J Qual Health Care* 2001;13(6):475-80.
5. Lilford RJ, Chilton PJ, Hemming K, et al. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. BMJ 2010;341.
6. Walshe K, Higgins J. The use and impact of inquiries in the NHS. BMJ: B*ritish Medical Journal* 2002;325(7369):895.
7. Bevan G, Hood C. What's measured is what matters: targets and gaming in the English public health care system. P*ublic administration* 2006;84(3):517-38.
8. Bevan G, Hamblin R. Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK. J*ournal of the Royal Statistical Society: Series* A (S*tatistics in Society*) 2009;172(1):161-90.
9. Bevan G, Hood C. Health Policy: Have targets improved performance in the English NHS? BMJ: B*ritish Medical Journal* 2006;332(7538):419.
10. Walshe K, Wallace L, Freeman T, et al. The external review of quality improvement in health care organizations: a qualitative study. I*nternational Journal for Quality in Health Care* 2001;13(5):367-74.

11. Ackermann RT, Thompson TJ, Selby JV, et al. Is the number of documented diabetes process-of-care indicators associated with cardiometabolic risk factor levels, patient satisfaction, or self-rated quality of diabetes care? The Translating Research into Action for Diabetes (TRIAD) study. *Diabetes Care* 2006;29(9):2108-13.
12. Harman JS, Scholle SH, Ng JH, et al. Association of Health Plans' Healthcare Effectiveness Data and Information Set (HEDIS) performance with outcomes of enrollees with diabetes. *Med Care* 2010;48(3):217-23.
13. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, et al. Review: relation between quality-of-care indicators for diabetes and patient outcomes: a systematic literature review. *Med Care Res Rev* 2011;68(3):263-89.
14. Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009;301(8):831-41.
15. Lingsma HF, Steyerberg EW, Scholte op Reimer WJM, et al. Statin treatment after a recent TIA or stroke: is effectiveness shown in randomized clinical trials also observed in everyday clinical practice? *Acta Neurol Scand* 2010;122(1):15-20.
16. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, et al. A longitudinal study examining adherence to guidelines in diabetes care according to different definitions of adequacy and timeliness. *PloS One* 2011;6(9): e24278.
17. Stone MA, Charpentier G, Doggen K, et al. Quality of Care of People With Type 2 Diabetes in Eight European Countries Findings from the Guideline Adherence to Enhance Care (GUIDANCE) study. *Diabetes Care* 2013.
18. Lumley J, Chamberlain C, Dowswell T, et al. Interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst Rev* 2009;3(3).
19. OECD. Regulatory enforcement and inspections, OECD best practice principles for regulatory policy, 2014.
20. Klerks MCJL, Ketelaars CAJ, Robben P. Unannounced, compared with announced inspections: A systematic review and exploratory study in nursing homes. *Health Policy* 2013;111(3):311-19.
21. Tuijn S, Janssens F, Robben P, et al. Reducing interrater variability and improving health care: a meta-analytical review. *J Eval Clin Pract* 2012;18(4):887-95.
22. Huisman A, Robben P, Kerkhof A. An examination of the Dutch Health Care Inspectorate's supervision system for suicides of mental health care users. *Psychiatr Serv* 2009;60(1):80-85.
23. De Wilde RF, Besselink MGH, van der Tweel I, et al. Impact of nationwide centralization of pancreaticoduodenectomy on hospital mortality. *British Journal of Surgery* 2012;99(3):404-10.
24. Vandenbroucke JP, Hofman A, van Stiphout W. *Grondslagen der epidemiologie*: Elsevier/Bunge, 1999.
25. Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid). Supervising public interests, towards a broader perspective on government supervision (Toezien op publieke belangen, naar een verruimd perspectief op rijkstoezicht). Den Haag, 2013.
26. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*: Wadsworth Cengage learning, 2002.
27. van Gestel YRBM, Lemmens VEPP, Lingsma HF, et al. The hospital standardized mortality ratio fallacy: a narrative review. *Med Care* 2012;50(8):662-67.
28. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. BMJ 2010;340.
29. Lingsma HF. Measuring quality of care - methods and applications to acute neurological diseases. Erasmus University Rotterdam, 2010.
30. Bergman MM. *Advances in mixed methods research: Theories and applications*: Sage, 2008.
31. Creswell JW, Clark VLP. *Designing and conducting mixed methods research*, 2007.
32. Johnson RB, Onwuegbuzie AJ. Mixed methods research: A research paradigm whose time has come. *Educ Res* 2004;33(7):14-26.
33. Health Council of the Netherlands (Gezondheidsraad). Towards evidence-based supervision. Research on effects of supervision by the Dutch Health Care Inspectorate (Op weg naar evidence based toezicht. Het onderzoek naar effecten van toezicht door de Inspectie voor de Gezondheidszorg). Den Haag, 2011.
34. Berriet-Solliec M, Labarthe P, Laurent C. Goals of evaluation and types of evidence. *Evaluation* 2014;20(2):195-213.
35. Verdonk W. Supervisory bodies and impact assessment. *Impact assessment of the tobacco legislation - Effects of the workplace-smoking ban and the tobacco sales ban to minors*. Maastricht, 2014.

Summary

Samenvatting



SUMMARY

In healthcare, low quality of care can have serious consequences for patients and their relatives, but individual patients are generally unable to check the quality and safety of care. That is one of the reasons why the government regulates public and private parties in the field of healthcare. In the Netherlands, an independent governmental control agency, the Dutch Healthcare Inspectorate, is responsible for this supervision on quality and safety of care. For various reasons, activities of the inspectorate receive public and political attention on a regular basis. These reasons are for example that healthcare organisations desire less regulatory burden, politicians aim to reduce the budgets of the inspectorate and the public asks for maximum safety and highest quality of care. While continuously trying to deal with these contradictory of requests for both reduction and increase of supervision, the inspectorate aims to improve quality and safety of care.

Quality of care can be assessed by examining structures, processes and outcomes of care. Structures are the attributes of the setting in which care occurs, for example facilities and qualification of personnel. The term 'processes of care' refers to what is done when care providers are giving care, for example the practitioner's activities in making a diagnosis and recommending or implementing treatment. Outcomes of care include health status of patients and populations, for example mortality or functional outcome, but also patient's satisfaction with care.

Many strategies are described to improve quality of care; inspection is one of them. Inspections are a form of regulation that focusses on quality of care. In the Netherlands, government supervision in healthcare is conducted by the Dutch Healthcare Inspectorate (later: inspectorate), an independent agency of the Ministry of Health, Welfare, and Sport. The ultimate aim of the inspectorate is to improve population health by improving quality of care. The inspectorate supervises by using a combination of three methods: supervision in response to calamities, risk-based supervision, and theme-based supervision. In this thesis we will focus on theme-based supervision.

Little is known about the actual effects of the supervision programmes of healthcare inspectorates on quality of care and specifically health outcomes. Therefore the Health Council of the Netherlands and OECD recommended in their report to aim for evidence-based supervision. Information on the effects of supervision may help inspectorates in decision making and further improve their working methods. Research may also provide society with information on how successful supervision programmes have been.

The general aim of this thesis is to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes. We evaluated two cases of government supervision: supervision on integrated diabetes

care and supervision on quit-smoking counselling by midwives. We addressed two specific research questions:

1. What are the effects of these two government supervision programmes on structures and processes of care as well as on health outcomes?
2. In these cases of government supervision, what is the relation between structures and processes of care, and health outcomes?

Effect of government supervision on quality of care

In the first part of this thesis we focussed on the effect of government supervision on quality of care and health outcomes. In a cluster RCT we evaluated the effect of a supervision programme on the quality of integrated diabetes care (chapter 2). The supervision programme included announcements of inspections, site visits and sending individualised reports with specific recommendations to 20 randomly selected care groups. Indicators of effectiveness were derived from the structures, processes and outcomes of care. Structures and processes of care did not improve more in the intervention groups than in the control care groups. Moreover, health outcomes did not improve more in the intervention groups than in the control care groups. Although structures of care improved over time in the total population of intervention and control care groups, there were no changes in process of care or health outcomes. Therefore, we could not demonstrate improvements in quality of integrated diabetes care resulting from the supervision program. The result that structures of care did improve over time could not be attributed to the supervision programme.

In the following two chapters we evaluated a supervision programme on the provision of quit-smoking counselling by midwifery practices. The supervision programme consisted of 4 elements: 113 randomly selected practices were assessed using a questionnaire and sending individualised reports with specific recommendations; 10 other randomly selected practices were assessed through a site visit and sending individualised reports with specific recommendations; a deadline was announced by which all practices should comply with professional norms on such counselling; another 20 randomly selected practices were enforced through a site visit and sending individualised reports with specific recommendations.

In a quantitative study we evaluated the effects of the supervision programme on the provision of quit-smoking counselling by midwifery practices (chapter 3). We assessed provision of quit-smoking counselling through a minimal-intervention strategy. In practices that were assessed with a questionnaire, the provision of counselling improved partially compared to controls. After announcement of the deadline, Dutch midwifery practices reported significantly more provision of counselling. In conclusion, the provision of quit-smoking counselling improved spectacularly in Dutch midwifery practices. Despite some

limitations of our study, the supervision programme is likely to have contributed to the improvements in provision of counselling.
In an explorative qualitative study we identified factors related to guideline adherence after the supervision programme, and investigated whether the programme had helped improve adherence (chapter 4). We conducted semi-structured interviews with inspected and non-inspected midwives. Our results indicated that guideline adherence depends on several factors. Awareness and familiarity with the guideline are important, as is outcome expectancy. Additionally, motivation, guideline factors and environment factors were mentioned. Besides these previously documented factors, we found that professional collaboration also determined guideline adherence. Increased collaboration in counselling is associated with greater adherence to the guideline, such as provision of counselling and taking required training. The supervision programme helped improve quit-smoking counselling, by making midwives aware of the counselling and giving them an extrinsic motivation to provide counselling. In conclusion, of the factors related to guideline adherence, motivation and environmental aspects were the most important and professional environment was added as significant factor for guideline adherence. The improved adherence is partly attributable to the supervision programme.

Relation between quality of care indicators in supervision programmes

In the second part of this thesis we concentrated on the relation between structures and processes of care and health outcomes. In a cross-sectional study we aimed to quantify guideline adherence in general practices providing care to diabetes mellitus type 2 patients and explored the association between guideline adherence and patients' health outcomes (chapter 5). Guideline adherence was measured by comparing structure and process indicators of care with recommendations in the national diabetes care guideline. Health outcomes included biomedical measures and health behaviours. Guideline adherence varied between different recommendations. However, after adjusting for patient characteristics we found guideline adherence not to be associated with patients' health outcomes. We conclude that guideline adherence in Dutch general practices offering diabetes care was not optimal. Despite considerable variations between general practices, we found no clear relationship between guideline adherence and health outcomes.
In a quasi-experimental study we evaluated the provision of quit-smoking counselling by midwives and its effect on smoking behaviour and birth weight. The primary outcome parameter was 'quit smoking by end of pregnancy'. The secondary outcome parameter was birth weight. The midwives began counselling with 42% of the pregnant smokers, but seldom completed of

all the counselling steps. We found no difference in quit rate or birth weight between counselled women and those who were not. However, the data suggested that counselling is more effective when more steps of the minimal intervention strategy are completed. Therefore, provision of counselling can be improved, because it may lead to more quit smoking among pregnant women.

Discussion

In part 1, we studied the effectiveness of two supervision programmes on quality of care. Based on our findings and the methodological considerations described above, we conclude that the theme-based supervision programme on quit-smoking counselling probably was effective at improving quality of care. We found no evidence that the theme-based supervision programme on integrated diabetes care also improved quality of care, but given the methodological shortcomings we also cannot exclude the possibility that the supervision programme had an effect on structures and processes of care.

For part 2, we studied the relation between structures and processes of care and health outcomes. We found conflicting results in our two cases. Based on our findings, the methodological considerations and the existing literature, we conclude that the relation between structures and processes of care and health outcomes, which (by definition) can be assumed to exist in the case of evidence-based structures and processes, is difficult to demonstrate using observational data.

This thesis presents one of the first studies to generate empirical evidence of the effects of government supervision on quality of care and specifically health outcomes, and we will therefore also reflect on the possibilities to evaluate supervision. An RCT is a perfect design for this purpose, as long as sufficient contrast can be created between the intervention and control group. However, creating such contrast is rarely possible. Creating sufficient contrast between these groups is also difficult with quasi-experimental designs with a control group. In complex supervision programmes, a mixed methods approach provides additional information on the interpretation of quantitative results.

To improve future research on the effectiveness of supervision, we propose to extend the methodological criteria for this type of research. First, consider the complexity of the supervision programme under study. Second, consider the possibilities to create a control group that is sufficiently separated from the intervention group and assigned to receive no or less supervision. Third, consider if sufficient care providers are included in the study to perform the analysis with sufficient statistical power. Fourth, consider if structures and processes of care are measured; if the effect measures are closely related to the aim of the supervision programme and are measured at an appropriate moment in time.

This thesis also provides implications for policy. To increase its effectiveness, project plans should

also include analyses on the potential to improve quality of care. Next, the inspectorate should not enforce non-evidence-based guidelines, because it will limit the effectiveness of supervision. When evidence-based guidelines are enforced, health outcomes should only be used with caution, because they might not reflect structures and processes of care. To improve researchability, the inspectorate can adapt supervision programmes. These adaptations should focus on formulating concrete aims, conducting a proper problem analysis, executing well-planned programmes and facilitating data collection. Our recommendations and tightened criteria may be used by other Dutch inspectorates, as well as inspectorates from other countries, to contribute to evidence-based supervision.

SAMENVATTING

Als je een auto of hypotheek aanschaft, verwacht je dat deze van goede kwaliteit is. Het kan echter gebeuren dat na een tijdje het tegendeel waar blijkt te zijn, de auto gaat kapot of het ingelegde geld van de hypotheek verdampt. Was er opzet in het spel? Had je dit kunnen voorzien? In veel situaties kunnen burgers de kwaliteit van een product of dienst niet vooraf beoordelen. In het geval van slechte kwaliteit van de gezondheidszorg kan dit extra grote consequenties hebben. Daarom heeft de overheid een toezichthouder aangesteld die de burgers moet beschermen tegen slechte kwaliteit van de zorg. Dit proefschrift gaat over deze toezichthouder, de Inspectie voor de Gezondheidszorg: hoe deze optimaal en op basis van wetenschappelijke inzichten kan functioneren.

Kwaliteit is een belangrijke eigenschap van de gezondheidszorg en kan worden gemeten door middel van drie soorten kenmerken: structuur, proces en uitkomsten. Deze kenmerken worden ook wel indicatoren genoemd, omdat ze een indicatie geven van de kwaliteit van zorg. Structuurindicatoren zijn de kenmerken van de situatie waarin de zorg wordt verleend. Voorbeelden hiervan zijn de instrumenten die een huisarts beschikbaar heeft en het opleidingsniveau van de zorgverleners in de huisartsenpraktijk. Procesindicatoren beschrijven welke zorg wordt geleverd door zorgverleners. Bijvoorbeeld de huisarts die de diagnose stelt en daarna de behandeling uitvoert. Uitkomstindicatoren beschrijven de gezondheidstoestand van patiënten en populaties. Een voorbeeld is sterfte of lichamelijke conditie van een patiënt, maar ook de patiënttevredenheid met betrekking tot de zorgverlening.

De kwaliteit van zorg verbeteren kan op verschillende manieren. Eén manier is het aanstellen van een onafhankelijke en externe partij die toezicht houdt op de kwaliteit van zorg, zoals de Inspectie voor de Gezondheidszorg (IGZ). De IGZ is een overheidsinstantie die onafhankelijk toezicht houdt op ongeveer 40.000 instellingen en bedrijven, waar zo'n 1,3 miljoen personen werken. Het doel van de IGZ is het bevorderen van de volksgezondheid door effectieve handhaving van de kwaliteit van zorg. Daarvoor gebruikt de IGZ drie soorten van toezicht: thematoezicht, risico-indicatorentoezicht en incidententoezicht. Dit proefschrift gaat over thematoezicht.

In theorie kan de IGZ de kwaliteit van zorg verbeteren, maar is dat in de praktijk ook echt zo? Uit onderzoek bleek dat er bijna geen bewijs was gebaseerd op waarnemingen, zogenaamd empirisch bewijs. Daarom is het doel van dit onderzoek het verkrijgen van empirisch bewijs over de effecten van toezicht door de IGZ op kwaliteit van zorg. Dit is onderzocht aan de hand van twee programma's van de IGZ: toezicht op diabeteszorggroepen en toezicht op stoppen-met-rokenbegeleiding door verloskundigen.

Diabetes Mellitus type 2 is een chronische stofwisselingsziekte, waarbij de insuline die aanwezig is in het lichaam niet (voldoende) kan wor-

den gebruikt. Na verloop van tijd kan diabetes het hart, bloedvaten, ogen, nieren en zenuwen beschadigen. Dit type diabetes ontwikkelt zich langzaam en wordt vaak op latere leeftijd ontdekt.
Roken tijdens de zwangerschap is schadelijk voor moeder en kind. Eerstelijns verloskundigen kunnen rokende zwangere vrouwen helpen bij het stoppen met roken. Op dit moment rookt ongeveer 6% van de zwangere vrouwen, maar onder laagopgeleide vrouwen is dit ongeveer 14%.

Effect van toezicht op kwaliteit van zorg

In het eerste deel van dit proefschrift beschrijven we het onderzoek naar het effect van de twee programma's van de IGZ op kwaliteit van zorg. In het eerste programma hield de IGZ toezicht op de kwaliteit van verschillende schakels van zorgverleners (ook wel ketenzorg genoemd) voor diabetespatiënten in zorggroepen. In een experiment hebben we onderzocht wat het effect is van dit toezichtprogramma op kwaliteit van zorg (hoofdstuk 2 van dit proefschrift). In dit programma werd het toezicht aangekondigd, werd een deel van de zorggroepen bezocht door de inspectie en ontvingen zij daarna een individueel rapport met daarin specifieke aanbevelingen voor het verbeteren van de kwaliteit van zorg in hun zorggroep. In ons onderzoek hebben we geen effect van dit toezichtprogramma kunnen aantonen op kwaliteit van zorg. Dat betekent echter niet dat dit toezicht geen effect heeft gehad. We hebben aanwijzingen voor beperkingen in het onderzoek die kunnen verklaren waarom we geen effect konden aantonen. Een beperking zou kunnen zijn dat ons onderzoek niet lang genoeg heeft geduurd, totdat de zorggroepen hun zorg hadden verbeterd na het toezicht. Dus eventuele verbeteringen kunnen na ons onderzoek alsnog zijn opgetreden. Daarnaast was het verschil in toezicht tussen geïnspecteerde en niet-geïnspecteerde zorggroepen relatief klein.
Het andere toezichtprogramma, gericht op stoppen-met-rokenbegeleiding door verloskundigen, bestond uit meerdere delen: het afnemen van een toezichtvragenlijst bij verloskundigenpraktijken en het sturen van een individueel rapport met aanbevelingen (deel 1), het bezoeken van verloskundigenpraktijken en het sturen van een individueel rapport met aanbevelingen (deel 2), het afspreken van een uiterste datum waarop alle verloskundigenpraktijken de richtlijnen moeten opvolgen (deel 3) en opnieuw het bezoeken van praktijken en het sturen van een individueel rapport met aanbevelingen (deel 4). Ons onderzoek laat zien dat dit toezichtprogramma waarschijnlijk wel een effect heeft gehad op kwaliteit van zorg. We hebben gebruik gemaakt van zowel een kwantitatief onderzoek (hoofdstuk 3) als een kwalitatief onderzoek (hoofdstuk 4). Dit betekent dat we gebruik hebben gemaakt van zowel cijfers als van gesprekken en documenten. In dit onderzoek vonden we bijvoorbeeld dat de verloskundigenpraktijken die een vragenlijst hebben ingevuld voor de IGZ en daarna individuele

aanbevelingen van de IGZ ontvingen, vaker stoppen-met-rokenbegeleiding boden dan de praktijken die geen vragenlijst hadden ingevuld en dus ook geen individuele aanbevelingen kregen. Bovendien vonden we in een periode van 2 jaar belangrijke verbeteringen in stoppen-met-rokenbegeleiding bij vele verloskundigenpraktijken. Deze periode viel samen met de periode waarin de IGZ dit programma uitvoerde. In de gesprekken hebben we de verloskundigen ook gevraagd waarom zij hun stoppen-met-rokenbegeleiding verbeterden. De verloskundigen gaven aan dat zij door de IGZ bewust zijn geworden van het probleem van roken tijdens de zwangerschap en de mogelijkheid van het geven van begeleiding. Ook het feit dat deze begeleiding van de IGZ moest worden gegeven was een reden om dit daadwerkelijk te gaan doen. In dit onderzoek was het echter niet mogelijk om exact aan te geven welk deel van het toezichtprogramma deze verbetering heeft veroorzaakt en hoe groot het aandeel van de IGZ was in de verbeteringen in de stoppen-met-rokenbegeleiding door verloskundigenpraktijken. Ook andere organisaties en de verloskundigen zelf wilden de begeleiding te verbeteren.

Relatie tussen kwaliteitsindicatoren in het toezicht

Naast het onderzoek naar de effecten van toezicht hebben we ook onderzoek gedaan naar het verband tussen structuur, proces en uitkomstindicatoren van kwaliteit van zorg. Hiervoor hebben we opnieuw gebruik gemaakt van de ketenzorg voor diabetespatiënten (hoofdstuk 5) en stoppen-met-rokenbegeleiding door verloskundigen (hoofdstuk 6). Het verband tussen kwaliteitsindicatoren is belangrijk als de kwaliteit van zorg gemeten, beoordeeld en vergeleken wordt. In ons onderzoek naar diabetes ketenzorg vonden we dat lang niet alle patiënten zorg ontvingen volgens de richtlijnen. Per aanbeveling varieerde het percentage patiënten wat de aanbevolen zorg ontving tussen de 28% en 97%. We vonden ook grote variatie tussen huisartspraktijken. Echter, we hebben geen relatie gevonden tussen structuur, proces en uitkomstindicatoren.

Ook aan rokende zwangere vrouwen werd niet altijd de aanbevolen zorg verleend. In 42% van de rokende zwangere vrouwen werd stoppen-met-rokenbegeleiding gestart en slechts 5% van de vrouwen kregen tot het einde deze begeleiding. We hebben geen verschil gevonden in stopgedrag en geboortegewicht tussen vrouwen die begeleiding kregen en zij die geen begeleiding kregen. De resultaten van ons onderzoek suggereren wel dat betere stoppen-met-rokenbegeleiding, waarin meer stappen van stoppen-met-rokenbegeleiding werden geboden, leidde tot meer stoppen met roken onder zwangere vrouwen.

Discussie

In dit proefschrift hebben we onderzocht wat de effectiviteit van twee toezichtprogramma's op kwaliteit van zorg was. We concluderen dat het toezichtprogramma op stoppen-met-rokenbegeleiding in de

verloskundigenpraktijk waarschijnlijk effectief was in het verbeteren van kwaliteit van zorg. We hebben geen bewijs dat het programma voor diabetes ketenzorg ook de kwaliteit van zorg heeft verbeterd.
Daarnaast hebben we onderzoek gedaan naar de relatie tussen structuur, proces en uitkomstindicatoren. In zowel het diabetes als het stoppen-met-rokenonderzoek bleek het moeilijk een verband aan te tonen tussen structuur, proces en uitkomstindicatoren.
Dit proefschrift presenteert één van de eerste onderzoeken dat empirisch bewijs verzamelt over de effecten van toezicht op de kwaliteit van zorg. Daarom kunnen we aanbevelingen doen voor vervolgonderzoek naar de effecten van toezicht. Het is mogelijk een door middel van toeval toegewezen experiment uit te voeren, waarbij een deel van de zorgverleners onder toezicht staan (interventiegroep) en een ander deel niet (controlegroep). Echter, er moet voldoende verschil in toezicht zitten tussen deze twee groepen. Het is gebleken dat dit behoorlijk lastig is. Bij ingewikkelde toezichtprogramma's kan door middel van een combineerde methodiek van onderzoek, met zowel getallen en ervaringen als resultaat, meer inzicht worden verkregen voor de interpretatie van de resultaten.
In dit onderzoek zijn we met een aantal problemen geconfronteerd in onze poging tot het aantonen van het effect van toezicht. De Gezondheidsraad heeft criteria opgesteld die overwogen moeten worden als het effect van toezicht wordt onderzocht. Gebaseerd op onze ervaringen in dit type onderzoek, willen we graag extra criteria aanbevelen (**dikgedrukt**):

- Beschrijving probleem, doel, interventie en effecten: zijn (kunnen) het probleem, doel, de inhoud van de toezichtinterventie en de beoogde effecten helder (worden) beschreven?
- Werkingsmechanisme: is er een beschrijving van de keten van toezichtactiviteit naar effect, of kan deze achteraf beschreven worden; is het beoogde werkingsmechanisme plausibel **en is de toezichtinterventie niet te complex**?
- **Controlegroep: kan er een controlegroep gecreëerd worden die voldoende geïsoleerd is van de interventiegroep en die geen of minder toezicht toegewezen heeft gekregen?**
- Vergelijking interventie en referentie/controlegroep randomisatie: in hoeverre kunnen de interventie- en controlegroep gerandomiseerd worden, is het aantal eenheden voor randomisatie groot genoeg, is het ethisch-politiek en financieel haalbaar om te randomiseren? In hoeverre zijn de groepen vergelijkbaar?
- **Statistische power (onderscheidend vermogen): zijn er genoeg zorgverleners toegewezen aan de interventie- en controlegroep om een analyse uit te voeren met voldoende statische power?**
- Effectmaten: in hoeverre zijn (kunnen) de effectmaten betrouwbaar worden gemeten, zijn de metingen afkomstig uit een onafhankelijke databron, zijn ze gemeten

door een onafhankelijke onderzoeker? **Zijn structuur- en procesmaten gemeten? Zijn de effectmaten nauw gerelateerd aan het doel van het toezichtprogramma? Zijn de effectmaten gemeten op een geschikt moment in de tijd (niet te vroeg om een effect te verwachten en niet te laat wanneer het effect te sterk verdund is)?**

- Voor-na vergelijking/trend data: zijn er gegevens over de effectmaten beschikbaar om een voor-na vergelijking uit te voeren, of kunnen deze gegevens verzameld worden, zijn er meerdere meetmomenten voor- en na de toezichtactiviteit beschikbaar?
- Bias (vertekening): is er betrouwbare informatie beschikbaar over 'verstorende' variabelen die een mogelijk effect van de toezichtactiviteit op de effectmaten zouden kunnen verklaren, of kan deze informatie verzameld worden?

Naast deze aanvulling op de criteria, kunnen we ook aanbevelingen doen aan de IGZ.

- Voor het verhogen van de effectiviteit zou de IGZ in elk thematoezicht moeten nagaan wat de mogelijkheden zijn voor verbeteren van de kwaliteit van zorg.
- Daarnaast zou de IGZ geen richtlijnen moeten handhaven die niet wetenschappelijk onderbouwd zijn, want dat zou de effectiviteit van toezicht beperken. Als wetenschappelijk onderbouwde richtlijnen worden gehandhaafd, zou in het gebruik van gezondheidsuitkomsten voorzichtigheid moeten worden geboden, omdat gezondheidsuitkomsten niet altijd een goede voorspeller zijn voor structuur en processen van zorg.
- De IGZ kan onderzoek eenvoudiger maken door toezichtprogramma's aan te passen. Deze aanpassingen richten zich op het formuleren van concrete doelstellingen, het uitvoeren van een gedegen probleemanalyse, het uitvoeren van zorgvuldig geplande programma's en het mogelijk maken van dataverzameling voor onderzoek.

Onze aanbevelingen en aangescherpte criteria kunnen, naast de IGZ, ook worden gebruikt door andere inspecties en inspecties in andere landen, waardoor het kan bijdragen aan wetenschappelijk onderbouwd toezicht.


INSPECTIE VOOR INSPECTIE VOOR DE GEZONDHEIDSZORG
ALS ER NU OOK NOG EEN INSPECTIE VOOR DE INSPECTIE VOOR DE INSPECTIE KOMT, IS DE CRISIS METEEN OPGELOST!
AARGH

Dankwoord

About the author

Publications

PhD portfolio

DANKWOORD

De afgelopen vijf jaar was dit proefschrift mijn stip op de horizon. Ik wil iedereen bedanken die – bewust of onbewust – heeft bijgedragen aan het feit dat ik deze stip nu bijna heb bereikt.

Hester, samen met jou heb ik dit proefschrift geschreven en ik ben je ontzettend dankbaar voor alles wat je voor me hebt gedaan. Jouw chaotische enthousiasme zorgde er voor dat het schrijven van dit proefschrift altijd leuk was. Ik bewoner je analytisch vermogen en motiverende scherpte. Ik ben trots dat we dit samen hebben kunnen volbrengen. Het is voor mij een eer om jouw eerste promovendus te mogen zijn. Ik wens je alle geluk in je verdere wetenschappelijke carrière!

Paul, dankzij jou is dit proefschrift er gekomen. Als opdrachtgever, maar vooral als wetenschapper heb je er voor gezorgd dat het onderzoek elke fase heeft overleefd. Je tactische werkwijze en geduldige, maar ook kritische blik hebben het project doen slagen. Ik ben je enorm dankbaar voor alles wat je voor me hebt gedaan om te komen waar ik nu sta.

Johan, dankzij jou heeft het proefschrift dit niveau gehaald. Als ik het idee had dat ik alles wist en alle vragen kon beantwoorden, wist jij altijd een vraag te stellen die me verder uitdaagde. Ik bewonder je snelle denkvermogen. Dank voor de goede tijd op MGZ.

Vicki, het was fijn om tijdens Hester's afwezigheid op iemand terug te kunnen vallen. Je communicatieve vaardigheden hebben de respons op de onderzoeken in dit proefschrift enorm verhoogd, dank daar voor!

Prof. dr. Wagner, prof. dr. Janssens en prof. dr. Bindels, hartelijk dank voor het beoordelen van het manuscript van dit proefschrift. Ik dank u, en de overige leden van de grote commissie, voor het deelnemen aan de oppositie.

Beste coauteurs, dank voor jullie waardevolle inbreng in alle fases van het onderzoek. Ik ben trots dat jullie namen in mijn proefschrift staan.

Beste onderzoeksmedewerkers, jullie hebben werk uit handen genomen wat essentieel was voor het slagen van het onderzoek, ik had het zelf niet beter kunnen doen. Bedankt!

Zeer gewaardeerde deelnemers aan het onderzoek, jullie deelname heeft de basis gelegd voor dit proefschrift, zeer veel dank!

Alle inspecteurs en adviseurs bij de IGZ bedankt voor jullie hulp bij het zoeken naar de mogelijkheden voor onderzoek. Ook al moesten we, soms in een laat stadium, besluiten dat sommige onderzoeken niet mogelijk waren. Dit is toch waardevol geweest voor het onderzoek. En uiteraard ook heel veel dank aan iedereen bij de IGZ die heeft meegewerkt aan het onderzoek in dit proefschrift!

Vele organisaties en zorgverleners hebben zijdelings meegewerkt aan het onderzoek, waarvan ik enkele specifiek wil noemen: **KNOV, STIVORO, TRIMBOS, RIVM, VIDE, academische werkplaats toezicht IGZ, inspectieraad, verloskundigen en praktijkondersteuners huisarts**. Dankzij jullie kreeg ik regelmatig 'een kijkje in de keuken' van de zorg- en toezichtspraktijk en waardevolle feedback van experts.

Mannen van VRUMUN, jullie waren fantastisch in het faciliteren van de data-extractie, veel succes met het verdere uitbouwen van VRUMUN en eventuele nieuwe kansen.

Zonder technologie had de totstandkoming van dit proefschrift tig keer zo lang geduurd. **Google, Dell, Microsoft, IBM, SAS en Samsung**, bedankt!

Van de eerste tot de laatste stappen op MGZ hebben velen mijn onderzoek ondersteund:
Lieve secretaresses en adviseurs, het was super om met kleine en grote zaken altijd bij jullie terecht te kunnen.
Beste mannen van de ICT en databeheerders, dankzij jullie stond de ICT altijd klaar voor het onderzoek.
Ook voor de **statistische ondersteuning** niets dan lof, geen analyse kon door jullie niet worden opgelost.
Het onderzoeksbureau maakte de dataverzameling een stuk makkelijker.
Allemaal bedankt! Het was heerlijk om binnen te lopen met een vraag en daarna meteen weer met een oplossing naar buiten te lopen.

Tijdens een groot deel van mijn onderzoek heb ik mee mogen werken aan de medezeggenschap in het Erasmus MC. Dit heeft er voor gezorgd dat ik zo nu en dan verder kon kijken dan dit onderzoek. Speciale vermelding verdient de **OC Gezondheidswetenschappen, de Gezamenlijke Vergadering, de Commissie Onderwijs & Onderzoek van de Ondernemingsraad en bureau medezeggenschap**.

Aan de **studenten** die mijn pad zijn gekruist tijdens dit onderzoek: dankzij jullie weet ik dat lesgeven niet verschrikkelijk moeilijk en eng is en door jullie frisse blik vond ik altijd weer energie om door te gaan. Veel succes in jullie verdere opleiding en carrière!

Lieve kamergenoten, het was een fantastische tijd samen, misschien heb ik van jullie nog wel het meeste geleerd de afgelopen jaren. Ik heb me door jullie altijd enorm gesteund gevoeld, dank daar voor!

Alle MGZ collega's, bedankt voor de vele wetenschappelijke en sociale gebeurtenissen. De CMB werkbesprekingen, research meetings, taartjes, praatjes in de keuken, lunchwandelingen in het 'kleine' en 'grote' park en de vele kopjes thee hebben voor de nodige afleiding gezorgd. Hester en 'haar meiden', bedankt voor de gezellige etentjes!

Allerliefste paranimfen, dank dat jullie op deze speciale dag naast mij willen staan.

Heel veel dank ook aan al mijn **juffen, meesters, leraren en docenten** die het (on)mogelijk hebben gemaakt om hier te komen. Speciale dank aan de begeleider van mijn master scriptie en zijn collega's, dankzij hen durfde ik deze uitdaging aan te gaan.

Lieve vrienden, zonder jullie is het leven maar saai! Ik ben blij dat jullie er zijn.

Lieve familie, jullie zijn onmisbaar!
Liebe Verwandtschaft, vielen herzlichen Dank für eure großzügige Gastfreundschaft!
Lieve broer, ik ben trots dat ik je zus mag zijn!
Lieve pap en mam, jullie zijn de beste!

Samen waren jullie de drijfveren achter dit proefschrift.

BEDANKT!

En nu is het tijd voor een nieuwe stip op de horizon

ABOUT THE AUTHOR

Sandra Frederieke Oude Wesselink was born in Zwolle, the Netherlands on February 18th 1987. In the age of 8, she moved to Hengelo. After graduation of her secondary school education in 2005, she started a bachelor in Biology and Medical Laboratory Research at Saxion university of Applied Science in Enschede. After completion of the first year she moved to Maastricht University. At this university she obtained a Bachelor of Science degree in Health Sciences in 2009 and a Master of Science degree in Public Health in 2010. The specialisation of the master was Health Services Innovation and the topic of her master thesis was fear of falling in older people. In the same year she started a PhD trajectory on effects of governmental supervision at the department of Public Health of the Erasmus Medical Centre of Rotterdam, which resulted in this thesis. As part of the PhD education she obtained a second Master of Science degree in Health Sciences at the Netherlands Institute of Health Sciences in 2013. At this moment Sandra is working as risk detection and supervision development advisor at the Dutch Health Care Inspectorate, Utrecht.

LIST OF PUBLICATIONS

Kempen GIJM, Oude Wesselink SF, van Haastregt JCM, Zijlstra GAR. Long-term effect on mortality of a multicomponent cognitive behavioural group intervention to reduce fear of falling in older adults: a randomised controlled trial. Age and Ageing 2011;40(4):519-23.

van Dishoeck AM, Oude Wesselink SF, Lingsma HF, Steyerberg E, Robben PB, Mackenbach JP. Transparency: can the effect of governmental surveillance be quantified? (Transparantie: is het effect van toezicht te meten?). Nederlands Tijdschrift voor Geneeskunde 2013;157(16).

Oude Wesselink SF, Lingsma HF, Reulings PGJ, Wentzel HR, Erasmus V, Robben PBM, Mackenbach JP. Does Government Supervision Improve Stop-Smoking Counseling in Midwifery Practices? Nicotine & Tobacco Research 2015;17(5):572-9.

Oude Wesselink SF, Lingsma HF, Robben PBM, Mackenbach JP. Guideline adherence and health outcomes in diabetes mellitus type 2 patients: a cross-sectional study. BMC Health Services Research 2015;15(1):22.

Oude Wesselink SF, Lingsma HF, Ketelaars CAJ, Mackenbach JP, Robben PBM. Effects of government supervision on quality of integrated diabetes care: A cluster randomized controlled trial. Medical Care 2015 (in press).

Oude Wesselink SF, Lingsma HF, Robben PBM, Mackenbach JP. Provision and effect of quit-smoking counselling by primary care midwives. Midwifery 2015 (in press).

Submitted

Oude Wesselink SF, Stoopendaal A, Erasmus V, Smits D, Mackenbach JP, Lingsma HF, Robben PBM. Government supervision on quality of smoking-cessation counselling in midwifery practices: a qualitative exploration

PHD PORTFOLIO

Name: Sandra Oude Wesselink
Erasmus MC, Department of Public Health
Research School: Netherlands Institute for Health Sciences
PhD period: 2010 - 2015
Promotors: Prof. dr. Johan Mackenbach
Prof. dr. Paul Robben
Copromotor: Dr. Hester Lingsma

1. PhD training	Year	Workload (ECTS)
General courses		
Biomedical English Writing and Communication – David Alexander	2014	1
Research Integrity – Dr.ir. Medard Hilhorst	2014	0.3
Education in small groups – Yvonne Gruteke	2014	0.2
Scientific writing course MGZ – Dr. Frank van Lenthe	2013	1
Presentation skills Eigenwijs Presenteren (Ellen Looyestijn)	2012	0.1
Specific courses		
Master Public Health, NIHES	2011-2013	70
Health Policy and Politics, institute of Health Policy & Management	2011	5
Seminars and workshops		
Seminars department of Public Health, Erasmus MC	2010-2014	2
Meetings clinical decision making, department of Public Health, Erasmus MC	2010-2014	1
Colloquium, Dutch Health Care Inspectorate	2010-2014	1
Lectures, Dutch Health Care Inspectorate	2010-2013	1
Presentations		
Colloquium, Dutch Health Care Inspectorate	2015	1
European Partnership for Supervisory Organisations in Health Services and Social Care (EPSO)	2015	1
Academic Network Supervision	2014	1
Dutch Conference Public Healh Care	2014	1
Society of Medical Decision Making (Europe)	2014	1
Colloquium, Dutch Health Care Inspectorate	2012	1
European Partnership for Supervisory Organisations in Health Services and Social Care (EPSO)	2012	1
Effects of supervision, the ideal formula	2012	1
International conferences		
Society of Medical Decision Making (Europe)	2014	1
International Society for Quality in Health Care	2014	1
European Health Management Association	2013	1
2. Teaching		
Supervising Bachelor thesis	2014	1
Lecturer medical student VO theme 3.C.2 (primary prevention)	2013-2014	1
Supervising Master thesis	2012-2013	2
Supervisor medical students theme 3.C.4 (community projects)	2012-2014	1
3. Other		
Member "Gezamenlijke vergadering"	2013-2014	2
Member "Onderdeelcommissie thema Gezondheidswetenschappen"	2012-2014	10