PERFORMANCE OF MATERNITY CARE FROM THE CLIENTS PERSPECTIVE

Development & application of the ReproQuestionaire

Marisja Scheerhagen

PERFORMANCE OF MATERNITY CARE FROM THE CLIENT'S PERSPECTIVE

Development and application of the ReproQuestionnaire

Marisja Scheerhagen

This research project was funded by Stichting Miletus

ISBN: 978-94-028-1523-8

Cover/design by FAIR2 Media
Lay-out by Legatron Electronic Publishing, The Netherlands
Printed by Ipskamp Printing, The Netherlands

PERFORMANCE OF MATERNITY CARE FROM THE CLIENT'S PERSPECTIVE
Development and application of the ReproQuestionnaire

Performance van de geboortezorg vanuit het perspectief van de cliënt
Ontwikkeling en toepassing van de ReproQuestionnaire

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

June 5th, 2019 at 11.30 hrs

by

Marisja Scheerhagen
born in Amsterdam

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotors: Prof.dr. G.J. Bonsel
Prof.dr. A. Franx

Other members: Prof.dr. J.A. Hazelzet
Prof.dr. D. Delnoij
Prof.dr. C.J.M. de Groot

Copromotors: Dr. E. Birnie
Dr. H.F. van Stel†

Paranymphs Nico Ruitenbeek
Martin Scheerhagen

† Shortly before completion of this thesis, co-promotor Henk van Stel unexpectedly died during his summer holiday.

Facts are facts, but perception is reality

Contents

1

General Introduction

POOR OUTCOMES OF DUTCH MATERNITY CARE

Perinatal mortality rate in the Netherlands is still high compared to other European countries: in 2010 the Netherlands had the sixth highest fetal and neonatal mortality rate of the 29 evaluated countries. The prevalence of the major forms of perinatal morbidity, the so-called BIG4 (congenital anomalies, preterm birth, small for gestational age and low Apgar score) is high in the Netherlands too[1-3]. About 85% of perinatal mortality is preceded by at least one of the BIG4[4]. Before 2010, the dominant theory was that population factors were responsible such as high average maternal age at first childbirth, the high prevalence of multiple pregnancies, as well as the non-treatment policy in very premature births. Since 2010, the non-medical factors are thought to be of additional relevance[4-8]. Non-medical factors that influence the medical outcome are socio-demographic characteristics (e.g. ethnicity and socio-economic status) but also lifestyle-related factors (e.g. alcohol consumption and smoking). Moreover, it is thought that clients' experiences could affect health outcome[9-12]. For example, clients who truly understand the explanation of their caregiver are more likely to comply to treatment or to change lifestyle. Finally, organizational and professional factors probably play a role. Recent registry data based studies show that organizational and professional factors explain about 30% of perinatal mortality differences. These factors partly relate to the Dutch two-tier system maternity care, which is characterized by risk selection in all stages of antenatal, natal and postnatal care, strict division of service provision between different health care professionals, and to insufficient 24/7 continuity of hospital-based care[4,13,14].

Based on these findings, governmental, professional, and institutional stakeholders initiated a series of reform measures to improve maternity care performance in 2010. These measures were the following.

Firstly, maternity care is being organized in perinatal units (in Dutch: verloskundig samewerkingverband). A perinatal unit consists of a hospital with associated community midwife practices and maternity care organizations. The aim of perinatal units is to achieve more effective collaboration between all involved professionals[4,8], including sharing professional responsibility for clients rather than a strict division of tasks between the first and second tier and health care professionals involved[8,15-18]; one clinical perspective, one risk management approach and one client orientation is assumed, i.e. integrated tier-independent care and shared care[19].

Secondly, collective and individual preconception care should be implemented[20-23].

Thirdly, antenatal risk selection should improve to avoid delay of suitable medical care and late referral to secondary care. Risk selection should not only be based on medical risk factors, but also screening or detection of non-medical risk factors such as socio-medical risk factors and indicators of socio-economic status and deprivation. Specific instruments to achieve this are R4U[24] and Mind-2-Care[25].

Finally, implementation of setting continuity and integrating medical facilities in primary care during birth through implementation of birth centers[26-29]. Birth centres, for example, are aiming for a smoother transition between different types of health care professionals, and avoiding abrupt transfers between settings (22% of all deliveries)[30].

This process is still ongoing. In our view the improved system performance should be evaluated, which is currently absent.

WHO MODEL OF SYSTEM PERFORMANCE

In 2000 the WHO presented a comprehensive model to compare different health systems (global comparison), monitor its performance, and evaluate system changes[31,32]. According to this model, clients' experiences give an indication of the system responsiveness to the clients' values and expectations and are a reflection of honouring human rights[9-12,33]. Therefore, the World Health Organization (WHO) in 2000 stated a system's 1) responsiveness as an independent indicator of its performance, along with the systems 2) health and 3) fairness of financial contributions (Figure 1). We will shortly discuss the outcome indicators below.

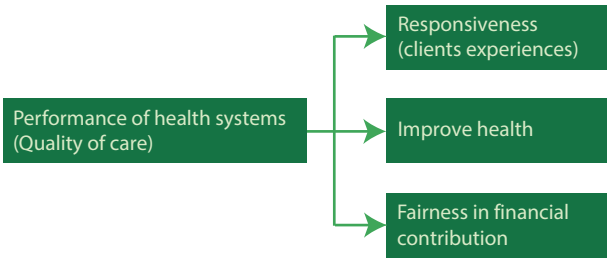
First, the WHO elaborated responsiveness as the way a client is treated by the professional and the environment in which the client is treated, where eight different domains are suggested to cover the concept. This model deliberately focuses on individual experiences rather than characteristics of processes or structures, acknowledging that between and even within countries the same client experiences may be arrived at by various means[31-34]. Next, health focus on the average levels of health and the distribution of health across individuals[31,32]. In the context of maternity care one may think of conventional measures of neonatal and maternal health, on short and long term, both expressed in averages and in gaps between groups (socio-economic status (SES)-based, ethnic-based). Finally, the last outcome of performance strives for fairness in regard to households having to bear the burden of payments to the health system[31,32]. In the maternity care context, one may think of the access to hospital care (limitations of choice by insurance policies in particular in so-called budget schemes), to prenatal tests, place of delivery and maternity care (limitations through co-payment); these policies in the Netherlands exert large, selective influence.

According to the WHO, the quality of a health care system is sufficient/good if the average levels of both health and responsiveness are high, there are no inequalities in health status and responsiveness, and there is a small distribution of across individuals in fairness in financial contribution[31,32]. This could be translated to the current situation in maternity care as follows: First, health (maternal, neonatal) and responsiveness levels during the entire process of care are at least average compared to comparable countries, and within the Netherlands the variation in health and responsiveness is limited according to unit (hospital, practice).

Second, gaps in health and responsiveness according to SES, ethnicity, religion, and place of living are limited, or absent if this proves to be an attainable goal. Third, access to mother and child essentially is free of charge. Finally, where obvious outcome deficits are present beyond chance, quality improvement procedures are into place.

The WHO model seems also appropriate to evaluate the system changes in the Dutch maternity care, outlined in section 1. As described, the performance of the Dutch maternity care is currently unknown, so the effectiveness of the implemented changes can not be evaluated. Although the system's health can be evaluated in terms of indicators of perinatal morbidity and perinatal mortality –which we know are suboptimal (see paragraph 1)–, the system's responsiveness and financial fairness have not been evaluated before. This thesis is especially dedicated to the responsiveness of the Dutch maternity care.

Figure 1. Framework for assessing the performance of healthy systems.



IMPROVING CLIENTS' EXPERIENCES

Evaluating the clients' experiences by a health care organization could initiate a two-stage quality cycle. The first stage, focuses on "determine our position", by ranking the scores of the health care organizations that collectively from a health care system. This allows health care organizations with outlying performance (or best and worst practices) to be identified. In the second stage, underperformers are invited to improve their results followed by an internal interpretation of the results and improvement of care accordingly. This process is also called benchmarking[35-38]. By routinely performing a benchmark the effectiveness of the improvement measures can be evaluated.

Prerequisite for successful benchmarking is the routine measurement of clients' experiences. The implementation of a uniform measurement procedure has to resolve several challenges due to the peculiarities of maternity care and the required suitability for quality cycles.

Regarding the peculiarities of maternity care, those challenges are:

First is the choice of the source of information. We may ask the mother for experiences, but for the baby some form of proxy measurement is mandatory, and we should decide to what extent the concept of client experiences translates in this situation.

A second challenge concerns the unit of measurement and analysis, which must be connected to the current transition in maternity care. The obvious choice would be to analyze the clients' experiences according to the perinatal unit where she received care: health care professionals have a shared responsibility for their clients' care and the large diversity in co-operation between the organizations is beneficial to nationally improve the clients' experiences. However, we expect that the co-operation between different health care professionals and organizations is often too fragile for a benchmark to reach its full potential.

Third challenge is the reference period to be evaluated. Maternity care covers different time windows (antepartum phase, childbirth, postpartum phase). From a managerial point of view one would consider separate measurements, to create feedback loops on the spot, but this is demanding. Moreover, both medical outcomes and patient experiences tend to influence the outcomes of the subsequent phase (also described in this thesis)[9,10,33,39,40].

In the context of application in quality cycles, the fourth challenge is to define and identify poor, average and good performance. This covers both the need for 'case mix' adjustment, and the definition of poor and good performers. With case mix correction the data is adjusted for determinants that 1) are beyond the influence and usually unrelated to the organization, but which 2) influence the outcome (here: clients' experiences) and 3) are distributed unequally across health care organizations[41]. 'Beyond the influence of an organization' is often wrongly understood. Of course, being of non-Western background or belonging to a low-SES group or living in a deprived area all are not subject to change ('beyond the influence'), but often the effects thereof can be succesfully mitigated. Next, defining poor, average and poor performance depends of the norm. Additional to the statistical approach, one could also categorize units based on a relevant difference (or minimally important difference (MID)) with the reference point[42]. The MID should take into account that a. the outcome variation among clients (after case mix adjustment) still is only to a limited degree subject to unit performance, and b. for a unit to be better or worse, it seems inefficient to require that *all* clients improve on average one MID.

The last and fifth challenge emerges if units are detected with consistently poor responsiveness. It appears difficult to relate particular poor (or good) outcomes to their origin. Hence processes, like detailing the data and discussing results with involved professionals, have to be put into place to create the translation from measured underperformance into action for improvement.

DEVELOPMENT OF THE INSTRUMENT

For equal measurements within and between health care organizations, the clients' experiences can best be collected by one or more surveys. To structurally evaluate the clients' experiences several instruments/questionnaires already exist, e.g. NHS en CQ instruments[43-46]. However, these questionnaires are unable to deal with the previously described challenges: they either focus on specific processes (NHS survey), monodisciplinary perspectives (CQ) or assume a specific maternity care organization, and lack a formal aggregate scoring system for the client's experience allowing a graded quality judgment[43-46]. Therefore, we developed and extensively tested a patient reported measure addressing the client's experience conform the WHO responsiveness model. The questionnaire should be suitable for the perinatal context, and comply with the theoretical considerations shown in Box 1. This questionnaire was coined the ReproQuestionnaire (ReproQ).

Box 1. Theoretical considerations

- WHO Responsiveness model as conceptual basis
- Symmetrical antepartum and postpartum version of the questionnaire, covering first antenatal visit up to postpartum maternity period
- Neutral toward provider or organization structures
- Perspective of the mother, but mother and child
- Performance-as-experienced then, and reported now, by the client
- Suitable for both stages of a two-stage quality cycle
- Suitable for clients with low educational level and clients with a non-Dutch background
- Digital base, but multimodal applicable
- Short in terms of time to complete

AIM OF THIS THESIS

The aim of this thesis is to give a *scientific* account of the development, testing and piloting of the ReproQ of which the development started in the end of 2011. The anticipated use was plural: for most monitoring quality of care, and effectiveness and inequality research.

During development three phases can be distinguished, each with its own research questions. Phase 1 focuses on the initial development and explores several essential psychometric analyses. Phase 2 assesses the ReproQ's suitability for a benchmark and determines its discriminative power. Phase 3 focuses the implementation and application of the ReproQ after development.

In this PhD thesis, the following research questions are answered:

Development

1) What is the content and construct validity?
2) What is the test-retest reliability of the postnatal ReproQ?
3) What is the Minimally Important Difference of the ReproQ?

Benchmarking

4) In a benchmark, which determinants should be considered for case mix adjustment, and which determinants attribute to the explanation of a low client experience score?
5) Is the ReproQ able to identify best practices and underperformers when used in a benchmarking?

Implementation and application

6) Can the antenatal experiences be measured validly after birth?
7) After development, is the ReproQ suited for quality improvement when taken into practice?

OUTLINE OF THIS THESIS

This thesis consists of three parts, following the three developmental phases; see Table 1.

Table 1. Outline of the thesis.

Development: construct, psychometrics, scoring	Benchmark, discriminatory power	Implementation and application in quality improvement
Chapter 2 Chapter 3	Chapter 4 Chapter 5	Chapter 6 Chapter 7 Chapter 8

In part 1 (**chapter 2 and 3**) the theoretical considerations are determined, after which a first concept was made in co-operation with professionals, health care professionals and health insurance companies. Next, several psychometrics of the instrument are tested, among the content and construct validity, its test-retest reliability and the minimally important difference (MID). All analyses focus on the quality of the questionnaire and are beneficiary the questionnaire's suitability for a benchmark. Some aspects of psychometrics are slightly different from conventional testing. For example a skewed score distribution of domains or the questionnaire as a whole not necessarily means a 'poor' of 'invalid' instrument. In this normative context it may be simply the case that particular aspects of care delivery on the one hand are regarded as essential, and on the other hand universally are carried out very well.

Moreover, we developed three different scoring models: the mean score, the median score (above/below the median) and the negative score (having at least one negative experience).

In part 2 (**chapters 4 and 5**) the suitability of the ReproQ was tested in a two-stage quality cycle. Suitability rests on many additional requirements to be met beyond standard psychometrics. To be suited for the first stage, the questionnaire should be able to identify care providers that perform above or below some norm. The questionnaires discriminative power rests on the combined result of response (absolute number, representativeness), true outcome variation of clients, case mix adjustment, performance related variation, and measurement error. Next, the questionnaire results should give guidance where and what (or whom) to improve.

Part 3 focuses on the implementation and application of the ReproQ after its development. **Chapter 6** explores whether measuring the antenatal experiences in retrospect is valid, and consequently the number of measurements. **Chapters 7 and 8**, are the reports on two different applications of ReproQ and its the outcomes (in the second phase of the quality cycle) for maternity care improvement. **Chapter 7** describes how the results of ReproQ can be used as basis for quality improvement. **Chapter 8** studies the use of ReproQ as evaluation instrument for health care interventions (here: the implementation of Birth Centers) and its role in quality improvement.

Finally, **chapters 9 and 10** discuss and summarize the findings in the previous chapters. Additionally recommendations are offered for implementing the ReproQ and future research.

The ReproQ (key version) is added as **appendix**.

REFERENCES

1. Mohangoo AD, Buitendijk SE, Szamotulska K, et al. Gestational age patterns of fetal and neonatal mortality in Europe: results from the Euro-Peristat project. *PLoS One.* 2011;6(11):e24727.
2. Zeitlin J, Mohangoo AD, Delnord M, Cuttini M, Committee E-PS. The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *J Epidemiol Community Health.* 2013;67(12):983-985.
3. Zeitlin J, Szamotulska K, Drewniak N, et al. Preterm birth time trends in Europe: a study of 19 countries. *Bjog.* 2013;120(11):1356-1365.
4. Bonsel GJ, Birnie E, Denktas S, Poeran J, Steegers EAP. *Trends in the perinatal mortality. Description study pregnancy and birth 2010.* Rotterdam: Erasmus MC; 2010.
5. Buitendijk SE, Nijhuis JG. [High perinatal mortality in the Netherlands compared to the rest of Europe]. *Nederlands tijdschrift voor geneeskunde.* 2004;148(38):1855-1860.
6. de Graaf JP, Ravelli AC, de Haan MA, Steegers EA, Bonsel GJ. Living in deprived urban districts increases perinatal health inequalities. *J Matern Fetal Neonatal Med.* 2013;26(5):473-481.
7. Mohangoo AD, Buitendijk SE, Hukkelhoven CW, et al. [Higher perinatal mortality in The Netherlands than in other European countries: the Peristat-II study]. *Nederlands tijdschrift voor geneeskunde.* 2008;152(50):2718-2727.
8. Steering committee Pregancy and Birth. *A good start. Save care during pregnancy and birth.* Utrecht2009.
9. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Soc Sci Med.* 2000;51(11):1611-1625.
10. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997;45(12):1829-1843.
11. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med.* 1998;47(10):1573-1588.
12. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med.* 1994;38(4):509-516.
13. Poeran J, Borsboom GJ, de Graaf JP, Birnie E, Steegers EA, Bonsel GJ. Population attributable risks of patient, child and organizational risk factors for perinatal mortality in hospital births. *Matern Child Health J.* 2015;19(4):764-775.
14. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the Dutch obstetric care system. *BMC health services research.* 2014;14:622.
15. Offerhaus PM, Hukkelhoven CW, de Jonge A, van der Pal-de Bruin KM, Scheepers PL, Lagro-Janssen AL. Persisting rise in referrals during labor in primary midwife-led care in the Netherlands. *Birth.* 2013;40(3):192-201.
16. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC pregnancy and childbirth.* 2009;9:18.
17. Wiegers TA, van der Zee J, Keirse MJ. Transfer from home to hospital: what is its effect on the experience of childbirth? *Birth.* 1998;25(1):19-24.
18. Christiaens W, Gouwy A, Bracke P. Does a referral from home to hospital affect satisfaction with childbirth? A cross-national comparison. *BMC health services research.* 2007;7:109.
19. Garcia-Barbero M. Trends in integrated care-reflections on concetual issues. . *Nature.* 1988;332.
20. Shawe J, Delbaere I, Ekstrand M, et al. Preconception care policy, guidelines, recommendations and services across six European countries: Belgium (Flanders), Denmark, Italy, the Netherlands, Sweden and the United Kingdom. *The European journal of contraception & reproductive health care : the official journal of the European Society of Contraception.* 2015;20(2):77-87.
21. van der Zee B, de Beaufort I, Temel S, de Wert G, Denktas S, Steegers E. Preconception care: an essential preventive strategy to improve children's and women's health. *Journal of public health policy.* 2011;32(3):367-379.

22. van Voorst S, Plasschaert S, de Jong-Potjer L, Steegers E, Denktas S. Current practice of preconception care by primary caregivers in the Netherlands. *The European journal of contraception & reproductive health care : the official journal of the European Society of Contraception.* 2016;21(3):251-258.
23. van Voorst SF, Vos AA, de Jong-Potjer LC, Waelput AJ, Steegers EA, Denktas S. Effectiveness of general preconception care accompanied by a recruitment approach: protocol of a community-based cohort study (the Healthy Pregnancy 4 All study). *BMJ open.* 2015;5(3):e006284.
24. Posthumus AG, Birnie E, van Veen MJ, Steegers EA, Bonsel GJ. An antenatal prediction model for adverse birth outcomes in an urban population: The contribution of medical and non-medical risks. *Midwifery.* 2016;38:78-86.
25. Quispel C, Schneider TA, Bonsel GJ, Lambregtse-van den Berg MP. An innovative screen-and-advice model for psychopathology and psychosocial problems among urban pregnant women: an exploratory study. *Journal of psychosomatic obstetrics and gynaecology.* 2012;33(1):7-14.
26. van der Kooy J, De Graaf JP, Kolder ZM, et al. A newly developed scavenging system for administration of nitrous oxide during labour: safe occupational use. *Acta anaesthesiologica Scandinavica.* 2012;56(7):920-925.
27. van der Kooy J, de Graaf JP, Birnie DE, Denktas S, Steegers EA, Bonsel GJ. Different settings of place of midwife-led birth: evaluation of a midwife-led birth centre. *SpringerPlus.* 2016;5(1):786.
28. Hermus MAA, Boesveld IC, Hitzert M, et al. Defining and describing birth centres in the Netherlands - a component study of the Dutch Birth Centre Study. *BMC pregnancy and childbirth.* 2017;17(1):210.
29. Wiegers T, de Graaf H, van der Pal K. The rise of birth centres and their role in health care [in Dutch: De opkomst van geboortecentra en hun rol in de zorg]. . *Tijdschrift voor gezondheidswetenschappen.* 2012;90(8):475-478.
30. Stichting Perinatale Registratie Nederland. *Perinatal zorg in Nederland 2014. (Perinatal care in the Netherlands 2014).* Utrecht: Stichting Perinatale Registratie Nederland;2014.
31. Murray CJL, Frenk L. Summary measures of population health in the context of the WHO framework for health system performance assessment. In: Murray CJL, Evans DB, eds. *Health Systems Performance Assessment*. Geneva, Swiss: World Health Organisation,; 2002.
32. Darby C, Valentine NB, Murray CJL, de Silva A. *Wolrd health organization (WH): strategy on measuring responsiveness.* Geneva: World Health Organization,;2000.
33. Valentine N, de Silva A, Kawabata K, Darby C, Murray C, Evans B. Health system responsiveness: concepts, domains and measurement. In: Murray C, Evans B, eds. Geneva, Switzerland: World Health Organization; 2003:573-596.
34. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
35. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
36. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
37. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
38. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
39. Pohl RF, Bender M, Lachmann G. Hindsight bias around the world. *Exp Psychol.* 2002;49(4):270-282.
40. Ruoss M. [Pain patients show a higher hindsight bias]. *Z Exp Psychol.* 1997;44(4):561-588.
41. Sixma H, Hendriks M, De Boer D, Delnoij D. Handboek CQI ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument [Manual CQI development: guidelines and regulations for the development of a CQI measurement]. 2008; http://www.nivel.nl/sites/default/files/bestanden/Handboek-CQI-Ontwikkeling.pdf.

42. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7(5):541-546.
43. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483
44. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* Oxford, United Kingdom: National Perintal Epidemiology Unit;2010.
45. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg
46. Wiegers TA, Keirse MJ, Berghs GA, van der Zee J. An approach to measuring quality of midwifery care. *J Clin Epidemiol.* 1996;49(3):319-325.

Part 1

Development

2

Measuring Client Experiences in Maternity Care Under Change: Development of a Questionnaire Based on the WHO Responsiveness Model

M. Scheerhagen, H.F. van Stel, E. Birnie, A. Franx, G.J. Bonsel

PLoS One. 2015; 10(2): e0117031

ABSTRACT

Background. Maternity care is an integrated care process, which consists of different services, involves different professionals and covers different time windows. To measure performance of maternity care based on clients' experiences, we developed and validated a questionnaire.

Methods and findings. We used the 8-domain WHO Responsiveness model, and previous materials to develop a self-report questionnaire. A dual study design was used for development and validation. Content validity of the ReproQ-version-0 was determined through structured interviews with 11 pregnant women (≥28 weeks), 10 women who recently had given birth (≤12 weeks), and 19 maternity care professionals. Structured interviews established the domain relevance to the women; all items were separately commented on. All Responsiveness domains were judged relevant, with Dignity and Communication ranking highest. Main missing topic was the assigned expertise of the health professional. After first adaptation, construct validity of the ReproQ-version-1 was determined through a web-based survey. Respondents were approached by maternity care organizations with different levels of integration of services of midwives and obstetricians. We sent questionnaires to 605 third trimester pregnant women (response 65%), and 810 women 6 weeks after delivery (response 55%). Construct validity was based on: response patterns; exploratory factor analysis; association of the overall score with a Visual Analogue Scale (VAS), known group comparisons.

Median overall ReproQ score was 3.70 (range 1 – 4) showing good responsiveness. The exploratory factor analysis supported the assumed domain structure and suggested several adaptations. Correlation of the VAS rating and overall ReproQ score (antepartum, postpartum) supported validity ($r=0.56$; 0.59, $p<0.001$ Spearman's correlation coefficient). Pre-stated group comparisons confirmed the expected difference following a good vs. adverse birth outcome. Fully integrated organizations performed slightly better (median=3.78) than less integrated organizations (median=3.63; $p<0.001$). Participation rate of women with a low educational level and/or a non-western origin was low.

Conclusions. The ReproQ appears suitable for assessing quality of maternity care from the clients' perspective. Recruitment of disadvantaged groups requires additional non-digital approaches.

INTRODUCTION

Performance of maternity care is primarily determined by its health outcomes, in particular mortality and morbidity of mother and child over the short and long term. Such outcomes differ globally, countrywise, and also within countries where health care quality differences may be in part responsible[1-5].

Another dimension of maternity care performance is the way that clients (here primarily the women involved) experience the care provided. This includes whether they feel secure, feel treated with respect, feel adequately informed; are facilities in a broad sense accessible and client-friendly. These client experiences with health care provision are supposed to be important for two reasons: 1) client experiences represent an independent outcome of performance, which may guide choices of health care provider if outcomes are similar[6]; 2) client experiences may affect clinical outcomes through several ways, hence may act as determinant of the aforementioned outcomes in mother and child[7-10]. According to the World Health Organization (WHO), which developed an influential concept to measure client experiences, adequate client orientation ultimately relates to respecting human rights, specified for the context of health care provision[6,11,12].

To achieve uniform measurement of client experiences as a performance indicator, the WHO elaborated the so-called Responsiveness model, after comprehensive preparatory studies and consultation. Following this model, responsiveness is defined as the way a client is treated by the professional and the environment in which the client is treated, where eight different domains are suggested to cover the concept. This model deliberately focuses on individual experiences rather than characteristics of processes or structures, acknowledging that between and even within countries the same client experiences may be arrived at by various means. The model has been shown to enable comparison of experienced performance within and between countries on a general level[6,13].

So far, the responsiveness questionnaires were never specified to a health care subsystem, such as maternity care. We selected the WHO responsiveness model to measure client experiences in maternity care in the Netherlands, for reasons explained below.

Measurement of maternity care performance in general is a challenge, because maternity care consists of different services (e.g. antenatal check-ups, care during the delivery); different time windows (antepartum phase, childbirth, postpartum phase) and involves several professions; and professionals (e.g., obstetricians, midwives, and maternity nurses) where many tasks are executed interchangeably.

Seen from the client's perspective, the health system in many countries shows considerable variety in health care arrangements, the location of organizations (e.g. urban vs. rural), and overall integration.

This is particularly true in the Netherlands where currently the maternity care is changing from a two-tier system to an integrated care system[14-18]. The current dominant two-tier system is based on strict division of tasks, with primary care though midwives and

general practitioners for assumed low-risk pregnant women, and secondary/tertiary care for assumed high-risk women in hospitals and perinatal centers. Primary care and secondary care professionals each have their own professional autonomy, responsibilities, and financial arrangements, and integration of processes and risk standards is limited. In view of the unsatisfactory performance of the Dutch maternity care system (perinatal outcome, maternal outcome, system weaknesses e.g. in risk management and 24/7 hospital quality), maternity care shifts towards integrated care, following the 2010 advice of a National Committee on Perinatal Care established by the Ministry of Health[3,4,16,19,20]. Integrated care combines the delivery and organization of health services; it assumes one clinical perspective, one risk management approach and one client orientation[21].

Existing indicators and questionnaires all appeared limited for our purposes. They either focus on specific processes, monodisciplinary perspectives or assume a specific maternity care organization; they usually contain additional modules on outcomes and procedural facts, and lack a formal aggregate scoring system for the client's experience allowing a graded quality judgment[22]. For example, the questionnaires of the British National Health Service (Women's Experience of Maternity Care)[22] and the National Perinatal Epidemiology Unit[23] include only part of the responsiveness domains, focusing on the personal quality of services. The Dutch Consumer Quality Index for primary maternity care[24] and a similar survey for postnatal care[25] focus on the care delivered by one professional group (community midwife, maternity nurse) for specific phases (antenatal, delivery, first postnatal week) assuming monodisciplinary care as standard, i.e. without any involvement of hospital, gynaecologist or paediatrician. Two other comprehensive interviewer-based instruments are obviously not suited for self-report. The Maternity Experiences Survey from Canada assumes additional explanatory support of an interviewer, and its length precludes routine application[26]. Prior to the ReproQ, we developed a structured face-to-face interview based on the WHO responsiveness concept to evaluate care in an integrated birth centre, which includes clinical postdelivery services[27]. Like the Maternity Experiences Survey this interview was too long for routine application, and results suggested that after a complicated delivery, bias could occur in the report of client experiences antenatally ("carry back" effect[28]). Other surveys, not listed here, primarily ask for the presence of structural features or care processes rather than for the performance-as-experienced. International comparisons of health services[29] have made clear that one cannot easily rely on the structural features, as a proxy for the actual client centeredness of services, in particular in case of disadvantaged groups. The WHO model seemed appropriate and suitable in this case as starting point for a uniformly applicable questionnaire on client experiences, as it allows for measurement regardless of the particular organizational and professional characteristics. We expect that this questionnaire is sensitive for performance characteristics that benefit from integration, such as – in terms of the WHO domains – Communication, Prompt attention, Information continuity, etc. The questionnaire may also be sensitive for potential negative aspects of

integration such as decreased autonomy if care becomes more rule-based. Existing indicators and questionnaires either focus on processes and structural features (from a professional point of view) of maternity care, or are to some extent restricted to one organizational structure[22-24], justifying our comprehensive approach on the base of a proven concept.

The study presented here describes the development of a client experiences questionnaire on the basis of the WHO responsiveness model, and presents basic psychometric evidence.

METHODS & MATERIALS

The development of the questionnaire, called the ReproQ, covered three phases: 1) overall design and specific item generation for the client experiences following the WHO concept; 2) interview study involving relevant stakeholders to determine the content validity of the null version of the ReproQ; 3) survey study in 4 different regions to enable constructive psychometric analysis. Prior to the description of the methods used in these phases, we describe the seven theoretical considerations on which the ReproQ is based. The phasing is shown in Figure 1.

Theoretical considerations

Content

1) The WHO responsiveness model was the conceptual basis. This model consists of four domains concerning the interaction between the client and health professional (Dignity, Autonomy, Confidentiality, and Communication), and four domains concerning the organizational structure (Prompt attention, Access to family and community support, quality of Basic amenities, and Choice and continuity of care)[6,13].
2) In agreement with the WHO model, the operationalization of the concept into experience items avoided any implicit preference toward provider or organization structures, leaving room to different organization structures and different levels of integrated care (high/low). We did not measure integral working as such; moreover, we assumed performance in terms of the WHO responsiveness concept would benefit from more integration, if performed well.
3) The questionnaire focussed on performance-as-experienced by the client, rather than on structural features or processes.

Coverage

4) The mother is the principle bearer of experiences, because choices and decision-making in maternity care delivery generally rest with the mother or mother-to-be. In addition, the child's father may not invariably be a desirable or available co-respondent. Obviously, responsiveness cannot be reported by the neonates themselves.
5) From a system's point of view, maternity care actually consists of service delivery that is different during pregnancy, during childbirth and postpartum care. The

antepartum phase can be defined as monitoring intermittent preventive care, mostly in an ambulatory facility. Screening is a particular feature at onset of antenatal care. The delivery is a single, high impact process, which shows many features of acute curative care. Postpartum care aims at monitoring the health of both mother and child, and at empowering the parents for the future. In these three phases, the interaction with health care professionals, facilities, and the time axis of experiences are quite different. We developed two "mirror" versions of the questionnaire; one to measure experiences during pregnancy (antepartum) and one to measure experiences during delivery and thereafter (postpartum).

Both versions are symmetrical, in that the same type of experiences are asked for and the way these are asked for is also identical, yet each item is adapted to the context (antepartum vs. postpartum). In each version we asked the client to judge each item during two reference periods: in the antepartum questionnaire the first and second half of pregnancy, in the postpartum questionnaire the event of labour and birth, and the subsequent postpartum week. Consequently, responses on all responsiveness items existed for 4 different reference periods.

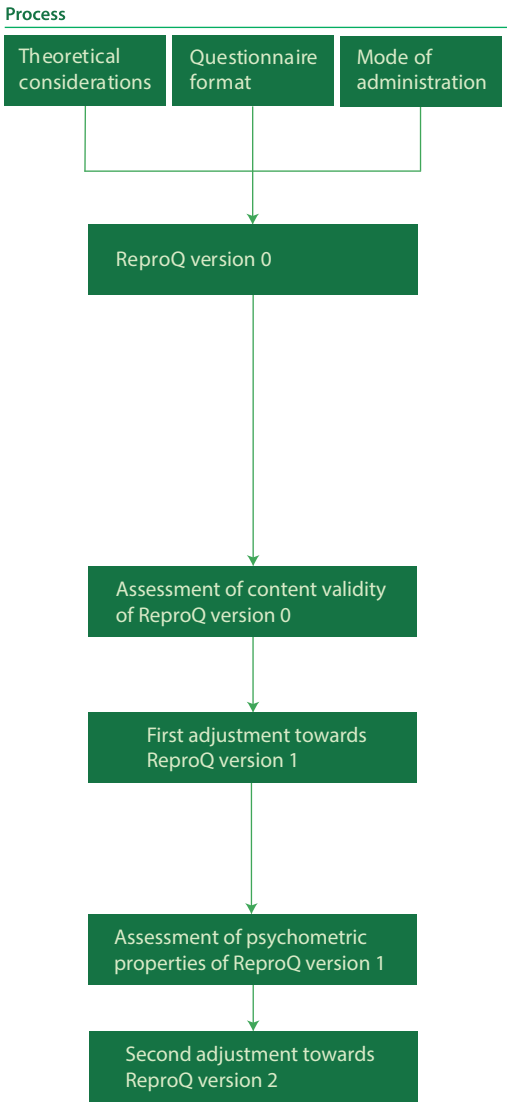
Feasibility

6) The questionnaire was intended for self-report of clients, without support, and was primarily developed as online survey. A paper version should be also available, limiting the possible complexity of the digital version.
7) The questionnaire was suitable for clients with low educational level (defined as duration ≤6 years for migrant women and ≤8 years for women of Dutch origin) and migrants and clients of non-western origin. This was achieved by the following: a) the response mode uniformly used 4 simple categories: "never", "sometimes", "often", and "always", with a numerical range of 1 (worst) to 4 (best); b) items consist of short sentences; c) common language was used (reading level B1, checked by word frequency lists[30]); testing by members of the target group. We are aware that illiterate clients need another approach, most likely a structured interview.

Survey structure and item generation

The questionnaire consists of five sections, i.e.: 1) information about the current care process, the location of care (e.g. home or hospital) and the dominant health professional delivering care (e.g. midwife or obstetrician); 2) the clinical outcome of both mother and child, as perceived by the mother in non-medical terms; 3) the client experiences in terms of the eight key domains of the responsiveness model; 4) information about previous pregnancies; 5) socio-demographic characteristics of the client.

Figure 1. Flowchart of the developmental process of the ReproQ.



Section 3 is the key section of the ReproQ. For the generation of the items of this section we used four sources. First we looked at the responsiveness items of the World Health Survey and Multicountry Survey model[6], adapting items with contextual information of maternity care. Second, we used items generated for a previously developed face-to-face interview[27]. Third, we explored published questionnaires on the same or related concepts concerning maternity care[22-24]. Finally, we used the manual of the Dutch Consumer Quality Index method to measure client experiences[31].

The other sections were developed to enable interpretation of the experiences, and supplementary discriminative content validation, as reported in this paper. The elaboration of these sections was based on existing formats and will not be discussed further.

Content validity: interviews

Content validity of the pilot version of the ReproQ (version 0) was determined through structured interviews, supported by questionnaires, with 11 pregnant women, 10 women who recently had given birth (≤12 weeks postpartum) and 19 maternity care professionals (7 midwifes, 4 obstetricians, 2 maternity nurses, 4 executives and 2 perinatal health officers). In Spring 2012, the participating clients were approached in three different maternity care organizations in The Netherlands with different levels of integration: 1) a fully integrated midwifery practice and a peripheral hospital (Roosendaal); 2) a fully integrated midwifery practice and a university hospital (Utrecht); 3) a clinic from the university hospital in Rotterdam, with an adjacent birth centre (Rotterdam). The hospitals involved, and the birth centre provided care to clients of several associated primary care midwifery practices and clients, which were already under care from the hospital. Clients were approached either by their professional or a member of the research team. The maternity care professionals were recruited from the same facilities through their team manager.

We intended to perform a group interview with each group of relevant stakeholders in each center, resulting in altogether nine group interviews. We intended to include a minimum of six participants per interview. All interviews were chaired and performed by the research team. The number of participants for each organization is shown in table A1 in the appendix.

The group interviews of about 2 hours followed a common structure: 1) prioritisation of the responsiveness domains; 2) two comments on each item (a. contents and b. grammar/readability); these were first written down for each item separately, and subsequently discussed in plenum; 3) systematic check for missing topics or perspectives of the questionnaire. Health professionals were additionally asked to rate the suitability of the experience items of the questionnaire (ReproQ core, section 3) from the perspective for women with a low educational level and non-western women. Because they regularly encounter many of these women during their consultation hours, we assumed that they could give a reasonable judgment of the suitability. They separately rated the suitability for

women with a low educational level and for women with different ethnicities on a five-point scale [strongly agree-strongly disagree].

More in detail, the client interview first invited the participants to individually describe their wishes and possible improvements concerning the maternity care they had received. Discussion could follow. Second, clients were asked which two of the eight Responsiveness-domains were most important to them. Finally, the clients were asked to fill out the null version of the questionnaire; comments were noted and discussed plenary. Each group client interview lasted about 2 hours. We performed some individual interviews, when the number of participating clients was less than the required six participants per group interview. Each participant received a compensation of €20 ($27, £16).

The group interviews with maternity care professionals lasted on average 1.5 hours and were unrewarded. In the group interviews with clients, 7 pregnant women and 9 women who recently had given birth participated. In addition, we interviewed 4 pregnant women and 1 woman who recently had given birth individually. In the group interviews with health professionals, 7 midwives, 4 obstetricians, 2 maternity nurses, 4 executives and 2 perinatal health officers participated.

The null version of the ReproQ was adjusted based on the joint comments, where comments of clients and health professionals were regarded as equally relevant. We assumed that the item content to be valid if the comments involved no or minor changes in item wording or response categories.

Survey study to obtain psychometric characteristics

We obtained psychometric characteristics of the adjusted questionnaire in a subsequent survey study. Pregnant women and women who recently had given birth were asked for participation when they visited their care provider. After written informed consent, they received an invitation by email to fill out the web-based questionnaire. Patients were locally recruited with the support of the organisation.

To qualify for the antepartum questionnaire, women should have a gestational age less than 34 weeks; to qualify for the questionnaire concerning the delivery and postpartum care, women should have given birth less than 6 weeks earlier. The antepartum questionnaire was sent in the 34th week of their pregnancy, the postpartum questionnaire was sent 6 weeks after the expected date of delivery. Non-responding women received an e-mail reminder 2 weeks after they received the initial questionnaire.

Four maternity care organizations participated for client recruitment. Three of these also participated in the interview study. The additional organization included four hospitals and four midwifery practices.

Altogether a wide range of organisational structures and client populations was covered. To determine the psychometric characteristics of the questionnaire, we aimed at a minimum of 300 completed antepartum and 300 completed postpartum questionnaires. Because the

questionnaire exists of two versions, that are not identical, we aimed at a sample size of 300 respondents for both versions of the questionnaire. The sample size was based on the Dutch manual to develop Consumer Quality questionnaires[31].

Analysis

Interviews relevant stakeholders to determine the content validity

The prioritised domains will be reported in percentage of domains ranked first or second.

The items were primarily adapted based on the detailed individual written comments. Combining the comments per item resulted in 1) items needing no change; 2) items to be simplified or changed to avoid textual ambiguities; 3) adaptation of the response mode in specific cases, e.g. through addition of the option "not applicable", or changes in the labels of the response levels; 4) items to be removed, if the item did not sufficiently fit to the concept or if the item showed too much overlap with other items questions.

The comments on missing domains or items are reported if multiple comments indicated such missing.

The response mode of the five point suitability-questions for women with a low educational level, and of non-Dutch origin were later reduced to three categories: agree-neutral-disagree, as extreme categories were rarely used.

Survey study followed by psychometric analyses

We invited 605 pregnant women, of whom 396 responded (65%), and invited 810 women who recently had given birth, of whom 483 responded (55%). We excluded 45 pregnant women and 50 women who recently had given birth, because 50% of their answers were missing in 2 of more domains. The first step in the analysis was the checking for response patterns, such as a floor-ceiling-effect, the computation of the percentage missing-values per item, and the computation of the digitally measured response time. The second step involved analysis of the construct validity using Exploratory Factor Analysis (EFA)[32]. The main goal was to identify items that required replacement to another domain, rewording, or removal. Because we use a so-called formative measurement model (pre-stated domain structure) and not a reflective model, the decisions on which item belongs to which domain finally are based on content and the EFA combined, rather than EFA alone.

The analyses were intended to be performed separately for the four phases of maternity care, namely first half pregnancy, second half pregnancy, birth, and postnatal care. However, as answer patterns for the first and second half of the pregnancy were close to identical, we only present data of the second half of pregnancy, and data of birth and postnatal care (3 reference periods).

In the EFA for labour and birth, and postnatal care, the three questions of the domain Basic amenities were not included, because the number of respondents was too small due to routing in the questionnaire. The EFA was conducted as a principal components analysis

followed by orthogonal rotation (Varimax)[32]. The factors were determined by the Kaiser criterion (i.e. an Eigenvalue >1). In addition, we computed Cronbach's alpha to determine the internal consistency of each factor. Note that internal consistency of items may be empirically low despite a close relation in terms of contents: e.g. items on the accessibility all refer to one basic concept, yet the travelling distance to the facility is not empirically associated to the accessibility by phone.

Third, convergent validity was tested by the association between an overall 10-point VAS rating with the overall client experience of women, combining all domain responses. This 10-point VAS rating was based on the recent recommendations of the National Patient Survey Coordination Centre[33]. The overall client experiences score was obtained by first computing an average score per domain (where the 1, 2, 3 or 4 response was treated numerically), and then computing an unweighted average across the 8 domain scores, resulting in an overall experience score with range 1 – 4. The association of women's global rating with their experience as a client was expressed by Spearman's correlation coefficient (rho).

The last step was a preliminary assessment of the discriminative validity of the ReproQ by three so-called known group comparisons. The client experience was compared applying the following groupings: 1) pregnant women versus women who recently had given birth; 2) women with better vs. worse clinical outcome of their baby depending on perceived health problems by the mother and hospitalization of the baby (altogether 4 groups); and 3) women who received care in fully integrated facilities versus women who received care in less integrated facilities.

We calculated domain scores (giving a profile) and an overall ReproQ score. Domain scores were declared missing when less than half of the items of that domain were filled out. We refrained from imputation of missing data. If more than half of the domain scores were missing, no overall score was computed. Because the experience data did not show a normal distribution, we report the overall median (MD) and the interquartile range (IQR) of all Responsiveness domains. To explore if differences in performance were significant between groups, we performed a Mann-Whitney test or Kruskal-Wallis test depending on the number of determinant categories (2 or 4, respectively). Significance level was $p<0.05$, without adjustment for multiple testing, as this was an explorative study, without prior sample size calculation. For the statistical analyses we used SPSS 21.0.

General

The development process was supervised by a steering committee. This group consisted of representatives from health professionals, health insurance companies, a client-patient association, and members of the research team. Besides the steering committee, we were advised by a senior officer of the WHO engaged in the development of responsiveness measurement, with sufficient knowledge of the Dutch language.

The Medical Ethical Review Board of the University Medical Centre Utrecht approved the study protocol (study number MEC-2012-435).

RESULTS

Item generation

The test version of our antepartum questionnaire contained 30 experience items. The postpartum questionnaire contained 36 experience items. The difference is explained by items in the domains Prompt attention and Basic amenities concerning specific elements of the delivery and postnatal care, such as the facilities during hospitalization after the delivery or the presence of a maternity nurse.

Interviewing stakeholders to determine the content validity

The mean age of the participating women was 32.3 years (SD=5.5). Of the 21 women, 6 reported to be of non-Dutch origin (29%). Most women had a high education; 8 women had a low/middle education (38%). All women were married or living together. Half of the women gave birth for the first time (52%). 13 of the 21 women received care in an integrated facility (62%). The characteristics are described in Table 1. All responsiveness domains were confirmed as being relevant in general. The domains Dignity and Communication were selected as most important by clients, by health professionals from their own perspective, and from the proxy-perspective of clients with low educational level or migrant status as expressed by these professionals. Clients and health professionals gave altogether 266 comments about the items in the Responsiveness domains (roughly 1 out of 5 items received a comment). 93 (35%) of these comments were related to the clarity of the wording of items. The participants stated problems with specific terms e.g. “personal attention”, “home situation” and the meaning of “several options” in the item “Could you choose from several options for postnatal care?”.

Of the 266 comments, 119 comments (45%) concerned the relevance of items. Women noted difficulty in giving response if they had not been in a situation as described. Health professionals doubted whether some items could be judged by clients in case of high urgency of the care provided. They suggested adaptations of question or response (adding “not applicable”) in some instances.

54 comments (20%) suggested literal improvements in text of items or the response.

The topics claimed more than once to be missing were the client’s judgment of the health professional’s expertise and specific items on cultural customs and traditions of migrant women. As the ReproQ is to be used in connection to medical outcome measures, we refrained from adding an item on assigned expertise.

The suitability for women with low educational level was judged as sufficient by 10 of the 18 health professionals, while one health professional thought the questionnaire was

unsuitable for women with low education. All professionals emphasized to be cautious with the application of standard survey data collection techniques in respondents with a low educational level.

Based on all comments, we left 11 items unchanged; 7 items were slightly rephrased; 5 items were deleted; 2 items were added; and the response mode of 10 items was rephrased (adding "not applicable").

Survey study to determine psychometric characteristics

The characteristics of pregnant women and women who recently had given birth are presented in Table 1. As differences were minimal, characteristics are described combined. The participating women had a mean age of 33.1 years (SD=4.4). Of the 784 women who responded, 72 (9%) reported to be of non-Dutch origin. 71 women were not living together with the father of the child, or did not have a relationship with the father at all (9%).

The response pattern of the women generally showed high responsiveness to the client. The response modus "never" (representing an adverse experience) was not used in several items. "Never" was most often used in the item concerning "choice of health care professional" (19.7%) in the antepartum and postpartum questionnaire (18.2%). The response modus "always" was least often used in the item concerning waiting in the antepartum questionnaire (20.3%), and most by the item concerning privacy (94.3%). In the postpartum questionnaire, the response modus "always" was least often used in the item concerning the furnishing of the maternity care organizations (36.1%), and most often in the item "treated with respect" (88.6%). The per item missing rates were all below 5%. Filling out the antepartum questionnaire lasted on average 16 minutes (95% confidence interval (CI): 11 – 21min). The postpartum questionnaire took on average 14 minutes (95% CI: 11 –17min).

The EFA revealed 9 factors in the antepartum questionnaire; 7 factors in experience with delivery, and 5 factors in postnatal care. Table 2 shows the factor loadings of each item (after rotation) for pregnancy, labour and birth, and postnatal care phase separately. Factor loadings of items that deviate from the dominant factor (i.e. the domain on which most of the items of the domain loaded) are shown in italics. The factors that included two items or more had a Cronbach's alpha varying between 0.68 and 0.89. From the EFA it appears that the factor solution shows considerable commonality across the three phases.

The median score was 3.69 for the antepartum version (IQR 3.39 – 3.87). The median score of the postpartum version was 3.74 (IQR 3.45 – 3.88). In Figure 2 the global 10-point VAS rating was related to the overall ReproQ client experience score, to determine the convergent validity. A low VAS rating was associated with a lower ReproQ score in both the antepartum ($r=0.59$; $p<0.001$) and the postpartum questionnaire ($r=0.56$; $p<0.001$).

The average score combining all domains per individual has a median of 3.68 (IQR=3.40 – 3.87) antepartum, and a median of 3.73 (IQR=3.44 – 3.88) postpartum ($p=0.23$). Domain-wise, Autonomy, Dignity, Confidentiality were experienced better in pregnant

women compared women who recently had given birth (p between 0.021 and <0.0001). Women who recently had given birth had better experiences with Prompt attention, Social considerations and Choice and continuity (p between 0.033 and <0.0001).

Table 1. Characteristics of the participating women in the preparatory interview study and the ReproQ survey study.

Characteristics	Interview study	Survey study	
	% (N=21)	Antepartum % (N=351)	Postpartum % (N=433)
Age			
≤25	19	5	8
26-30	29	32	29
31-35	24	38	42
>=36	29	21	16
Missing	0	4	5
Ethnic background			
Dutch	71	85	79
Non-Dutch	29	9	9
Missing	0	6	12
Education			
Low	0	1	1
Middle	38	40	36
High	62	54	57
Education abroad	0	0	2
Missing	0	5	4
Marital status			
Married/living together	100	92	91
Relationship, not living together	0	3	3
No relationship	0	1	2
Missing	0	5	4
Parity			
Primiparous	48	51	49
Multiparous	52	45	47
Missing	0	3	4

Table 1. *Continued*

Characteristics	Interview study	Survey study	
	% (N=21)	Antepartum % (N=351)	Postpartum % (N=433)
Integrated care			
Integrated care facility	71	46	52
Non-integrated care facility	29	49	44
Missing	0	4	4
Maternity care organization			
1	52	23	25
2	29	24	27
3	–	39	23
4			
Missing		4	5

Figure 2. Convergent validity: association between overall rating of maternity care, and ReproQ score (all domains combined).

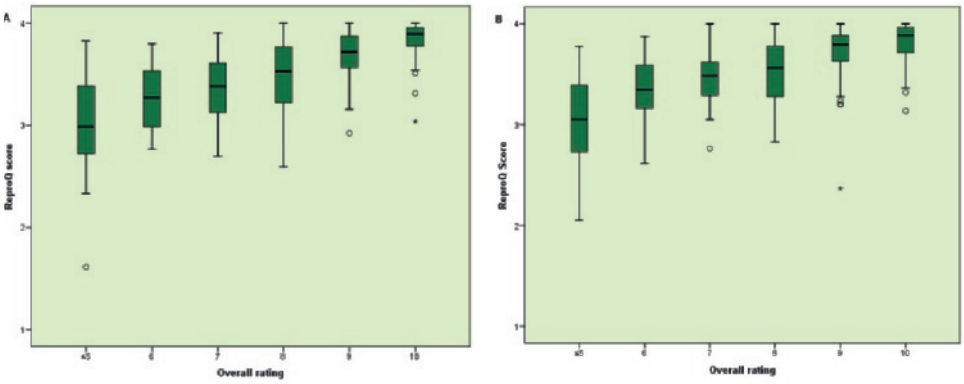
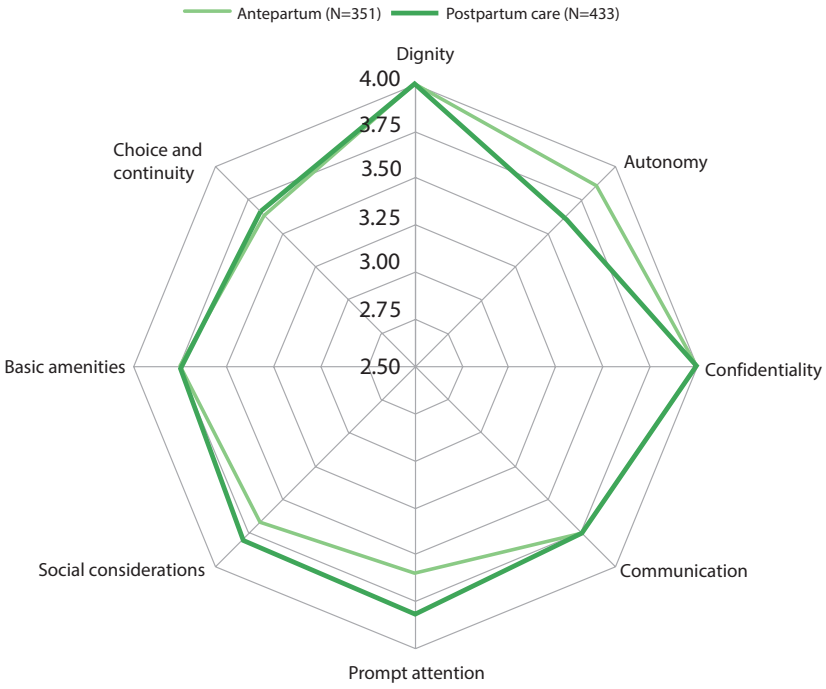


Figure 2A shows the results during the antenatal phase; Figure 2B shows the results during labour and postnatal care. The overall rating (10-point VAS scale) was significantly associated with the overall ReproQ score (i.e. the unweighted summation [range 1 – 4] of the individual eight domains), in both the antepartum ($p<0.001$) and the postpartum phase ($p<0.001$).

Figure 3. Discriminative validity: median domain-specific ReproQ-score, for antepartum and postpartum questionnaire.

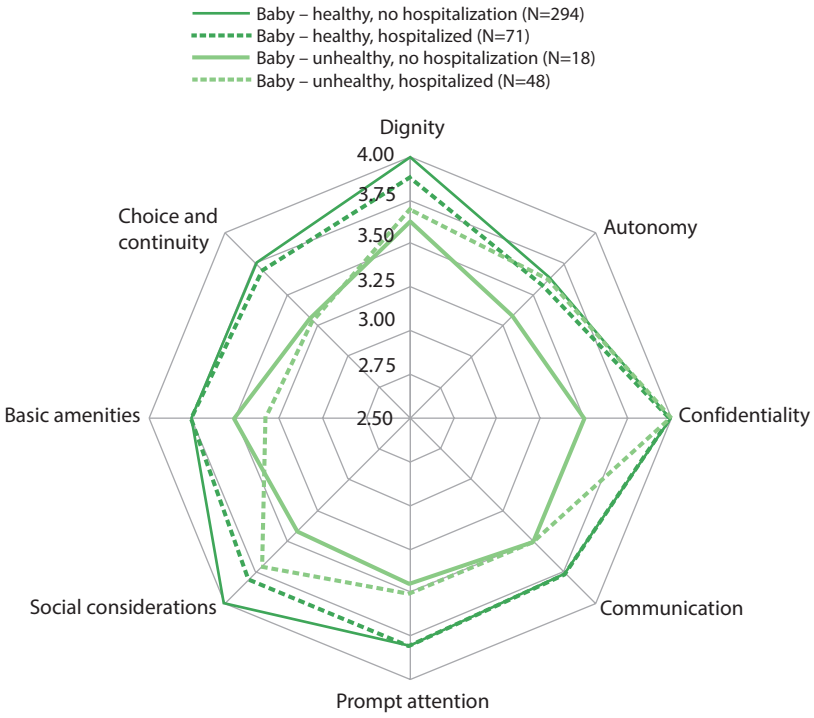


The figure shows the median domain-specific ReproQ score for antepartum (A) and postpartum (P) questionnaire. The interquartile range of the domains are as follows: Dignity: 3.8 – 4.0 (A); 3.7 – 4.0 (P); Autonomy: 3.4 – 4.0 (A); 3.3 – 4.0 (P); Confidentiality 4.0 – 4.0 (A); 3.5 – 4.0 (P); Communication: 3.5 – 4.0 (A); 3.5 – 4.0 (P); Prompt attention: 3.4 – 3.8 (A); 3.4 – 4.0 (P); Social considerations: 3.0-4.0 (A); 3.4 – 4.0 (P); Basic amenities 3.3 – 4.0 (A); 3.3 – 4.0 (P); Choice and continuity: 2.9 – 4.0 (A); 3.1 – 4.0 (P).

The results on discriminative validity showed that overall, pregnant women and women who recently had given birth had a similar overall ReproQ score (MD=3.68, IQR=3.40 – 3.87 vs. MD=3.73, IQR=3.44 – 3.88; p=0.23), see Figure 3. Domain-wise, Autonomy in pregnant women was experienced better compared to Autonomy in women who recently had given birth. Women who recently had given birth had better experiences with Prompt attention.

Women who perceived no health problems in their baby reported best overall ReproQ score (see figure 4), independent whether their baby was hospitalized (MD=3.78, IQR=3.52 – 3.87) or not (M=3.72, IQR=3.45 – 3.85). Within these groups, women with their baby hospitalized showed more negative experiences concerning Dignity and Social considerations than women whose baby was not hospitalized. Women who did perceive health problems in their baby, but whose baby was not hospitalized showed a lower median overall score (MD=3.47, IQR=3.16 – 3.80). Women whose baby had been hospitalized with (perceived) health problems showed the lowest overall scores (MD=3.42, IQR=3.10 – 3.72).

Figure 4. Discriminative validity: median domain-specific ReproQ-score according to perceived neonatal outcome and neonatal hospital admission.



The figure shows the median domain-specific ReproQ score for 4 groups: 1) women who did not perceive health problems with their baby and whose baby was not hospital-admitted (MD=3.78, IQR=3.52 – 3.87); 2) women who did not perceive health problems with their baby, but whose baby was hospital-admitted [e.g. for monitoring] (MD=3.72, IQR=3.45 – 3.85); 3) women who perceived health problems with their baby, but the baby was not hospital-admitted (MD=3.47, IQR=3.16 – 3.80); and 4) women who perceived health problems with their baby and whose baby was hospitalized (MD=3.42, IQR=3.10 – 3,72) (p>0.001). Note that the client's perception of neonatal outcome may differ from clinical judgement; here we assume the client's perspective to be primarily important.

All domains and the overall score differed significantly between the four subgroups (Kruskal-Wallis, all p<0.001).

During pregnancy, the overall ReproQ score of women who received care in a full integrated facility (MD=3.65, IQR=3.37 – 3.86) showed no significant difference with women who received care in a less integrated facility (MD=3.74, IQR=3.42 – 3.88; p=0.14) (see Figure 4). In the delivery and postpartum phase women who received care in integrated facilities had a slightly higher score compared to less integrated facilities (Md=3.78, IQR=3.53 – 3.90 vs. Md=3.63, IQR=3.34 – 3.84; p<0.001). All domains except "Choice and continuity" (p=0.062) differed significantly.

Table 2. Exploratory factor analysis of the ReproQ (antenatal, delivery, and post partum responses separately).

Domain	Item	Adaptation	Factor number and factor loading*		
			Antepartum (N=351)	Delivery (N=433)	Postpartum (N=433)
Dignity	Respecting privacy**	NC	F1 – 0.46	F1 – 0.52	*F4 – 0.59*
	Treating with respect	NC	F1 – 0.81	F1 – 0.74	F1 – 0.81
	Giving personal attention	NC	F1 – 0.67	F1 – 0.61	F1 – 0.66
	Treating with kindness	NC	F1 – 0.81	F1 – 0.72	F1 – 0.78
Autonomy	Involving client in decision-making	NC	F7 – 0.80	F3 – 0.70	F4 – 0.63
	Acceptance of treatment refusal	NC	F7 – 0.82	F3 – 0.70	F4 – 0.69
	Considering personal wishes regarding pregnancy and birth	DD	*F1 – 0.64*	*F1 – 0.48* F3 – 0.58	*F1 – 0.55* F4 – 0.48
	Offering Down's syndrome screening as free choice	NC	*F9 – 0.86*	–	–
	Involving client in decision-making on pain relief	NC	-	*F7 – 0.80*	*F5 – 0.83*
	Involving client in decision-making on setting of birth	AI AR	*F3 – 0.35* *F5 – 0.36*	*F6 – 0.67*	*F5 – 0.50*
Confidentiality	Trustworthy as health professional	DD	*F1 – 0.64*	*F1 – 0.59*	*F1 – 0.62*
	Secured provision of medical information to others	DEL	*F4 – 0.32*	*F4 – 0.42* *F5 – 0.54*	*F3 – 0.45* *F4 – 0.49*
Communication	Responsive to client questions	NC	F1 – 0.46	*F1 – 0.43* F4 – 0.48	F1 – 0.53 *F2 – 0.53*
	Consistency of advice across professionals	NC	F1 - 0.46	F4 – 0.65	F1 – 0.47
	Comprehensibility of explanation	NC	*F5 – 0.55*	F4 – 0.77	F1 – 0.50
	Provision of information while treated	NC	*F5 – 0.57*	*F1 – 0.43* F4 – 0.48	F1 – 0.46 *F2 – 0.42*
Prompt attention	Access for appointment/ contact in urgent situations	AR	*F8 – 0.63*	F2 – 0.65	*F1 – 0.45* *F2 – 0.42* F3 – 0.46
	Access for appointment/ contact without urgency	NC	*F2 – 0.37* F4 – 0.38 *F8 – 0.39*	F2 – 0.59	*F1 – 0.42* *F2 – 0.40* F3 – 0.47
	Waiting time for service	AI AR	F4 – 0.57	F2 – 0.43	F3 – 0.62
	Physical accessibility of setting	AI	F4 – 0.65	F2 – 0.65	F3 – 0.57
	Prompt phone response of health professional	AI	F4 – 0.79	F2 – 0.71	F3 – 0.48

Table 2. *Continued*

Domain	Item	Adaptation	Factor number and factor loading*		
			Antepartum (N=351)	Delivery (N=433)	Postpartum (N=433)
Social considerations	Attention for family	DEL	F2 – 0.80	F3 – 0.59	F2 – 0.65
	Taking into account of family duties when making appointments	AI	F2 – 0.76	F3 – 0.57	F2 – 0.71
	Involvement of the partner in care provision	NC	F2 – 0.76	F3 – 0.47	F2 – 0.70
Basic amenities	Decoration of setting	AI	F3 – 0.68	–	–
	Hygiene of setting	AI	F3 – 0.75	–	–
	Comfort of setting	AI	F3 – 0.66	–	–
	Playground children or other facilities	DEL	F3 – 0.46	–	–
Choice and continuity	Making service time available on request of the client	DD	–	*F2 – 0.55*	*F2 – 0.53* F3 – 0.50
	Continuity of care provision when change of individual professional (same discipline)	AI AR	F6 – 0.77	F5 – 0.53	F3 – 0.62
	Continuity of care provision when change of professional (across disciplines)	AI AR	F6 – 0.84	F5 – 0.60	F3 – 0.67
	Allowance for selecting a preferred type of health professional	NC	*F5 – 0.66*	*F6 – 0.63*	F3 – 0.43
	Being explicit on which health professional is actual in charge	NC	F6 – 0.61	*F4 – 0.49*	*F2 – 0.60*

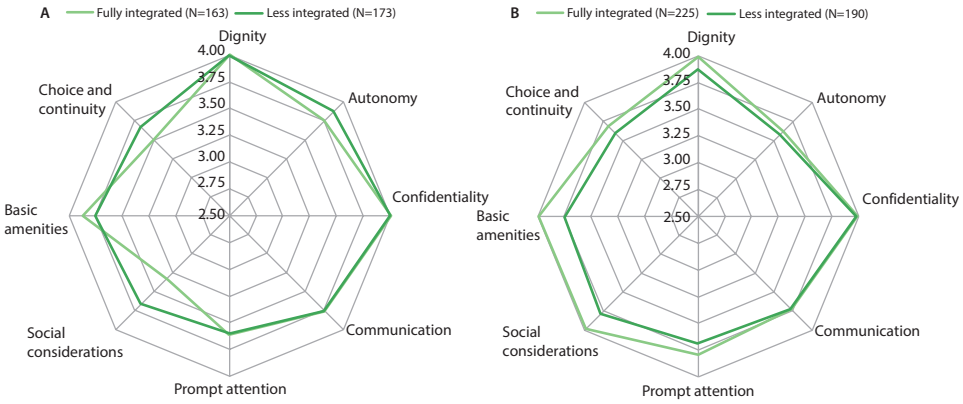
Adaptation: NC=no change; AI=adjusted item; AR=adjusted response mode; DD=assigned to different domain based on EFA; DEL=removed

* The last three columns represent 3 separate factor analyses. The number of the factor is listed in the order of the output (F1=first factor, F2=second factor, etc). Only results with a factor loading > 0.3 are shown. If an item corresponds to a factor numbers in italics, then the factor analysis apparently placed the item into another domain, as we assumed constructing the item cf. the WHO domain structure.

** For reasons of brevity, we indicate the contents of each item as a professional or maternity care organization characteristic. The question to the respondent refers to the actual experience. E.g. the first item indicated with "respecting privacy", in the ReproQ reads as "Did the caregiver consider your privacy during care provision?"

Based on the results of the statistical analyses we left 19 items unchanged; 10 items were slightly rephrased; 3 items were deleted; 4 new items from the original item pool were added; 3 items were left unchanged but formally assigned to another domain; and the response mode of 4 items was slightly rephrased (see Table 2). No changes were needed after the second round of stakeholders experts for final verification that wording of item and response were strictly neutral and unequivocal for types of organizations. The questionnaires can be accessed online: antepartum version, postpartum version.

Figure 5. Discriminative validity: median domain-specific ReproQ-score according to integration level of care facilities.



The figures show the median domain-specific ReproQ score during the antenatal phase (5A) en during labour and postpartum care (5B), for women who received care in fully integrated organizations and for women who received care in less integrated organizations. These differences were not significant during antenatal care (p=0.142), but were significant in care during labour and birth, and during postnatal care (p<0.001).

DISCUSSION

We developed a self-report questionnaire, the ReproQ, to measure the performance of maternity care from the perspective of clients. We used the WHO responsiveness model, which evaluates experienced client interactions with professionals and the care providing organisation. Content validity of the instrument was judged appropriate with balanced contribution of the WHO domains. According to participants, more attention could be given to sensitivity for the cultural background and traditions of the client, and the experienced professional expertise. The null version of the questionnaire was adjusted based on all comments, but we did not include a domain on the experienced expertise as in our opinion this should be primarily reflected in the clinical outcome, which is difficult to assess by the client.

The construct validity of the improved ReproQ version 1 was established in a survey study involving pregnant and recently delivering women. The response pattern showed overall good to excellent overall ReproQ scores, averaging over domains. The effect that participants rarely use the most negative response modus, is known from other self-report instruments in maternity care and may be partially caused by the fact that pregnancy and childbirth are not a disease and generally have good outcome[23,34-36]. The most positive response category ("always") was used the most, demonstrating a ceiling effect also shown in the maternity experience survey of the National Perinatal Epidemiology Unit[23].

The exploratory factor analyses largely confirmed the pre-stated domain structure. However, the EFA strongly suggested to rearrange and reword items from the Confidentiality domain, because these items loaded on different factors for the different stages of maternity care. Testing convergent validity, we established a clear association between the overall VAS rating and the overall ReproQ score.

The known group comparisons revealed literature-expected differences between women perceiving good vs. bad outcomes in their baby, being aware that this may be a "cross-over" effect rather than actually reflect poor responsiveness. In clients who received care in fully integrated facilities vs. less integrated facilities, we observed differences during birth and postnatal care, but not during pregnancy as might be expected as integration effects from the perspective of the client is most clearly experienced at that stage.

Strengths

The ReproQ focuses on the actual experiences of women with maternity care while existing questionnaires mainly focus on procedural aspects[22-24]. While following the adequate procedure can contribute to responsiveness, it does not replace or predict care provision, which is client-centered. For example, the provision of written information can be a valuable standard procedure, but it requires verification of utilization and understanding of the information.

The ReproQ is unique in the coverage of the eight responsiveness domains, which were all considered valuable. The questionnaire of the National Health Service in the United Kingdom included only 6 of the 8 Responsiveness domains, often using one specific item within a domain. Prompt attention was e.g. indicated by the item "were you given the help you needed?"[22]. This item was similar in the questionnaire of the National Perinatal Epidemiology Unit concerning women's experience with maternity care[23]. As it combines promptness and perceived adequacy, response is difficult to interpret.

To prevent cross-over effects from labour and birth to the antenatal experiences[28], we created two separate questionnaires to measure the experiences during pregnancy and the experiences during delivery and postpartum care. This facilitates quality improvement as the services involved usually are different.

Limitations

This study had several limitations.

First, fewer clients participated in the group interviews than anticipated. In order to cover all relevant perspectives and to maximize input on the issue of comprehensibility for the deprived, we conducted additional individual interviews. In both forms all participants first wrote down their individual comments (positive/negative/change) on contents and readability for each questionnaire item of the ReproQ core separately. No discussion or exchange was allowed in the group session at this stage. In the group sessions, these

items were then presented one by one, and discussed if asked for. The items were primarily adapted based on the detailed individual written comments, which frequently converged; occasionally the plenary discussion was used to solve an arbitrary wording choice. We assume the combination of group and individual sessions did not compromise the results.

Second, there is no reference standard available to measure performance from the perspective of clients, which makes it hard to establish the quality of the measured concept. We believe however that the responsiveness model provides a solid conceptual base, confirmed by extensive testing during its development and thereafter[6,13]. The comprehensiveness and cross-cultural suitability has been confirmed in our study.

Third, women with a low educational level were underrepresented in our studies despite repeated and considerable efforts to engage them. An explanation may be lack of perceived control of these women, which is reflected in reluctance to participate: they do not believe that participation or responding matters[37].

Fourth, a minority of the non-western women participated. This percentage (9%) is lower than the percentage of non-Dutch pregnant women in the Netherlands (non-Dutch: 16%)[38]. Possible explanations include a language barrier[39,40] and our reliance on an anonymous digital procedure. Perhaps the frequent coexistence of low education and non-Western ethnicity plays a role[41]. To increase their participation, the questionnaire could be adapted by adding specific questions or by translating the questionnaire into other languages. For both non-response prone groups the questionnaire could also be presented differently. For example, with assistance of an independent third person, or by using a face-to-face interview. Another option would be to ask key figures of their local society to promote participation.

Future use

The resulting questionnaire may be used in various types of evaluation studies, dedicated to compare specific interventions or specific organization structures, or health care providers.

From its conceptual base – a complement to medical outcome – it follows that outcomes, like mortality of both mother and child, or compound measures like the Perinatal Adverse Outcome Index[42] are unconditionally required for overall judgement. Interpretation of the relevance of average ReproQ differences requires further study.

Future research

Further research is needed on the discriminative capacity of the ReproQ to show differences between care providers, and on the interpretation and relevance of observed differences. Also, testing the proposed domain structure in a new sample using confirmatory factor analysis is needed.

CONCLUSION

We developed a client experience questionnaire ("ReproQ") to measure maternity care performance based on the WHO responsiveness model. After content analysis the improved ReproQ questionnaire showed acceptable convergent and satisfactory discriminative validity. Participation of disadvantaged groups in measurement of client experiences may require additional approaches.

ACKNOWLEGDEMENTS

We are grateful for all maternity care organizations that took part in our study, with special thanks to Richard Pal, Jan Oostenbrink, Marlies Bartels, Hanneke de Graaf, Jannie de Vos, Valentine Geluk, Anky van Hecke and Karin Scheele. They inspired the maternity care organizations to participate and contributed to the translation of the WHO concept into the survey presented. We thank Jacoba van der Kooy and Nicole Valentine (WHO) for their extensive support designing the questionnaire. We also thank Marit Hitzert, who helped performing the interviews and the survey study. In addition we are grateful for the support of Augustinus JP Schrijvers and Eric AP Steegers for valuable organisational and professional support.

FUNDING

This study was funded by Stichting Miletus, www.stichtingmiletus.nl. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. de Graaf JP, Ravelli AC, de Haan MA, Steegers EA, Bonsel GJ. Living in deprived urban districts increases perinatal health inequalities. *J Matern Fetal Neonatal Med.* 2013;26(5):473-481.
2. de Graaf JP, Schutte JM, Poeran JJ, van Roosmalen J, Bonsel GJ, Steegers EA. Regional differences in Dutch maternal mortality. *Bjog.* 2012;119(5):582-588.
3. Mohangoo AD, Buitendijk SE, Szamotulska K, et al. Gestational age patterns of fetal and neonatal mortality in Europe: results from the Euro-Peristat project. *PLoS One.* 2011;6(11):e24727.
4. Zeitlin J, Mohangoo AD, Delnord M, Cuttini M, Committee E-PS. The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *J Epidemiol Community Health.* 2013;67(12):983-985.
5. Zeitlin J, Szamotulska K, Drewniak N, et al. Preterm birth time trends in Europe: a study of 19 countries. *Bjog.* 2013;120(11):1356-1365.
6. Valentine N, de Silva A, Kawabata K, Darby C, Murray C, Evans B. Health system responsiveness: concepts, domains and measurement. In: Murray C, Evans B, eds. Geneva, Switzerland: World Health Organization; 2003:573-596.
7. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Social science & medicine (1982).* 2000;51(11):1611-1625.
8. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Social science & medicine (1982).* 1997;45(12):1829-1843.
9. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Social science & medicine (1982).* 1998;47(10):1573-1588.
10. Williams B. Patient satisfaction: a valid concept? *Social science & medicine (1982).* 1994;38(4):509-516.
11. de Silva A. A framework for measuring responsiveness. *Global Programme on Evidence for Health Policy Discussion Paper 32* 2000.
12. Gostin L. The domains of health responsiveness: a human rights assessment. 2002.
13. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
14. Christiaens W, Gouwy A, Bracke P. Does a referral from home to hospital affect satisfaction with childbirth? A cross-national comparison. *BMC health services research.* 2007;7:109.
15. Offerhaus PM, Hukkelhoven CW, de Jonge A, van der Pal-de Bruin KM, Scheepers PL, Lagro-Janssen AL. Persisting rise in referrals during labor in primary midwife-led care in the Netherlands. *Birth.* 2013;40(3):192-201.
16. van der Velden K. *Safe care during pregnancy and birth.* 2009.
17. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC pregnancy and childbirth.* 2009;9:18.
18. Wiegers TA, van der Zee J, Keirse MJ. Transfer from home to hospital: what is its effect on the experience of childbirth? *Birth.* 1998;25(1):19-24.
19. Bonsel GJ, Birnie E, Denktas S, Poeran J, Steegers EAP. *Trends in the perinatal mortality. Description study pregnancy and birth 2010.* Rotterdam: Erasmus MC; 2010.
20. Ravelli AC, Jager KJ, de Groot MH, et al. Travel time from home to hospital and adverse perinatal outcomes in women at term in the Netherlands. *Bjog.* 2011;118(4):457-465.
21. Gröne O, Garcia-Barbero M. *Trends in Integrated Care - Reflections on Conceptual Issues.* . 2002.
22. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483

23. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* 2010.
24. Wiegers TA, Keirse MJ, Berghs GA, van der Zee J. An approach to measuring quality of midwifery care. *J Clin Epidemiol.* 1996;49(3):319-325.
25. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg
26. Dzakpasu S, Kaczorowski J, Chalmers B, Heaman M, Duggan J, Neusy E. The Canadian maternity experiences survey: design and methods. *J Obstet Gynaecol Can.* 2008;30(3):207-216.
27. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the Dutch obstetric care system. *BMC health services research.* 2014;14:622.
28. Christensen-Szalanski JJ. Discount functions and the measurement of patients' values. Women's decisions during childbirth. *Medical decision making : an international journal of the Society for Medical Decision Making.* 1984;4(1):47-58.
29. Murray CJL, Evans BE. *Health systems performance assessment: debates, mothods and empriricism.* Geneva: World Health Organization,; 2003.
30. BureauTaal. Dictionairy language level B1 and A2. 2009; http://www.zoekeenvoudigewoorden.nl.
31. Koopman L, Sixma H, Hendriks M, de Boer D, Delnoij D. *Manual CQI Development: Guidelines and requirements for the development of a CQI measurement.* 2011.
32. Stevens JP. *Applied multivariate statistics for the social sciences.* New York: Routledge; 2009.
33. Graham C, Maccormick S. *Overarching questions for patient surveys: development report for the Care Quality Commission (CQC).* Picker Institue Europe;2012.
34. Peterson WE, Charles C, DiCenso A, Sword W. The Newcastle Satisfaction with Nursing Scales: a valid measure of maternal satisfaction with inpatient postpartum nursing care. *J Adv Nurs.* 2005;52(6):672-681.
35. Smith LF. Development of a multidimensional labour satisfaction questionnaire: dimensions, validity, and internal reliability. *Qual Health Care.* 2001;10(1):17-22.
36. Smith LF. Postnatal care: development of a psychometric multidimensional satisfaction questionnaire (the WOMBPNSQ) to assess women's views. *Br J Gen Pract.* 2011;61(591):e628-637.
37. Bobak M, Pikhart H, Hertzman C, Rose R, Marmot M. Socioeconomic factors, perceived control and self-reported health in Russia. A cross-sectional survey. *Social science & medicine (1982).* 1998;47(2):269-279.
38. Ravelli AC, Eskes M, Tromp M, et al. [Perinatal mortality in The Netherlands 2000-2006; risk factors and risk selection]. *Nederlands tijdschrift voor geneeskunde.* 2008;152(50):2728-2733.
39. Agyemang C, Denktas S, Bruijnzeels M, Foets M. Validity of the single-item question on self-rated health status in first generation Turkish and Moroccans versus native Dutch in the Netherlands. *Public Health.* 2006;120(6):543-550.
40. Fransen M, Harris VC, Essink-Bot ML. Low health literacy in ethnic minority patients: understandable language is the beginning of good healthcare. *Nederlands Tijdschrijft voor Geneeskunde.* 2013;157(14):A5581.
41. Engelhard D. *Met kennis van feiten. Vluchtelingen, nieuwkomers en gezondheid in cijfers [With knowledge of facts: figures about refugees, migrants and health].* Utrecht: Pharos;2007.
42. Mann S, Pratt S, Gluck P, et al. Assessing quality obstetrical care: development of standardized measures. *Joint Commission journal on quality and patient safety.* 2006;32(9):497-505.

APPENDIX

Table A1. The number of participants in the focus groups for each organization.

	Maternity care organization 1*	Maternity care organization 2*	Maternity care organization 3*
Group interview – antepartum	5	4	0
Group interview – postpartum	2	2	3
Individual interview – ante/postpartum	4	0	1
Experts – ante/postpartum	8	7	6

* Description of the participating maternity care organizations:
1) an integrated midwifery practice and clinic from the university hospital (Utrecht)
2) a clinic from the university hospital, and a postnatal birth centre (Rotterdam)
3) an integrated midwifery practice and clinic from a peripheral hospital (Roosendaal)

3

Applicability of the ReproQ Client Experiences Questionnaire for Quality Improvement in Maternity Care

M. Scheerhagen, H.F. van Stel, D.J.C. Tholhuijsen, E. Birnie, A. Franx, G.J. Bonsel

PeerJ. 2016: 4; e2092.

ABSTRACT

Background. The ReproQuestionnaire measures the client's experience with maternity care, following the WHO responsiveness model. In 2015, the ReproQ was appointed as national client experience questionnaire and will be added to the national list of indicators in maternity care. For using the ReproQ in quality improvement, the questionnaire should be able to identify best and worst practices. To achieve this, ReproQ should be reliable and able to identify relevant differences.

Methods and findings. We sent questionnaires to 17,867 women 6 weeks after labor (response 32%). Additionally we invited 915 women for the retest (response 29%). Next we determined the test-retest reliability, the Minimally Important Difference (MID) and six known group comparisons, using two scorings methods: the percentage women with at least one negative experience and the mean score. The reliability for the percentage negative experience and mean score was both 'good' (Absolute agreement=79%; intraclass correlation coefficient=0.78). The MID was 11% for the percentage negative and 0.15 for the mean score. Application of the MIDs revealed relevant differences in women's experience with regard to professional continuity, setting continuity and having travel time.

Conclusions. The measurement characteristics of the ReproQ support its use in quality improvement cycle. Test-retuest reliability was good, and the observed minimal important difference allows for discrimination of good and poor performers, also at the level of specific features of performance.

INTRODUCTION

Client experiences are considered to be important independent indicators for health care performance[1,2]. Being relevant for its own sake, client experiences also affect clinical outcome through several pathways[3-6]. For example, clients who truly understand the explanation of their caregiver are more likely to comply to treatment or to change lifestyle, and arguably a patient-unfriendly clinical staff or an intimidating hospital-setting will not support recovery.

The routine measurement and use of client experiences play an indispensable role in systematic quality improvement[7,8]. For that purpose, the client information can be used in a two-stage quality cycle. In the first stage, care providers that perform above or below average are identified. This process is also called benchmarking[9-12]. In the second stage assumed underperformers are invited to improve their results followed by an internal quality cycle, where above-average performers ('best practices') may give guidance. Translated technically, the quality cycle starts with quantification of individual client experiences and clinical outcomes (casemix-adjusted), followed by the ranking across providers. Next, after defining thresholds, under- and best-performing units are defined. Finally, client experiences and other outcomes are analyzed in more detail. Preferably this break down of data is combined with face-to-face interactions among professionals. This more refined analysis offers tangible targets for improvement, unlike the global outcomes used in benchmarking.

To include clients' experiences with maternity care in the routine care quality evaluation and quality improvement, and in view of professionals, clinical organizations and health insurance companies who urged for measuring quality from the perspective of the client, we developed the Repro Questionnaire (ReproQ)[13]. This integral measure covers the period from first antenatal intake up to the postpartum period. The ReproQ consists of 8 domains (33 experiences items), following the so-called WHO Responsiveness model[1,2]. All items strictly focus on service delivery from the clients' perspective.

Previously we demonstrated the feasibility, internal consistency and construct validity of the ReproQ[13]. In current paper we focus on psychometric properties needed to assess the suitability of the ReproQ for the two-stage quality improvement process. This suitability rests on two pillars: 1) are the judgments of pregnant women reliable, or stated otherwise, if the client survey is repeated do we get the same average and the same ranking of units? And 2) if we observe a quantitative difference between the average judgments of two care units – say 0.2 ReproQ points in our case – is this difference a relevant one? Epidemiologists developed a robust method to decide which differences are relevant in case of difficult to grasp clinical outcomes, the so-called minimal important difference (MID) approach. We tested these properties of the ReproQ to establish whether the ReproQ is suitable for both global benchmarking with a summary score (hence the assignment of below- and above-average performance), and for detailed profiling of providers or units or client groups once the underperforming units of client groups have been identified using the MID.

The data presented were collected during the provisional implementation of ReproQ measurement in about 1/3 of all perinatal units (hospitals with nearby midwife practices) in the Netherlands between October 2013 and January 2015. In 2015, the ReproQ was appointed as a national client experience questionnaire and will be added to the national list of indicators in maternity care[14]. Before the ReproQ was added to this national list, indicators only measured clinical outcomes (e.g. mortality, morbidity or complications) or parameters of professional performance. Adding the ReproQ to this list of indicators meets the WHO's recommendation to measure performance of health care systems also from the client's perspective. As indicator of performance the ReproQ should meet the conditions for a successful quality improvement cycle. This study explores two of these conditions: the ReproQ's reliability of the performance measurements and the MID as an aid to identify relevant differences between clients or perinatal units. The focus in this paper is on client's experiences with labor, because this is the key-event in maternity care. Antenatal care aims to create the best possible situation or starting point for labor. Antenatal risk assessment will be performed, and if necessary preventive measures and treatment of these risks are embedded. Provision of postnatal care is provided to reflect the outcome of the delivery for mother and child. Additionally, care during delivery is comparable in most Western countries, while antenatal and postnatal care are subject to more variation across countries or health systems.

MATERIALS & METHODS

Repro Questionnaire

The questionnaire consists of two analogous versions: version A covers the experiences during pregnancy (antenatal) and version B covers the experiences during birth and the postnatal period. Version A is presented at about eight months gestational age, version B about six weeks postpartum. Each version asks for experiences at two instances, in case of version B postpartum experiences during labor, and experiences in the subsequent postpartum week respectively. As questions only differ with respect to the context referred to (say, experienced respect is asked for first antenatal visits, late in pregnancy, during labor, and during post-partum care), the resulting dataset represents a similar measurement covering four time intervals. In this article we focus on data from version B on the experiences during labor, the 3rd time point.

The 8-domain WHO responsiveness model is the conceptual basis of the ReproQ. Responsiveness is the way a client is treated by the professional and the environment in which the client is treated. Responsiveness is operationalized as four domains represent interactions with health professionals (Dignity, Autonomy, Confidentiality, and Communication), and four domains that reflect experiences with the organizational setting (Prompt attention, Access to family and Community support, Quality of Basic amenities, and

Choice and continuity of care) (see Table 1)[1,2]. The response mode of the experience items uniformly used four categories: "never", "sometimes", "often", and "always", with a numerical range of 1 (worst) to 4 (best). An additional question which two domains are considered the most important, allows for a personalized scoring. Additional questions provide information on 1) the rating of the global experience, 2) the process of care process, the location of care (e.g. home or hospital) and the primary health professional being responsible (e.g. midwife or obstetrician); 3) the clinical outcome of both mother and child, as perceived by the mother; 4) information about previous pregnancies; and 5) client's socio-demographic characteristics.

Content validity of the ReproQ-version-0 was determined through structured interviews with pregnant women, women who recently had given birth, and health care professionals. All Responsiveness domains were judged relevant. Construct validity of the adapted ReproQ-version-1 was determined through a web-based survey, and based on response patterns; exploratory factor analysis; association of the overall score with a Visual Analogue Scale; and known group comparisons. The exploratory factor analysis supported the assumed domain structure and suggested several adaptations. Correlation of the VAS rating and overall ReproQ score supported validity for the antenatal and postnatal versions of the ReproQ. Further details are described elsewhere[13].

Table 1. Description of the eight WHO Responsiveness domains.

Domain	Description
Dignity	Receiving care in a respectful, caring, non-discriminatory setting.
Autonomy	The need to involve the individuals in the decision-making process to the extent that they wish this to occur; the right of patients of sound mind to refuse treatment for themselves.
Confidentiality	The privacy of the environment in which consultations are conducted by health providers; the confidentiality of medical records and information about individuals.
Communication	The notion that providers explain clearly to the patient and family The nature of the illness, and details for the required treatment and options. It also includes providing time for patients to understand their symptoms and to ask questions
Prompt attention	Care provided readily or as soon as necessary
Social considerations	The feeling of being cared for and loved, valued, esteemed and able to count on others should the need arise.
Basic amenities	The extent to which the physical infrastructure of a health facility is welcoming and pleasant
Choice and continuity	The power or opportunity to select, which requires more than one option.

Data collection

In current study, data were obtained from three sources. The majority of data were collected by three postnatal care organizations (organizations that deliver postnatal care over a period of seven to 10 days). Additional data were collected by the national Birth Centre Study (a university-based research organization), and from 10 perinatal units (a hospital with associated midwifery practices). There were no exclusion criteria regarding organization, health care professional or client.

Data collection implied that clients were invited to participate by their care provider on behalf of the research team. With their consent, name and e-mail address were obtained and provided to the organization that distributed the digital survey. Women provided formal informed consent at the beginning of the questionnaire. For the Birth Centre Study and 10 perinatal units, the research team received client's name and e-mail information for recruitment after written informed consent had been obtained. The person who included the woman can, theoretically, be the same as the health care professional in charge of the delivery (usually gynecologist or community midwife), but this is highly unlikely to be the case and not typical of our obstetric care system.

During data collection, an extensive data privacy protocol applied. The Medical Ethical Review Board of the Erasmus Medical Center, Rotterdam, the Netherlands, approved the study protocol (study number MEC-2013-455).

Data were collected in two waves. The first wave was between October 2013 and January 2015. Six weeks after the expected date of labor, all participating women received an invitation to fill out the postnatal ReproQ questionnaire. Non-responding women received a reminder two weeks later. These data were used to determine the MID and compare the known groups. The second wave occurred during October 2014 and January 2015. All women who previously filled out the postnatal ReproQ measurement in the first wave were invited to fill out their experiences again for the test-retest comparison. Excluded from invitation were women whose answers in the postnatal ReproQ were largely incomplete. The intended test-retest interval was 14 days. Since women's situation might change during the test-retest interval, we added the following item for verification. "*Have you experienced something important in the last two weeks?*"

Participating women

Sample size was not formally calculated since we had no prior data to use as input data. Additionally, a formal sample size calculation seems questionable since statistical testing does not a play role in the estimation of the MID. Moreover, we anticipated that the provisional national implementation of this survey would provide sufficient numbers of responses for the study questions. For the MID and known groups comparison, we included all usable responses. For the test-retest, we aimed at 200 usable questionnaires.

In the first wave, we invited 17,867 women who recently had given birth, of whom 5,760 responded to the survey (32%). We excluded 877 women, because they filled out less than two of the following characteristics: ethnicity, educational level, care process, and experienced outcome of the mother and baby. We considered these background data as critical to describe the study participants in sufficient detail, and to understand and interpret the ReproQ scores and the associated MIDs. In the second wave, we invited 915 women for the retest, of whom 265 responded (29%). We excluded 57 women for the retest, because their situation changed negatively or was unknown. We did so because, a test-retest analysis requires that context and conditions between the test and retest situations remain unaltered[15]. To judge representativeness, we compared the characteristics of 208 women in the test-retest with the 4675 women who filled out the test once using standard Chi square tests.

ReproQ Score Model

We used two scoring models to summarize women's experiences: the proportion women with negative experience(s) (in short: 'percentage negative') and the mean score. Both were calculated for the eight individual domains, the four personal domains, the four setting domains and a total score across all domains. Percentage negative was defined as filling out the response category 'never' in at least one of the domains and/or filling out 'sometimes' in a domain that the client identified as most important. The percentage negative method avoids compensation of a negative experience by positive experiences on other items of domains, whereas the mean scores allow the compensation of negative experiences. The mean scores were computed as unweighted average-scores, treating never (1), sometimes (2), most of the time (3) and always (4) numerically.

Data analysis

Test-retest reliability

Test-retest reliability was assessed using three measures. 1) For the percentage negative, we used the percentage absolute agreement, classified as excellent' (90% – 100%), 'good' (75% – 89%), 'moderate' (60% – 74%), or 'poor' (<60%) (Singh et al., 2011). 2) For the mean scores, we used the Intraclass Correlation Coefficient (two way mixed model, absolute agreement, single average), classified as: 'excellent' (≥.81), 'good' (.61 – .80), 'moderate' (.41 – .60), 'poor' (≤.40)[16]. 3) Finally we created the Bland-Altman plot, calculating the bias (or mean difference between test and retest scores) and the limits of agreement, equal to the mean difference ±2*SD of that mean difference[17,18].

Minimally Important Difference

We determined the MID using 1) the anchor-based (or the difference in score between two adjacent levels of an anchor-question[19]) and 2) distribution-based method (or the difference in distribution of observed scores[20]), each having their merits.

As anchor-question we used the global rating of a client's experience: "*Overall, how would you rate the care received during your labor and care after birth?"* (in short: 'Global rating'). This anchor-question emerged as best option in a review study of the Picker Institute[21]. Women could respond to this question on a 10-point VAS. We determined the mean score and the percentage negative of the individual domains, personal, setting and total scores for the VAS ratings 7, 8 and 9. We used the global rating of '8' as reference category, being the mode in our data[19]. Next, the MID was calculated by subtracting these mean scores of the adjacent categories 7 and 9 from the mean score of the reference, being 8, to check if the differences 7 – 8 and 8 – 9 were equal[19]. The same procedure was used to calculate the MID of the percentage negative. The distribution-based MID was only calculated for the mean score. To determine the MID with distributed-based methods, we calculated the standard error of measurement (SEM)[22], and one half of the standard deviation (½SD)[23,24]. The SEM is estimated by the baseline SD of the measurement multiplied by the square root of 1 minus its reliability coefficient (ICC from the test-retest assessment)[22,25,26]. A difference larger than 1 SEM is thought to indicate a true difference between groups[19,20]. The ½SD margin is regarded as a relevant difference as well[19,20,24].

Clinical known-group comparison

We used six so-called known group comparisons (in terms of clinical outcome) to assess the discriminative validity of the ReproQ. Here we determine if women from different 'known groups' also have different mean experience scores and percentage negative (setting, personal, overall), and if these differences exceed the anchor-based MIDs for 7 – 8 and 8 – 9.

We made the following 'known groups': First, we compared the scores of women who before the labor did and did not meet the health care professional who supervised their labor, this being a proxy of professional continuity[27]. Second, for setting continuity, we compared the scores of women who were entirely low risk versus women who shifted from low-risk to high-risk during parturition. These women have the highest mortality and morbidity risk[28,29]. Third, we compared the scores of women who started their labor in office hours (8:00 am – 5:00 pm, Mondays to Fridays) versus past office hours[30-33]. Fourth, we compared the scores of women who had to travel 15 minutes or more with women who had to travel less than 15 minutes. In agreement with literature, we only included women in this comparison who were transferred from home to hospital during parturition and whose birth was unplanned[34,35]. Fifth, we compared the scores of women who had an emergency with women who had a planned caesarean section[36]. Finally, as proxy of concentration of care, we compared the scores of women who delivered in small hospitals (<750 labors annually (first quartile) vs. large hospitals (≥1500 labors annually (fourth quartile)[37-41].

RESULTS

Table 2 presents the characteristics of responding women who filled out the test (n=4675) and women who filled out the retest (n=208). Mean age was 31 years (SD=4.3); 398 (8%) women were of non-Western background; and 368 (8%) women reported to have a low educational level (both percentages slightly below national average). About half of the women gave birth for the first time (52%; about national average), and 2313 (48%) women did not know the health care professional who supervised labor. 527 (11%) women were referred to secondary care during their pregnancy; 1724 (36%) were referred during parturition (about the national average) and 618 (12%) women had a cesarean section (below the national average of 18%). The characteristics of women who filled out the retest differed significantly in terms of ethnic background (more Western women), setting continuity (more women were referred to secondary care during pregnancy), and global rating (women gave a higher global rating).

Test-retest reliability

Table 3 shows the test-retest reliability. All experience items combined, 47% of the women reported one or more negative experiences filling out the test. When filling out the retest, 40% of women reported one or more negative experiences. The absolute test-retest agreement of 'having a negative experience' was 78.8% (CI: 72.6% – 84.2%). The ICC of the total scores ($mean_{test}$=3.79; $mean_{retest}$=3.78) was 0.78, showing good reliability. The mean test-retest difference of the total score was 0.01; limits of agreement were +0.31 and -0.31). The reliability of the personal and setting scores was similar to the reliability of the total score.

The level of agreement regarding negative experiences within individual domains was excellent, except for the domains Autonomy and for Choice and continuity that showed good agreement. In these two domains, women also reported a higher level of negative experiences (Autonomy: 27.9%; Choice and continuity: 18.3%) than in other domains (<7%). The ICCs varied between moderate (0.49 for Confidentiality) and good (0.70 for Communication). The bias was minimal (≤0.05) and was highest in the domains Dignity and Basic amenities.

Minimally Important Difference

Table 4 shows the MID results, using the two scoring models, including the results for the 7 – 8 and 8 – 9 differences. Using the percentage negative experience, the MID was 11.0%, based on the difference between the global ratings of 7 and 8. This means that the respondents rating their overall experience with the global rating scale with 7 showed 11% more cases of negative experiences compared to the respondents with the rating 8. When comparing the rating of 8 with 9, the MID was 9.2%. If we focus on the personal score, the MID using the 7 – 8 difference was 8.5%, which was comparable to the MID of 8.9% using the 8 – 9 difference. For the setting score, the MID 7 – 8 was 5.4%, which was smaller than

the MID 8 – 9 (6.2%). The MIDs of the individual domains were all ≤8%. Using the ReproQ overall mean instead of the percentage negative experiences, the anchor-based MID based on the 7 – 8 distance was 0.15; when based on the 8-9 rating difference the MID was 0.10. The mean-MIDs of the personal score were slightly larger than the mean-MIDs of the setting score, and the domain MIDs showed some heterogeneity; both patterns were also observed using MIDs for negative experiences.

The use of the mean score also allowed the computation of a distribution-based MID. The distribution-based mean-MIDs of the 7 – 8 differences of the personal, setting and total score were similar to the anchor-based MIDs. In case of the individual domains, all distribution-based mean-MIDs were a somewhat larger than the anchor-based mean-MIDs.

Table 2. Characteristics of women who filled out the test (n=4675) and the retest (n=208)[$].

	Test (%)	Retest (%)
Socio-demographics		
Age		
≤24 years	6	4
25 – 29 years	30	31
30 – 34 years	42	48
≥35 years	22	17
Parity		
Primiparous	52	48
Ethnic background*		
Non-Western	9	3
Educational level		
Low	8	5
Middle	35	33
High	57	61
Marital status		
Married/living together	96	97
Relationship, not living together	2	2
No relationship	2	2
Care		
Professional continuity		
No	48	51
Setting continuity		
Primary care only	37	34
Secondary care only	16	15
Referral to secondary care during pregnancy	11	17
Referral to secondary care during parturition	36	34

Table 2. *Continued*

	Test (%)	Retest (%)
Onset of delivery		
Outside office hours	70	64
Travel time*		
None or by choice	70	79
<15 minutes during delivery	18	14
≥15 minutes during delivery	12	7
Cesarean section		
No	87	87
Planned cesarean	4	3
Emergency cesarean	8	10
Hospital size of the perinatal unit		
<750 deliveries per year	12	11
750 – 1499 deliveries per year	47	44
≥1500 deliveries per year	40	46
Quality		
Picker overall rating*		
≤6	8	4
7	16	8
8	34	39
9	26	30
10	16	18

[5] The percentage of missing data was below 5% in all characteristics, and will therefore not be presented.
* Significant difference between the participating women of the test and women participating the retest.

Clinical known-group comparison

Figure 1A shows the impact of six known groups with an assumed influence on client experiences, using the percentage of negative experiences as scoring model. Two out of six comparisons showed differences in agreement with expectations. Already knowing the professional who supervised labor (i.e. continuity of professional), had a considerable impact: the differences in total score and personal score of women who knew and did not know their professional were larger than the associated MIDs (7 – 8 difference). Similarly, referral during labor (i.e. discontinuity of setting) was associated with differences in total, personal and setting scores larger than the MID.

Table 3. Test-retest reliability of the experience during labor, on percentage women with a negative experience and mean score (n=208).

Score	Negative experience*			Mean experience				
	Test (%)	Retest (%)	Absolute agreement (%)	Test mean (SD)	Retest mean (SD)	ICC	Bias	Limits of agreement^
Total score	**46.6%**	**39.9%**	**78.8%**	**3.79 (0.21)**	**3.78 (0.23)**	**0.78**	**0.01**	**0.31**
Personal domains	**31.7%**	**27.4%**	**82.2%**	**3.75 (0.27)**	**3.74 (0.28)**	**0.74**	**0.02**	**0.42**
Setting domains	**25.0%**	**22.6%**	**83.2%**	**3.82 (0.21)**	**3.83 (0.22)**	**0.74**	**-0.01**	**0.31**
Dignity	3.4%	2.9%	94.7%	3.89 (0.23)	3.84 (0.27)	0.62	0.05	0.48
Autonomy	27.9%	26.0%	86.5%	3.46 (0.59)	3.50 (0.50)	0.65	-0.04	0.88
Confidentiality	1.9%	1.4%	96.6%	3.84 (0.36)	3.81 (0.35)	0.49	0.02	0.74
Communication	2.4%	1.4%	98.1%	3.81 (0.34)	3.79 (0.34)	0.70	0.02	0.55
Prompt attention	6.7%	4.8%	94.2%	3.81 (0.30)	3.82 (0.29)	0.64	-0.01	0.49
Social considerations	2.4%	1.9%	98.6%	3.89 (0.28)	3.90 (0.24)	0.54	-0.01	0.49
Basic amenities	1.4%	1.0%	99.5%	3.89 (0.25)	3.89 (0.27)	0.58	0.05	0.48
Choice and continuity	18.3%	18.3%	85.6%	3.69 (0.44)	3.69 (0.45)	0.62	0.00	0.78

* Most negative experience (never) in a domain and/or 'sometimes' in the individually chosen 2 most important domains.

^ The Bland Altman plot of the total score are presented in supplementary file 2.

Table 4. Minimally important difference of the experience during labor based on the mean scores and the percentage women that had a negative experience (n=3841)[$].

Global rating per score	N	Negative experience*		Mean experience				
		% Neg	Anchor based MID	Mean	Anchor based MID	SD	Distribution-based MID	
							SEM	½ SD
Total score								
7 (→8)	584	60.4%	11.0%	3.59	0.15			
8 (ref)	1,322	49.4%		3.74		0.29	0.14	0.14
9 (←8)	1,021	40.2%	9.2%	3.84	0.10			
Personal score								
7 (→8)	584	44.9%	8.5%	3.52	0.17			
8 (ref)	1,322	36.4%		3.69		0.35	0.18	0.17
9 (←8)	1,021	27.5%	8.9%	3.80	0.11			
Setting score								
7 (→8)	584	36.6%	11.6%	3.66	0.13			
8 (ref)	1,322	25.0%		3.79		0.28	0.14	0.14
9 (←8)	1,021	18.8%	6.2%	3.87	0.08			
Dignity								
7 (→8)	584	11.6%	6.9%	3.66	0.18			
8 (ref)	1,322	4.7%		3.84		0.34	0.21	0.17
9 (←8)	1,021	2.0%	2.7%	3.93	0.09			
Autonomy								
7 (→8)	584	36.0%	5.4%	3.22	0.17			
8 (ref)	1,322	30.6%		3.39		0.58	0.35	0.29
9 (←8)	1,021	25.6%	5.0%	3.56	0.17			
Confidentiality								
7 (→8)	584	5.7%	2.0%	3.64	0.17			
8 (ref)	1,322	3.7%		3.80		0.46	0.33	0.23
9 (←8)	1,021	1.8%	1.9%	3.88	0.08			
Communication								
7 (→8)	584	5.3%	2.4%	3.55	0.17			
8 (ref)	1,322	2.9%		3.73		0.40	0.22	0.20
9 (←8)	1,021	0.7%	2.2%	3.84	0.11			
Prompt attention								
7 (→8)	584	10.3%	4.4%	3.62	0.13			
8 (ref)	1,322	5.9%		3.75		0.35	0.35	0.18
9 (←8)	1,021	3.9%	2.0%	3.85	0.10			

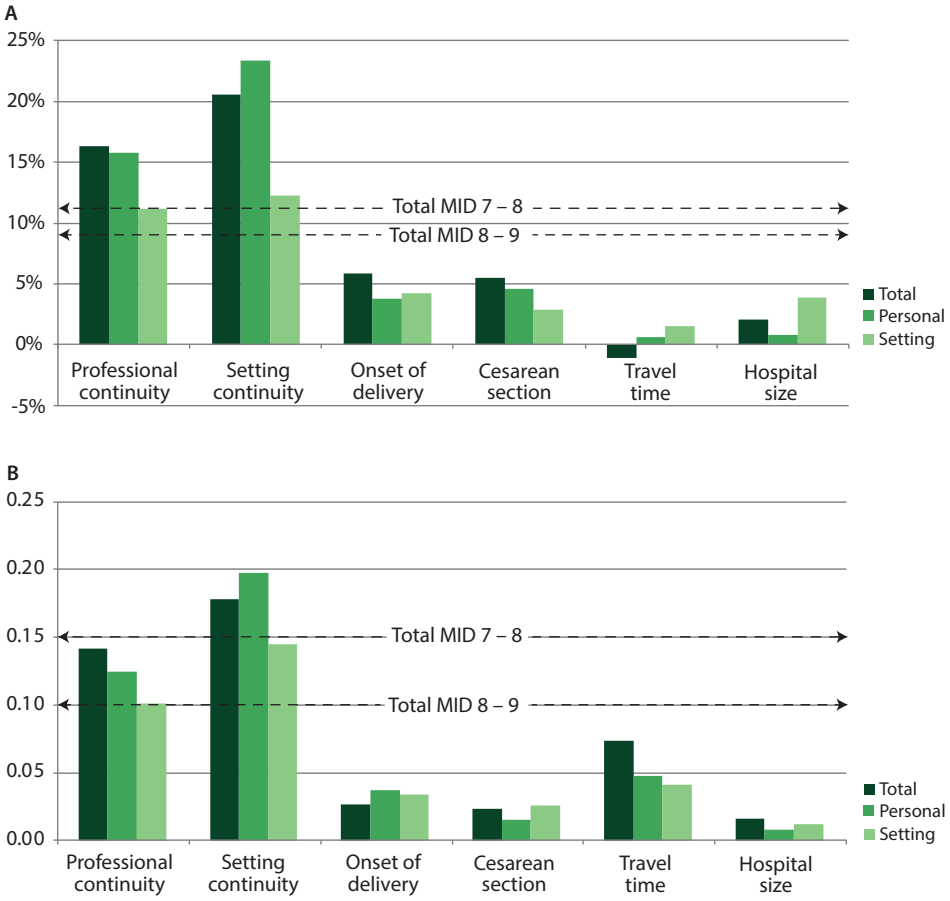
Table 4. *Continued*

Global rating per score	N	Negative experience*		Mean experience				
		% Neg	Anchor based MID	Mean	Anchor based MID	SD	Distribution-based MID	
							SEM	½ SD
Social considerations								
7 (→8)	584	3.3%	1.0%	3.75	0.12			
8 (ref)	1,322	2.3%		3.87		0.33	0.22	0.16
9 (←8)	1,021	1.5%	0.8%	3.91	0.04			
Basic amenities								
7 (→8)	584	4.6%	2.5%	3.80	0.07			
8 (ref)	1,322	2.1%		3.87		0.30	0.20	0.15
9 (←8)	1,021	1.6%	0.5%	3.92	0.05			
Choice and continuity								
7 (→8)	584	26.2%	8.0%	3.46	0.19			
8 (ref)	1,322	18.2%		3.65		0.52	0.32	0.26
9 (←8)	1,021	13.3%	4.9%	3.78	0.13			

$ Due to a software problem this item was not presented to 20% of the participating women.
* Most negative experience (never) in an domain and/or 'sometimes' in the individually chosen 2 most important domains.

Figure 1B shows the same known groups comparison, now using the mean ReproQ scores and the associated MID. The difference in mean overall, setting and personal scores between women who received only primary care and women who were transferred during parturition was larger than the corresponding MIDs (7 – 8 difference). All three differences scores of personal continuity and setting continuity were larger than the MIDs (8 – 9 difference). Further details are presented in supplementary file 1.

Figure 1. Difference in ReproQ in terms of negative score (A) and mean score (B), between the least preferred and the most preferred state, in 6 known-groups (n=4883).



Professional continuity – difference supervisor of the delivery is known vs. unknown (52% / 48%)
Setting continuity – difference primary care only vs. referred during labor (37% / 36%)
Cesarean section – difference planned vs. emergency cesarean section (4% / 8%)
Onset of delivery – difference in vs. outside office hours (30% / 70%)
Travel time – difference women had to travel < 15 minutes vs. ≥ 15 minutes, when transferred from home to hospital during labor (17% / 11%).
Hospital size – difference <750 deliveries per year vs. ≥1500 deliveries per year (3% / 12%).

DISCUSSION

To determine the suitability of ReproQ in the two-stage quality improvement cycle, we assessed its test-retest reliability and determined the MID according to two methods. Test-retest reliability was good for both scoring models. The anchor-based MID of the percentage negative experiences was 11%; the anchor-based MID of the mean score was 0.15 (on a range of 1 – 4). The distribution-based MIDs (SEM) proved about similar to the anchor-based mean-MID of the overall, personal and setting scores. However, for the domain scores the SEM exceeded the anchor-based mean-MIDs. The known-group comparisons showed that knowing the professional that supervised your labor and not being referred during labor had considerable impact on the experiences scores. As the observed ReproQ scores deviated more than the MID, we believe this instrument can be used as a benchmark with an interpretation of meaningful differences beyond statistical significance. Thus, the ReproQ can successfully identify areas that need improvement in subgroups of clients. One should be aware that the MID cannot be used to identify changes in (poor) experiences within clients.

Applying the percentage negative on the test-retest reliability showed that the reliability of the domains was higher than for the summery scores. This is surprising, because the likelihood to report a negative experience in both the test and retest is considerably larger for the summary scores than for the domains.

For the individual domains, fewer women reported a negative experience when filling out the retest than the test. The domains Autonomy and Choice and continuity showed similar percentages of negative experiences in the test and retest, though the reliability of these domains was low compared to the other domains. This indicates that women who reported a negative experience filling out the test are not the same women that reported a negative experience filling out the retest. Possible explanations for these effects are recall bias and/or response shifts, e.g. women adjust their opinion due to sharing their experiences with family and friends.

The summary scores showed higher reliability than the domain scores when using the mean score method. The explanation is that ICCs invariably increase when the summary scores include more items. When calculating a summary score, differences within a domain can be compensated by differences between the domains. This increases the stability of the summery scores.

The reliability of the domains Confidentiality and Social considerations was somewhat lower than for the other domains. It is possibly that women feel that these concepts are difficult to judge, which increases the fluctuations in domain scores.

Both the negative-MID and the mean-MID varied across adjacent response categories. When the global rating increases neither the percentage negative decreases nor the mean scores increase linear in all scores. This suggests that a gain in client experience as the result of quality improvement is not similar to the loss in client experience as the result of

deterioration. One explanation is that clients do not weigh all domains are equally in the global rating. Another explanation is that respondents are not inclined to use the extreme response categories.

The distribution-based MIDs (SEM) were similar to the anchor-based mean-MID of the overall, personal and setting scores. However, for the domain scores the SEM exceeded the anchor-based mean-MIDs, because the SDs of the domain scores were larger than the SDs of the summary scores, and because the domain ICC scores were lower.

The known-group comparisons were based on previously reported differences in clinical outcomes. Professional and setting continuity also resulted in large and relevant differences in experience scores. These differences are probably due to the deviation of the expected or planned process of care, which might result in a stressful event, even when this deviation is clinically necessary. Differences in experiences of the other clinical known-group comparisons were not relevant. It is possible that these experiences are not or only partly correlated with the clinical outcomes. Another explanation might be that the experiences were reported in retrospect. Perhaps women's experiences were biased afterwards by a good maternal or child outcome, by better, sufficient or intensive postnatal care when complications occurred during labor. It is also possible that women's experiences of the process of labor were affected by hormones and stress, or that women lacked information on what normal maternity care is. (Note that about half of the women was primiparous).

Strengths

To our knowledge, this is the first study to clarify the meaning or relevance of score differences obtained with client experience questionnaires. So far, current studies towards the MID mainly focussed on quality of life scores[19,42,43]. Secondly, the use of global ratings is debated due to their unknown validity and reliability[19]. By using the overarching question of the British National Patient Survey Coordination Centre as anchor-question we met this critique: this global rating is extensively tested and has a good content and construct validity[21]. Thirdly, we explored the differences between 7 – 8 and 8 – 9 changes in global rating. By doing so, we were able to check the assumption that both differences were similar. Inevitably, preference scales are to some extent non-linear in interpretation, which applies to both the ReproQ and to scales used for anchoring. At the upper or lower ends of the scale the interpretation of gains and losses may differ, and the 'degree of relevance' of one step higher (7 – 8) or lower (8 – 9) decreases. Since benchmarking is usually based on the comparison of averages, the impact of non-linearity is probably small.

We previously introduced the percentage negative experiences as an alternative scoring to the frequently used mean score. Three remarks should be made. Firstly, we deliberately focused on the percentage negative experiences instead of percentage positive experiences. Focusing on the latter may contribute to the validity of the findings. However, from a practical perspective, we chose to emphasize the percentage negative experiences

because in quality improvement cycles most benefit can be obtained when poor performing providers or centers are identified and improvements can be implemented. The percentage negative experiences seems therefore more relevant for quality improvement than the percentage positive experiences. We expect that the benefit of quality improvements for centers with a high percentage of positive experiences is less than the benefit for poor performing centers. Secondly, both the percentage negative experiences as well as the mean score can be used for benchmark purposes. Despite differences in approach, both may lead to the same identification of relevant differences in subgroups (see Figure 1). Finally, one could argue that our approach of the MID is conservative, as it actually defines the size of a relevant minimal difference between averages on the group level on the base of differences in individual global ratings.

Limitations

First, we sent the postnatal questionnaire six weeks after the expected date of labor, but it is unknown if this timing was optimal. An invitation later than six weeks could result in recall bias due to exposure to other influences (e.g. women return to work, assuming their usual habits and patterns), and/or in non-response because sharing one's birth experiences may seem less relevant. An invitation before six weeks is not necessarily a better option. It may result in better recollection of the experiences but the risk of mood swings and hormonal disturbances might affect responses and response rates.

Related to this: the postnatal questionnaire was not sent six weeks after the actual date of labor. We only had the expected date of delivery as anchor. To protect women's privacy, we were not allowed to collect the precise date of childbirth in the ReproQ. Since the expected date may deviate from the true date, women may have been surveyed earlier (but not more than 2 weeks earlier) when they delivered after the expected date, or about 4 to 5 weeks later for most women when they delivered before the expected date. In both cases, postnatal care had already ended, and it is unlikely that differences in timing of the invitation of the ReproQ may have resulted in different ReproQ scores between these groups.

Secondly, women with a low educational level and non-Western women were underrepresented despite considerable efforts to have them participate. Most likely this is selective non-response, as non-Western women report more negative experiences than Western women[44], and non-Western women are more often low educated and/or more health illiterate[45-47]. Addition of the non-response group is likely to widen the gap between poor and good experiences. This does not necessarily affect the estimated MID. Our non-Western women reported both a lower ReproQ score as well as a lower global rating than Western women. Repeating the MID calculations without this subgroup (non-Western women with a low educational level) resulted in about similar results. When this subgroup was excluded, the percentage negative MID decreases maximum -0.1% and increases max. +1.2%. Similar, the mean MID varies -0.01 to +0.03. Hence, the underrepresentation of

these subgroups has limited impact on the estimated MID. Regrettably, we could not find additional evidence on the influence of selection bias on psychometric properties of other client experience surveys with similar characteristics in terms of study population, length and mode of administration (e.g. read out loud by clinicians vs. stand alone, self-report)[25,26]. The impact of care process, birth outcome and socio-demographics on experiences scores, test-retest reliability and MID requires further study.

Thirdly, the MID is often used to identify changes in a patient's situation over time[19,42,43]. Given the small time window of the labor phase, it is unfeasible to validly assess changes in survey-based experiences within clients. Therefore, our MID estimates are based on cross-sectional comparisons. Our MID cannot be used to identify changes within a client, but only between health care providers, or within health care providers over time. These provider differences are more relevant than changes within clients for improving the quality of maternity care by the two-stage quality cycle.

Finally, we aimed at suitability of the ReproQ survey across countries, by using the universal WHO Responsiveness concept, by following an accepted strategy for survey development, and by avoiding any preferences towards providers, specific professionals or organizational structures. It is unclear if clients in other countries have the same importance ratings, experiences and MIDs as Dutch clients. Other self-report instruments in maternity care, such as the Women's Experience of Maternity Care Questionnaire of the NHS, overall indicate very good experiences[48-51]. Therefore, the MID in other developed countries will probably be in about same magnitude as our MID estimates.

Future use

The psychometrics of the ReproQ appear adequate for benchmarking for targeting quality improvement based on the profile of domain scores, and for monitoring of domain specific quality improvements. As part of a routine two-stage quality improvement cycle, as proposed by the ICHOM[52], we can identify relevant differences between birth care units who perform better or worse. The MID based percentage negative discriminates (known) groups better than the mean-MID. Furthermore, we recommend to use a multi-item questionnaire for benchmarking, such as the ReproQ, instead of a single-item benchmark: the reliability of a single-item benchmark is much lower and, unlike the ReproQ, single items are less effective in guiding specific improvements.

To increase the response rate, alternative modes of data collection should be explored. One suggestion is to invite women to directly fill out the questionnaire when waiting for their health care professional in the waiting room. Another suggestion to minimalize selection and response bias is to send all women the questionnaire including informed consent, without involvement of individual health care professionals. A third suggestion is to translate and provide the questionnaire in several languages for non-Western women.

Additionally, future use should pay attention to ethnicity and socio-economic background, beyond routine case-mix adjustment procedures. Adjustment always bears the risk that unintentionally worse experiences are neutralized, taking away the incentive for improvement.

With many benchmarking activities into place, the second part of the quality cycle urgently needs more attention and explicit implementation. Evidence-based routine quality cycles are still rare. Implementation requires true information-guided cycles in some detail. The benefit of such an approach has been demonstrated in the evaluation of innovations[7,8]. The introduction of the MIDs in quality cycles may convince stakeholders that progress through innovation is meaningful.

CONCLUSION

Maternity care is continuously developing, partly based on the measurement of client experiences. The ReproQ questionnaire, based on the WHO Responsiveness model, is suitable to be used in quality improvement cycles: we showed good test-retest reliability, and by determining the minimally important difference relevant differences can be identified.

ACKNOWLEGDEMENT

We are grateful for all maternity care organizations that took part in our study. We are especially grateful to Christa Cats, Hans Reinold and Ester van Dalen, who inspired the postnatal care organizations to participate in our research.

APPENDIX

See the open acces version of this article to access Appendix 1 and 2: https://peerj.com/articles/2092/

REFERENCES

1. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
2. Valentine NB, De Silva A, Kawabata K, Darby C, Murray CJ, Evans BE. Health system responsiveness: concepts, domains and measurement. In: Murray CJ, Evans BE, eds. *Health systems performance assessment*. Geneva: World Health Organization; 2003:573-596.
3. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Soc Sci Med.* 2000;51(11):1611-1625.
4. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997;45(12):1829-1843.
5. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med.* 1998;47(10):1573-1588.
6. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med.* 1994;38(4):509-516.
7. Haugum M, Danielsen K, Iversen HH, Bjertnaes O. The use of data from national and other large-scale user experience surveys in local quality work: a systematic review. *Int J Qual Health Care.* 2014;26(6):592-605.
8. Weinick RM, Quigley DD, Mayer LA, Sellers CD. Use of CAHPS patient experience surveys to assess the impact of health care innovations. *Jt Comm J Qual Patient Saf.* 2014;40(9):418-427.
9. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
10. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
11. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
12. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
13. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring Client Experiences in Maternity Care under Change: Development of a Questionnaire Based on the WHO Responsiveness Model. *PLoS One.* 2015;10(2):e0117031.
14. CPZ. Uniform gebruik cliëntervaringslijst ReproQ in de Geboortezorg [Uniform use client experience list ReproQ in maternity care]. 2015; http://www.goedgeboren.nl/netwerk/h/215/0/3596/rubriek--9/Uniform-gebruik-cli-ntervaringslijst-ReproQ-in-de-Geboortezorg, 2015.
15. De Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine. Cambridge University Press; 2011.
16. Singh AS, Vik FN, Chinapaw MJ, et al. Test-retest reliability and construct validity of the ENERGY-child questionnaire on energy balance-related behaviours and their potential determinants: the ENERGY-project. *Int J Behav Nutr Phys Act.* 2011;8:136.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.
18. Streiner DL, Norman GR. *Health measurement scales.* Oxford: University Press; 2008.
19. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7(5):541-546.
20. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109.
21. Graham C, Maccormick S. *Overarching questions for patient surveys: development report for the Care Quality Commission (CQC).* Picker Institue Europe;2012.
22. Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res.* 2002;11(1):1-7.
23. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2):171-184.

24. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003;41(5):582-592.
25. Rejas J, Ruiz MA, Pardo A, Soto J. Minimally important difference of the Treatment Satisfaction with Medicines Questionnaire (SATMED-Q). *BMC Med Res Methodol.* 2011;11:142.
26. Vernon MK, Revicki DA, Awad AG, et al. Psychometric evaluation of the Medication Satisfaction Questionnaire (MSQ) to assess satisfaction with antipsychotic medication among schizophrenia patients. *Schizophr Res.* 2010;118(1-3):271-278.
27. Saultz JW, Lochner J. Interpersonal continuity of care and care outcomes: a critical review. *Ann Fam Med.* 2005;3(2):159-166.
28. Evers AC, Brouwers HA, Hukkelhoven CW, et al. Perinatal mortality and severe morbidity in low and high risk term pregnancies in the Netherlands: prospective cohort study. *Bmj.* 2010;341:c5639.
29. Poeran J, Borsboom GJ, de Graaf JP, Birnie E, Steegers EA, Bonsel GJ. Population attributable risks of patient, child and organizational risk factors for perinatal mortality in hospital births. *Matern Child Health J.* 2015;19(4):764-775.
30. Gould JB, Qin C, Chavez G. Time of birth and the risk of neonatal death. *Obstet Gynecol.* 2005;106(2):352-358.
31. Gould JB, Qin C, Marks AR, Chavez G. Neonatal mortality in weekend vs weekday births. *Jama.* 2003;289(22):2958-2962.
32. Stephansson O, Dickman PW, Johansson AL, Kieler H, Cnattingius S. Time of birth and risk of intrapartum and early neonatal death. *Epidemiology.* 2003;14(2):218-222.
33. Urato AC, Craigo SD, Chelmow D, O'Brien WF. The association between time of birth and fetal injury resulting in death. *Am J Obstet Gynecol.* 2006;195(6):1521-1526.
34. Poeran J, Borsboom GJ, de Graaf JP, et al. Does centralisation of acute obstetric care reduce intrapartum and first-week mortality? An empirical study of over 1 million births in the Netherlands. *Health Policy.* 2014;117(1):28-38.
35. Ravelli AC, Jager KJ, de Groot MH, et al. Travel time from home to hospital and adverse perinatal outcomes in women at term in the Netherlands. *Bjog.* 2011;118(4):457-465.
36. Elvedi-Gasparovic V, Klepac-Pulanic T, Peter B. Maternal and fetal outcome in elective versus emergency caesarean section in a developing country. *Coll Antropol.* 2006;30(1):113-118.
37. Finnstrom O, Berg G, Norman A, Otterblad Olausson P. Size of delivery unit and neonatal outcome in Sweden. A catchment area analysis. *Acta Obstet Gynecol Scand.* 2006;85(1):63-67.
38. Moster D, Lie RT, Markestad T. Relation between size of delivery unit and neonatal death in low risk deliveries: population based study. *Arch Dis Child Fetal Neonatal Ed.* 1999;80(3):F221-225.
39. Moster D, Lie RT, Markestad T. Neonatal mortality rates in communities with small maternity units compared with those having larger maternity units. *Bjog.* 2001;108(9):904-909.
40. Phibbs CS, Bronstein JM, Buxton E, Phibbs RH. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *Jama.* 1996;276(13):1054-1059.
41. Tracy SK, Sullivan E, Dahlen H, Black D, Wang YA, Tracy MB. Does size matter? A population-based study of birth in lower volume maternity hospitals for low risk women. *Bjog.* 2006;113(1):86-96.
42. Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes.* 2006;4:69.
43. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77(4):371-383.
44. Scheerhagen M, Van Stel HF, Birnie E, Franx A, Bonsel GJ. *Het discriminerend vermogen van de ReproQuestionnaire [the discriminative power of the ReproQuestionnaire].* Rotterdam, The Netherlands: Erasmus Medical Center;2015.

45. Agyemang C, Denktas S, Bruijnzeels M, Foets M. Validity of the single-item question on self-rated health status in first generation Turkish and Moroccans versus native Dutch in the Netherlands. *Public Health.* 2006;120(6):543-550.
46. Fransen M, Harris VC, Essink-Bot ML. [Low health literacy in ethnic minority patients: understandable language is the beginning of good healthcare]
47. Beperkte gezondheidsvaardigheden bij patienten van allochtone herkomst: alleen een tolk inzetten is meestal niet genoeg. *Ned Tijdschr Geneeskd.* 2013;157(14):A5581.
48. Engelhard D. *Met kennis van feiten. Vluchtelingen, nieuwkomers en gezondheid in cijfers [With knowledge of facts: figures about refugees, migrants and health].* Utrecht: Pharos;2007.
49. Peterson WE, Charles C, DiCenso A, Sword W. The Newcastle Satisfaction with Nursing Scales: a valid measure of maternal satisfaction with inpatient postpartum nursing care. *J Adv Nurs.* 2005;52(6):672-681.
50. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* Oxford, United Kingdom: National Perintal Epidemiology Unit;2010.
51. Smith LF. Development of a multidimensional labour satisfaction questionnaire: dimensions, validity, and internal reliability. *Qual Health Care.* 2001;10(1):17-22.
52. Smith LF. Postnatal care: development of a psychometric multidimensional satisfaction questionnaire (the WOMBPNSQ) to assess women's views. *Br J Gen Pract.* 2011;61(591):e628-637.
53. ICHOM. How to measure. 2015; http://www.ichom.org/measure/. Accessed June 26th, 2015.

Part 2

Benchmarking

4

What Determines Women's Birth Experiences? Applications for a Benchmark

M. Scheerhagen, E. Birnie, A. Franx, H.F. van Stel, G.J. Bonsel

Submitted

ABSTRACT

Background. Clients' experiences are an important independent indicator for health care performance and are often measured as part of a quality improvement cycle, a so-called a benchmark. We studied the relevant determinants that can either attribute to the explanation of a low client experience score, or to a fair ranking of organizations.

Methods and results. The antenatal and postnatal experiences of 3,479 women were measured by the ReproQ, a validated 8-domain questionnaire.

Three scoring models were used to summarize women's experiences during childbirth: mean score, below/above the median score and having a negative experience.

To determine which determinants influence women's experiences during birth, we performed multiple linear and logistic regression analysis for all scoring models. The following determinants contributed to lower experiences: a low or middle educational level, this pregnancy not being planned, negative health outcome of mother and child, a low antenatal experience, women who did not experience professional and setting continuity, women who had an intervention during birth and women who were not able to give birth at the excepted place or who had no prior expectations where to give birth. The influence of organizational settings appeared modest.

Discussion. Relevant determinants that should be regarded in a benchmark are a client's socio-demographic characteristics, her antenatal experiences, the care process during pregnancy and birth and the client's reported outcome of mother and child.

INTRODUCTION

The World Health Organization (WHO) reckons the client's experience with health care provision as an important independent indicator for health care performance. According to the WHO, clients' experiences comprise both the way a client is treated by the professional and the environment in which the client is treated during an health care encounter[1,2]. At the individual level, a client's experiences may guide her choice of a health care organization when health outcomes are similar and may affect clinical outcomes in several ways[2-4]. For example, clients who understand the explanation of their caregiver are more likely to comply to treatment or to change lifestyle, which, in turn, could reduce health risks[2].

Clients' experiences are also used as performance indicator of health care organizations. The performance score of a health care organization can be obtained as the aggregated experience scores of their clients. After the aggregated scores are calculated, they can be used to improve health care delivery as part of a two-stage quality cycle. In the first stage, the aggregated scores are used to rank the experiences scores of health care organizations, so health care organizations with outlying performance (or best and worst practices) can be identified. In the second stage, assumed underperformers are invited to improve their results followed by an internal interpretation of the results and improvement of care accordingly. This process is also called benchmarking[5-8].

A fair ranking at the organizational level in stage one requires that case mix correction needs to be considered. With case mix correction the data is adjusted for determinants that are beyond the influence and usually unrelated to the organization, but which influence the outcome (here: clients' experiences) is distributed unequally across health care organizations[9]. When the case mix corrected overall ranking of organizations result in a low ranking due to a low experience score (where a low experience score is undesirable), this may initiate internal quality improvement of the organization. The internal quality improvement usually needs more detailed the information to provide clues for action, as an overall low client experience score, may be the result of all clients rating lower, or some clients rating considerable lower. Detailing the overall score therefore provides a clear focus for internal quality improvement (stage 2).

To our knowledge, this two-stage quality cycle is not yet operational in maternity care, even though measuring clients' experiences during birth is common practice in several countries[10-14]. Applying the two-stage quality cycle could improve the experiences with maternity care nationally, and simultaneously benefit the health outcome of mother and child. For an optimal quality cycle the need for case mix correction should be explored, just as possible explanations for a low experience score on organizational level.

Our aim was to study the impact of several groups of determinants of women's experiences during birth, 1) to identify determinants that need to be considered for case mix correction and 2) to determine if a low organizational score should be assigned to a lower experience score all clients or to a lower score in specific subgroup of clients. We hypothesize

that five groups of determinants are relevant for a client's experiences during childbirth: 1) her socio-demographics, 2) a client's antenatal experiences, 3) the care process during birth including possible interventions, 4) the perceived health outcome of mother and child, and 5) organizational settings, such as protocols, consultation hours, and cooperation between health care organizations. In our view women's socio-demographics (1) and the perceived health outcome (4) are candidates for case mix correction, if they indeed impact the clients' experiences and are not under the influence of the organization. The other three determinant groups provide information to interpret a low experience score.

The performance measure used in this paper was the Repro Questionnaire, a validated client experience measure which was selected (2015) as national performance indicator for the birth care system in the Netherlands[15].

METHODS

ReproQuestionnaire

The Repro Questionnaire (ReproQ) is an instrument, consisting of a 33-item client-experience module, to which modules are added for case mix correction and interpretation. These additional modules address the client's socio-demographic characteristics, details about the birth care process thus far, interventions during birth, and clinical outcomes of mother and child in non-medical terms as perceived by the mother and reported as patient-reported outcome measure (PROM). A relevance item is included on which two, out of the eight domains, are considered most important by the client, to enable optional weighting of responses. This allowed for the computation of personalized scores.

The experience items of the ReproQ cover the antenatal experiences and the experiences during birth and postnatal period, divided in an antenatal and postnatal version of the questionnaire. Both versions are identical, in the sense that the same type of experiences are asked for, but items are adapted to the specific antenatal or postnatal context. In both versions, the client is invited to give a judgement on the experience as described by the item, for two separate reference periods: in the antenatal version the first and second half of pregnancy, in the postnatal version the event of birth and the subsequent postnatal week. Altogether a typical ReproQ item is judged for four consecutive periods, where response is obtained at two moments.

To measure the experiences in a structured way, the WHO responsiveness model was selected as the conceptual basis of the ReproQ. According to the model, clients' experiences in health care consist of eight domains: four domains address the interactions of the client with the health professional(s) involved (Dignity, Autonomy, Confidentiality, and Communication), while the remaining four domains address the client orientation of the organizational setting (Prompt attention, Access to family and Community support, Quality of Basic amenities, and Choice and continuity of care)[1,2]. Additionally, we added a

relevance (weighting) question on which two out of the eight domains were considered most important by the client. This allowed for the computation of personalized scores.

From previous psychometric analysis followed that content and construct validity was adequate, as was the test-retest reliability of the questionnaire. Details of the development of the questionnaire are described elsewhere[16,17]

ReproQ scoring models

Three different scoring models exist to summarize women's experiences. The first is a binary negative experience score that represents the presence/absence of a so-called negative experience (for details see below). The second is a continuous score (1 – 4 scale) that represents the mean item score. The third is a binary median experience score that represents a score below the median. These three models may be applied to an individual domain (domain score), to the summary scores of the four personal and four setting domains (personal and setting score), or to a summary score of all domains (total score).

A 'negative' experience was defined as ticking the category 'never' in at least one of the items of the domains (a very poor experience), and/or filling out 'sometimes' in a domain that the client identified as most important, thereby creating a personalized score. Since the likelihood of a negative experience depends on the number of items, absolute percentages of negatives cannot be compared across domains. The negative score model assumes that on the individual client level, a negative experience cannot be compensated by very good experiences on other items or domains. This is contrary to the mean score where good experiences can compensate negative experiences.

The 'mean score' was defined as the unweighted average score of items within a domain, treating the item response categories numerically. The summary scores are not the mean of all items involved in the domains, but the mean of the mean domain scores involved in that summary measure. For the calculation of summary scores, each domain has the same weight, even if the domains rest on a different numbers of items.

We added a second binary outcome transforming the continuous score into a score above or below the median. This outcome was added because of the skewed distributions of the experience domains and summary scores.

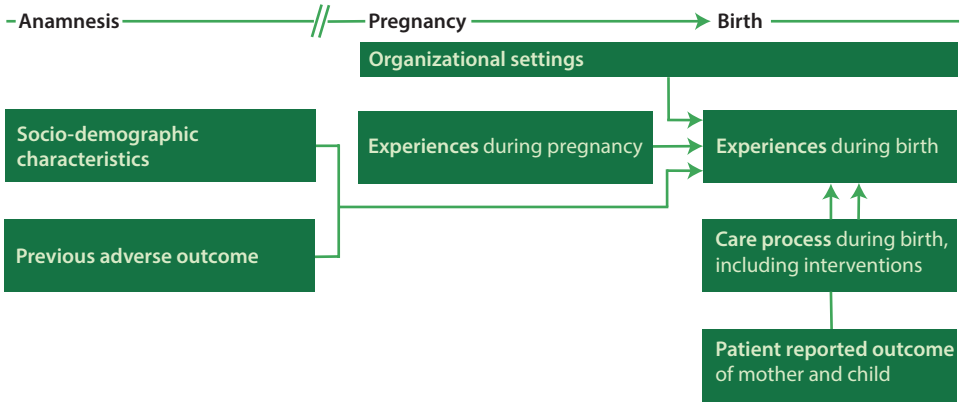
Outcome measure

Although the ReproQ covers all three phases of maternity care (pregnancy, birth and postnatal care), in this paper we focus on clients' experiences during birth, because this is the key-event in maternity care. Additionally, care during birth is comparable in most Western countries, while antenatal and postnatal care are subject to more variation across countries or health systems.

Determinants

The four groups of determinants cover a broad spectrum of individual determinants that can influence the experiences during childbirth (see Figure 1). Below we describe the definitions of determinants that are not self-evident. First, antenatal experiences refer towards the reported experience of the second half of the pregnancy, because in psychometric analysis the experiences during first and second half of pregnancy are highly associated (ICC=0.80). Second, professional continuity was defined as women who before the labor did and did not meet the health care professional who supervised their labor; we assumed this to be a proxy of professional continuity[18]. Third, setting continuity is a variable which assumes that women may give birth in settings with different levels of care depending on her risk status. A change in risk status results in setting discontinuity, as women receive care in different settings and at different locations. The most important setting discontinuity emerges if an assumed low-risk pregnant woman giving birth at home is transferred to the hospital setting[17]. Fourth, 'realization of the expected place of birth' was based on the comparison of the woman's expected location of birth one month before birth (home, hospital) with and the observed location of birth. Fifth, we defined 'a previous adverse outcome' if any of the following had happened with a previous pregnancy: a stillbirth, miscarriage, or abortion. Finally, the health care organizations in our context are so-called perinatal units. A perinatal unit consists of a hospital with associated community midwife practices, being responsible for 800 – 4000 deliveries annually.

Figure 1. Framework of determinants associated with the experiences during birth.



Data collection

Data were obtained as part of a national study on clients' experiences with maternity care in the Netherlands. Eligible for participation were all pregnant women of two maternity care organizations and 10 hospitals and collaborating midwifery practices. There were no exclusion criteria for health care organizations or clients. Neither were any rewards involved.

Clients were invited to participate on behalf of the participating organization. For the maternity care organizations, we received the name, e-mail address and expected date of birth of their clients who agreed to be approached. These women gave informed consent at the beginning of the digital ReproQ. For the 10 hospitals, the research team received contact details and expected date of birth after written informed consent had been obtained.

The antenatal version of the ReproQ was sent at 34th weeks of gestation. The postnatal ReproQ was sent six weeks after the expected date of childbirth.

The Medical Ethical Review Board of the Erasmus Medical Center, Rotterdam, the Netherlands, approved the entire study protocol (study number MEC-2013-455).

Data analysis

No formal sample size was considered as it could be anticipated that the data collection, as provisional national implementation of this survey would provide sufficient numbers of responses for statistical power.

We performed two separate analyses to determine what influences the experiences during childbirth: first we focus on all determinant groups except organizational setting. Next, we only focus on the influence of the organizational setting, which may be assumed a determinant of the aggregated experience score of the health care organization.

First, the impact of the five determinant groups on the total experience score during birth was estimated by regression analyses. For the dependent variables the absence of a negative experience, and a score above the median, we used binary logistic regression analysis. For the continuous score as the dependent variables we applied linear regression analysis. The following sets of determinants were included with an exploratory approach (forced entry): socio-demographic characteristics, the antenatal experiences, care process during pregnancy and birth including interventions during birth, and perceived health outcomes of mother and child (see Figure 2). A determinant was overall judged as significant if its estimated coefficient was statistically significant in at least two of the analyses, as a conservative approach.

For the binary logistic regression analysis, the goodness of fit was assessed using the proportion of correct predictions. For linear regression we used the adjusted R^2. The analyses on the total scores were repeated with the personal and setting score, to verify if the determinants relevant to the total score also apply for the personal and setting score.

Second, we explored the influence of organizational settings by analyzing the experiences during birth (total, personal and setting; for the negative, median and mean scores) separately for the three perinatal units with most respondents. Comparable experience scores across different organizations indicate that the impact of organizational setting on the total experience score is limited.

RESULTS

Client characteristics

We invited 24,752 women to both versions of the ReproQ, of whom 4,055 (16%) responded to both invitations. We excluded 567 responses if less than 2 of the following key characteristics were filled out: ethnicity, educational level, setting continuity, and experienced outcome of the mother and child. A flow diagram of the study is presented in Figure 2.

Figure 2. Flow diagram of study.

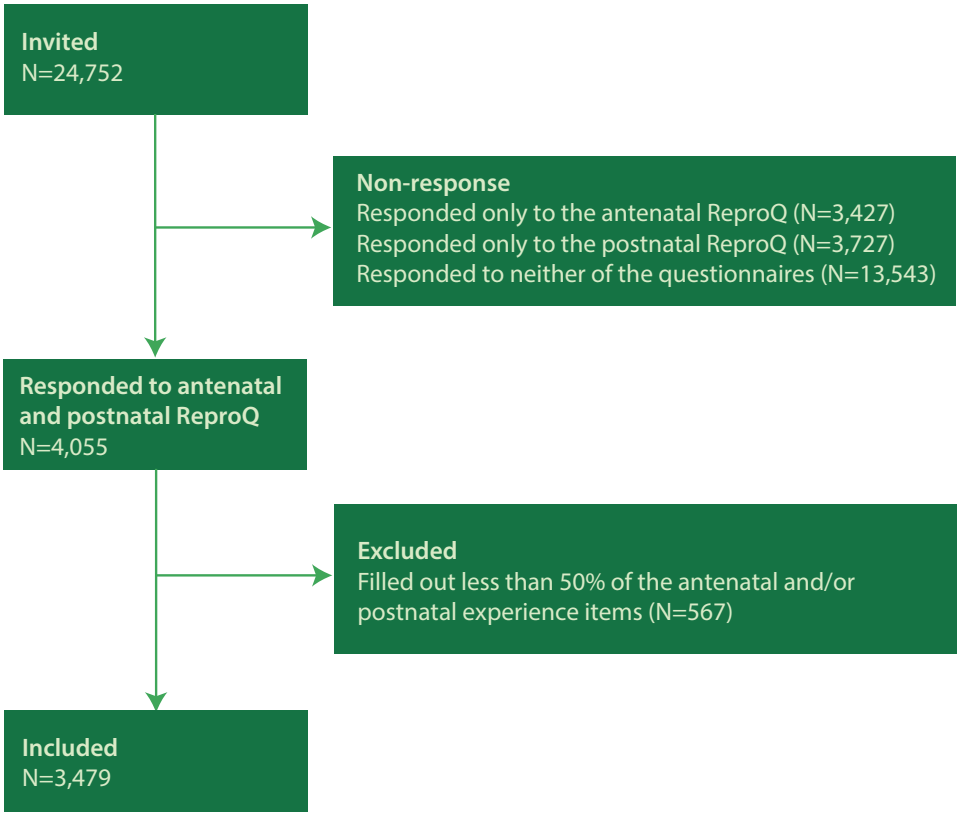


Table 1 presents the characteristics of the included women (n=3,479). Mean age was 31 years (SD=4.2). About half of the women gave birth for the first time (52%). 199 (6%) women were of non-Western background; and (another) 191 (6%) women reported to have a low educational level. 1,739 (51%) women did not know the health care professional who supervised their birth (professional continuity). 473 (14%) women were referred to

Table 1. Characteristics of women who filled out both the antenatal and postnatal ReproQ (n= 3,479)[$].

	N	%	Case mix potential
Socio-demographic characteristics			
Age (y)			
≤24	159	5	Yes
25-29	1,030	30	
30-34	1,487	43	
≥35	778	23	
Parity			
Primiparous	1,815	52	Yes
Multiparous	1,663	48	
Ethnic background			
Western	3,264	94	Yes
Non-Western	199	6	
Educational level			
Low	191	6	Yes
Middle	1,074	31	
High	2,199	64	
Marital status			
Married/living together	3,342	96	Yes
Not living together or no relationship	120	4	
Planned pregnancy			
Yes	3,147	90	Yes
No	331	10	
Care process			
Professional continuity			
Yes	1,698	49	No
No	1,739	51	
Setting continuity			
Primary care entirely	1,167	34	No
Secondary care entirely	593	17	
Primary care at first booking, referral to secondary care during pregnancy	473	14	
Primary care at first booking, referral to secondary care during parturition without urgency	918	27	
Primary care at first booking, referral to secondary care during parturition with urgency	263	8	

Table 1. *Continued*

	N	%	Case mix potential
Expected place of birth was realized			
Yes	2,037	59	No
No	1,314	38	
No prior expectations	86	3	
Day of birth			
Monday to Friday	2,532	73	No
Saturday/Sunday	939	27	
Time of birth			
8 AM to 5 PM	1,470	42	No
5 PM to 8 AM	1,996	58	
Intervention			
Induced labor			
No	2,601	75	No
Yes	865	25	
Intervention			
None/episotomy	2,598	75	No
Vacuum or forceps extraction	366	11	
Planned cesarean	214	6	
Emergency cesarean	298	9	
Pain medication			
No	2,103	61	No
Yes	1,363	39	
Patient reported outcome			
Outcome baby			
Healthy and not hospitalized	2,340	68	Yes
Healthy, but hospitalized	531	15	
Unhealthy, but not hospitalized	137	4	
Unhealthy and hospitalized	455	13	
Outcome mother			
Healthy and not hospitalized	1,940	56	Yes
Healthy, but hospitalized	230	7	
Unhealthy, but not hospitalized	733	21	
Unhealthy and hospitalized	570	16	
Adverse outcome previous pregnancies			
No	2,490	72	Yes
Yes	989	28	

[5] The percentage of missing data was below 3% for all characteristics, and will therefore not be presented.

secondary care during their pregnancy, and 1181 (35%) were referred during parturition (setting continuity). 512 (15%) women delivered by cesarean section. 455 (13%) women reported that their baby was unhealthy and hospitalized.

Client determinants

Table 2 shows the results of the logistic and linear regression analyses, using the total, personal and setting score as dependents. The results for the personal and setting score are shown in Online Resources 1 and 2. Overall, all four groups of determinants influence to some extent the total, personal and setting experience scores during childbirth.

None of the socio-demographic characteristics influenced the total and personal scores. Repeating the analysis for the setting score identified a client's educational level and whether or not this pregnancy was planned as significant.

The antenatal experiences were of significant influence for the total, personal and setting score in all three scoring models.

Of the care process determinants, only professional continuity, setting continuity and realization of the expected the place of birth were relevant in at least two of the summary scores (total, personal and setting score). Additionally, having an intervention during labor was relevant for the total and personal scores; induced labor impacted the setting score. The reported health outcome of mother and child were of significant influence for the total, personal and setting score. Having an adverse outcome in a previous pregnancy did not influence the total, professional and setting scores.

The goodness of fit was 64% for having a negative experience, 64% for the median score and 29% for the mean score.

Organizational settings (perinatal unit)

Figure 3 shows the experiences scores across perinatal units for each of the three summary scores. For each of the three scoring models, the summary scores were comparable across the perinatal units. The largest difference between the experience scores of the perinatal units was 4.7% for not having a negative experience (total as well as setting scores) (range of the unit total scores: 51.0% – 55.7%; setting scores: 75.2. % – 79.9%) and 14.2% in the median scores (setting score) (range of the unit scores: 42.5% – 56.7%). For the mean score, the maximum difference in experience scores across perinatal units was 0.06 for the setting score (range of the unit scores: 3.78 – 3.84).

Table 2. Impact of antenatal experiences, care process, interventions during birth, and patient reported outcomes on the total experience score during birth, expressed as having a negative experience, above/below the median score of 3.79 and mean score (n=3,479).

Goodness of fit	Overall sign $	Negative experience°			Median experience^			Mean experience		
		64%			64%			28%		
		OR	95% CI	p	OR	95% CI	p	β	95% CI	p
Socio- demographic characteristics										
Age										
≤24		0.74	0.50 – 1.10		0.73	0.48 – 1.10		-0.03	-0.07 – 0.01	
25-29		0.82	0.66 – 1.01		0.98	0.78 – 1.23		0.00	-0.02 – 0.03	
30-34		0.94	0.78 – 1.14		0.95	0.78 – 1.17		0.00	-0.02 – 0.02	
≥35 (ref)										
Parity										
Primiparous (ref)										
Multiparous		0.90	0.76 – 1.06		1.03	0.86 – 1.23		0.00	-0.02 – 0.02	
Ethnic background										
Western (ref)										
Non-Western		1.11	0.81 – 1.52		0.85	0.61 – 1.19		0.01	-0.02 – 0.05	
Educational level				*						
Low		1.36	0.97 – 1.90		1.15	0.81 – 1.64		0.01	-0.03 – 0.05	
Middle		0.88	0.74 – 1.03		1.17	0.98 – 1.39		0.01	-0.01 – 0.03	
High (ref)										
Marital status										*
Married/living together (ref)										
Not living together or no relationship		0.99	0.66 – 1.48		1.01	0.66 – 1.56		-0.09	-0.09 – 0.00	
Planned pregnancy										*
Yes (ref)										
No		0.86	0.67 – 1.12		0.87	0.66 – 1.14		-0.05	-0.08 – -0.03	
Experiences	Y			*			*			*
Antenatal experience		2.10	1.80 – 2.45		3.24	2.77 – 3.79		0.43	0.40 – 0.47	
Care process										
Professional continuity	Y			*			*			*
Yes (ref)										
No		0.75	0.63 – 0.89		0.78	0.65 – 0.93		-0.04	-0.06 – -0.02	

Table 2. *Continued*

Goodness of fit	Overall sign $	Negative experience°			Median experience^			Mean experience		
			64%			64%			28%	
		OR	95% CI	p	OR	95% CI	p	β	95% CI	p
Setting continuity	Y			*			*			*
Primary care entirely (ref)										
Secondary care entirely		0.56	0.42 – 0.74 0.47 – 0.85		0.38	0.28 – 0.52		-0.06	-0.09 – -0.03	
Primary care at first booking, referral to secondary care during pregnancy		0.63			0.34	0.25 – 0.47		-0.04	-0.08 – -0.01	
Primary care at first booking, referral to secondary care during parturition without urgency		0.68	0.52 – 0.88		0.38	0.28 – 0.50		-0.07	-0.10 – -0.04	
Primary care at first booking, referral to secondary care during parturition with urgency		0.59	0.42 – 0.84		0.42	0.29 – 0.61		-0.08	-0.12 – -0.04	
Expected place of birth was realized	Y			*			*			*
Yes (ref)										
No		0.82	0.68 – 0.99		0.74	0.61 – 0.91		-0.03	-0.05 – -0.01	
No prior expectations		0.55	0.35 – 0.87		0.49	0.30 – 0.80		-0.10	-0.15 – -0.05	
Day of birth										
Monday to Friday (ref)										
Saturday/Sunday		1.09	0.93 – 1.29		0.96	0.81 – 1.15		0.01	-0.01 – 0.03	
Time of birth										
8 AM to 5 PM (ref)										
5 PM to 8 AM		1.05	0.89 – 1.24		0.94	0.78 – 1.12		0.00	-0.02 – 0.02	
Intervention										
Induced labor										
No (ref)										
Yes		0.95	0.77 – 1.17		0.92	0.74 – 1.15		0.01	-0.01 – 0.03	
Intervention	Y			*				*		
None / episiotomy (ref)										
Vacuum or forceps extraction		1.00	0.78 – 1.29		1.13	0.86 – 1.48		0.00	-0.03 – 0.03	

Table 2. *Continued*

Goodness of fit	Overall sign $	Negative experience°			Median experience^			Mean experience		
			64%			64%			28%	
		OR	95% CI	p	OR	95% CI	p	β	95% CI	p
Planned cesarean		0.65	0.44 – 0.97		0.84	0.55 – 1.28		-0.03	-0.07 – 0.02	
Emergency cesarean		0.68	0.49 – 0.93		0.78	0.56 – 1.10		-0.08	-0.11 – -0.04	
Pain medication										*
No (ref)										
Yes		1.30	1.08 – 1.56		0.98	0.81 – 1.20		0.02	0.00 – 0.04	
Patient reported outcome										
Outcome baby	Y						*			*
Healthy and not hospitalized (ref)										
Healthy but hospitalized		0.92	0.74 – 1.14		0.84	0.67 – 1.05		-0.02	-0.04 – 0.00	
Unhealthy but not hospitalized		0.91	0.62 – 1.34		0.62	0.41 – 0.93		-0.07	-0.11 – -0.03	
Unhealthy and hospitalized		0.81	0.64 – 1.02		0.75	0.58 – 0.96		-0.07	-0.09 – -0.04	
Outcome mother	Y						*			*
Healthy and not hospitalized (ref)										
Healthy but hospitalized		1.07	0.76 – 1.52		1.06	0.74 – 1.52		0.00	-0.04 – 0.04	
Unhealthy but not hospitalized		0.82	0.68 – 0.99		0.54	0.45 – 0.66		-0.06	-0.08 – -0.04	
Unhealthy and hospitalized		0.96	0.75 – 1.22		0.74	0.57 – 0.95		-0.05	-0.08 – -0.03	
Adverse outcome previous pregnancies										
No (ref)										
Yes		1.08	0.91 – 1.27		1.13	0.95 – 1.34		0.01	-0.01 – 0.03	
Constant		0.71			0.64			2.31		

$ A determinant was overall judged as significant if its estimated coefficient was significant in at least two of the dependent determinants, noted by 'Y'

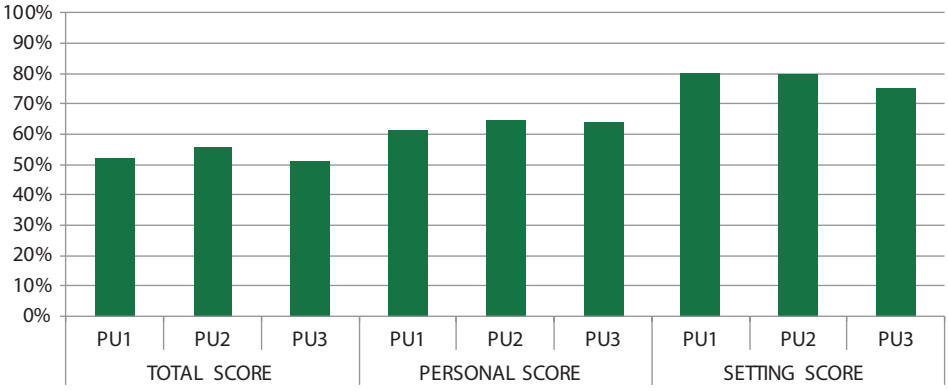
° Not having a negative experience. A negative experience was defined as 'never' in a domain and/or 'sometimes' in the individually chosen 2 most important domains.

^ Above or equal to the median of 3.79

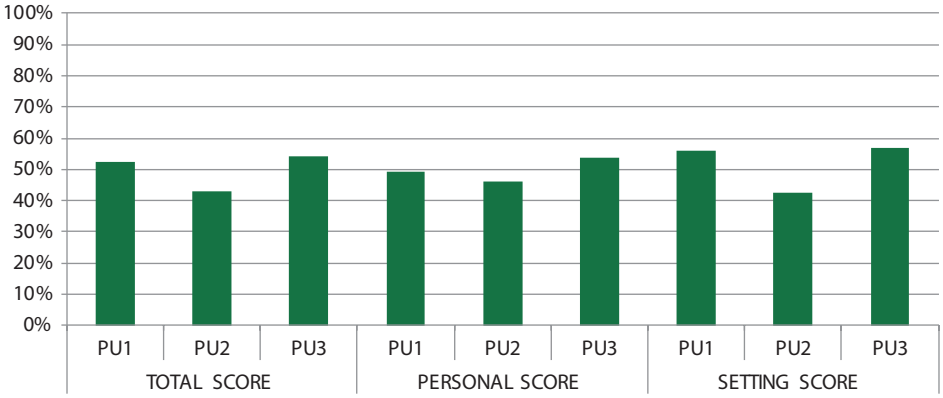
* $p<0.05$

Figure 3. Total, personal and setting scores for each of the three perinatal units with the most respondents, expressed as; a) percentage of women who did not have a negative experience; b) above or equal to the median score, and; c) the mean score (N_{PU1}=212, N_{PU2}=189, N_{PU3}=210).

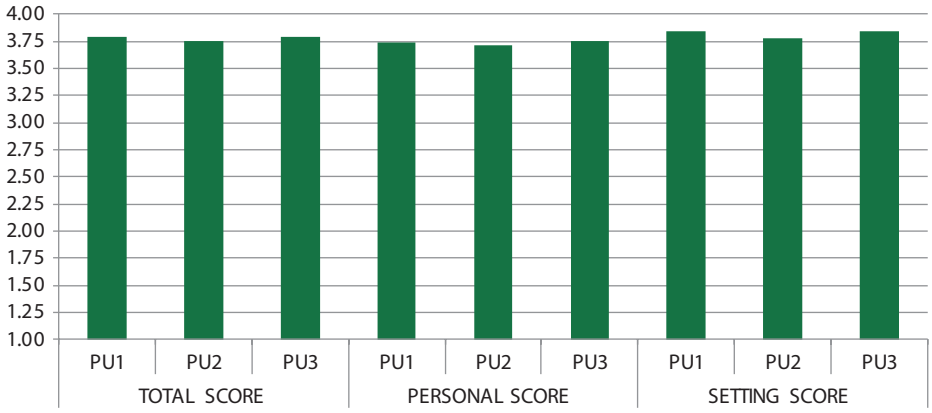
A. Not having a negative experience



B. Above or equal to the median



C. Mean score



DISCUSSION

Our study showed that a woman's experiences during birth were influenced by her socio-demographic characteristics, her experiences of the antenatal phase, the process of care during pregnancy and birth, and patient reported outcomes of mother and child. The impact of organizational settings on the experience scores during birth appeared modest. Consequently, for a fair benchmarking, the following determinants need to be considered for case mix correction: a woman's educational level, whether or not her pregnancy was planned, and the perceived health outcome of mother and child. Second, our findings imply that a low experience score at the organizational level (aggregated individual scores) can be assigned to variation in specific subgroups of clients, i.e.: women with a low antenatal experiences, women who did not experience professional and setting continuity, and women who did not have expectations towards the place of birth.

Strengths

To our knowledge, this is the first large quantitative study, in an unselected pregnant population, which focuses on the determinants of women's experiences during birth. Previous quantitative studies focused on birth satisfaction instead of experiences. To our opinion, experiences provide better clues to start quality improvement, as the experiences only focus on how care as was delivered, while satisfaction with care is the joint outcome of a woman's prior expectations and her actual experiences towards provided care. Satisfaction has the disadvantage that it does not provide specific clues for quality improvement without further investigation. For example, when a woman reports dissatisfaction with care, it is unclear why; perhaps she had unrealistic expectations, or her healthcare professionals did not treat her with respect, or perhaps she had worse PROMS, or a mix of these[19-21].

Second, we measured antenatal experiences separately from the experiences during birth. Most other studies measure both the antenatal and postnatal experiences after birth[10-12,14]. Separate measurement of antenatal and postnatal experiences has the advantage that the experiences during birth and the resulting perceived health outcome cannot influence the antenatal experiences in retrospect. The measurement of the antenatal experiences antenatally supports the validity of our findings.

Finally, the fact that our findings were mainly consistent across the total, personal and setting scores, and for all three scoring models, supports the robustness of our results.

Limitations

First, the sample was representative on the national level with respect to age, parity, induced labor, mode of birth, and use of pain medication[22], but women with a low educational level and non-Western women were underrepresented. Despite the relatively small sample of women with a low educational level, the odds ratios and beta-coefficients were similar to the trend we expected based on previous studies: women with a low educational level had

better experiences than women with a high educational level[23-25]. Similarities in these trends imply that the impact of the underrepresentation of women with a low educational level on the experience scores during birth is limited. For non-Western women, we cannot judge the impact of the underrepresentation of this subgroup, as literature is inconclusive. Our findings that non-Western women on average had better experiences than Western women in the professional domains and worse experiences in the setting domains were consistent with the findings of Raleigh et.al.[25]. However, our current findings are inconsistent with an unpublished stratified analysis that showed that non-Western women had slightly worse experiences than non-Western women for all scores (total, personal and setting)[26].

Second, the remaining determinants cover almost all aspects of the provided care (antenatal experiences, care process, perceived health outcome of mother and child, and organizational settings). Other studies suggest that a woman's level of empowerment might also play a role[20,27-30]. However, we expect empowerment to be of limited influence on our findings, as some of these studies also suggest that a woman's experiences are a determinant of empowerment instead of the other way around[29,30]. For example, having good experiences in terms of Autonomy increases a woman's feeling of being empowered.

Previous studies

Previously, a large Australian study (n=790) indicated that socio-demographics have no significant influence on satisfaction with care during birth[19]. This was confirmed in our study: none of the socio-demographic determinants appeared significant, except for the setting score, where lower educational level and having a planned pregnancy resulted in favorable experiences during childbirth. We have no clear explanation why educational level and a planned pregnancy do influence the setting score, while they do not influence the personal score, as we expected that a woman's educational level would influence both scores. For maternal age and marital status, the trend in odds ratios and beta coefficients were consistent with previous literature[23-25].

Our results also showed that antenatal experiences are one of the main determinants associated with experiences during birth. This was expected as care during pregnancy and birth are highly related: antenatal care aims to create the optimal situation or starting point for childbirth. There is no literature available to compare our results with, due to the conceptual difference between satisfaction and experiences.

The care process was also relevant to the experiences during childbirth. In particular, professional continuity, setting continuity, realization of the expected place of birth and having an intervention during birth were (highly) related to a woman's experiences. This is in agreement with literature[20,23,27,31-33]. This is probably due, at least partly, to a woman's perception of being healthy instead of being a patient. This is most likely is associated with feeling uncertain towards her own health and that of her baby.

As expected, a negatively perceived health outcome of mother and child was associated with a lower experience score during birth. At least three different mechanisms may explain this association. Firstly, good experiences directly contribute to healthy outcome: women are more likely to comply with treatment and care is quickly provided in case of an emergency[2-4]. Secondly, women recollect their experiences during birth when filling out the postnatal questionnaire, these experiences could distort in the direction of the (perceived) health outcome[34,35]. Thirdly, both experiences and health outcome could be influenced by a common factor. Since, the experiences during birth coincide with the health outcomes after birth, it is impossible to determine which of these explanations is most likely.

To our surprise, organizational setting had only limited influence of on women's experiences during childbirth. We expected that the influence of the organizational settings would be reflected in the setting scores as these are indirectly measured in these domains. For example, cooperation between health care professionals is included in Choice and continuity; and protocols and consultations hours are indirectly measured in Prompt attention. As the setting score is similar for all three perinatal units, we can assume that the influence of organizational settings is limited. Moreover, our results could slightly overestimate the true difference across these perinatal units, as case mix correction was not performed and we cannot rule out association between individual determinants and the organizational settings. 'Organizational setting' may represent multiple determinants needing further elaboration.

The degree to which health care organizations are able to influence the determinants of their experiences scores has consequences for the benchmark, as the lack of influence by health care organizations is an important criteria for determinants to be eligible for case mix correction. Applying the debate about the impact of health care organizations on determinants to the included groups of determinants, we argue that both socio-demographic characteristics and health outcome can, to some extent, be influenced by health care organizations: First, we belief that health care organizations should be able to provide care suitable for all clients, independent of their socio-demographic characteristics. Second, we also argue that health care organizations can influence health outcomes by the initiated care process and organizational settings.

Future use

Clients' experiences are often measured as part of a quality improvement cycle or benchmark. When performing a benchmark on the total experience scores of health care organizations, case mix correction need to be considered for the following determinants – in as far these determinants are unequally distributed across organizations: a client's educational level, whether or not this pregnancy was planned, and health outcome of mother and child. To improve the organizational experience score, health care organization should improve care for women with a low antenatal experience, women who did not experience professional

and setting continuity, women who had an intervention during birth and women who were not able to give birth at the excepted place or who had no prior expectations where to give birth. Performing a benchmark on the personal or setting scores require a slightly different set of determinants eligible for case mix correction and potential areas for improvement.

REFERENCES

1. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
2. Valentine NB, De Silva A, Kawabata K, Darby C, Murray CJ, Evans BE. Health system responsiveness: concepts, domains and measurement. In: Murray CJ, Evans BE, eds. *Health systems performance assessment*. Geneva: World Health Organization; 2003:573-596.
3. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Soc Sci Med.* 2000;51(11):1611-1625.
4. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997;45(12):1829-1843.
5. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
6. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
7. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
8. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
9. Sixma H, Hendriks M, De Boer D, Delnoij D. Handboek CQI ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument [Manual CQI development: guidelines and regulations for the development of a CQI measurement]. 2008; http://www.nivel.nl/sites/default/files/bestanden/Handboek-CQI-Ontwikkeling.pdf.
10. Dzakpasu S, Kaczorowski J, Chalmers B, Heaman M, Duggan J, Neusy E. The Canadian maternity experiences survey: design and methods. *J Obstet Gynaecol Can.* 2008;30(3):207-216.
11. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483.
12. Redshaw M, Heikkila K. Delivered with care: A national survey of women's experience of maternity care. 2010; https://www.npeu.ox.ac.uk/delivered-with-care.
13. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg
14. Wiegers TA, Keirse MJ, Berghs GA, van der Zee J. An approach to measuring quality of midwifery care. *J Clin Epidemiol.* 1996;49(3):319-325.
15. CPZ. Uniform gebruik cliëntervaringslijst ReproQ in de Geboortezorg [Uniform use client experience list ReproQ in maternity care]. 2015; http://www.goedgeboren.nl/netwerk/h/215/0/3596/rubriek--9/Uniform-gebruik-cli-ntervaringslijst-ReproQ-in-de-Geboortezorg, 2015.
16. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring Client Experiences in Maternity Care under Change: Development of a Questionnaire Based on the WHO Responsiveness Model. *PLoS One.* 2015;10(2):e0117031.
17. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ.* 2016;4:e2092.
18. Saultz JW, Lochner J. Interpersonal continuity of care and care outcomes: a critical review. *Ann Fam Med.* 2005;3(2):159-166.
19. Brown S, Lumley J. Satisfaction with care in labor and birth: a survey of 790 Australian women. *Birth.* 1994;21(1):4-13.
20. Christiaens W, Bracke P. Assessment of social psychological determinants of satisfaction with childbirth in a cross-national perspective. *BMC pregnancy and childbirth.* 2007;7:26.
21. John J. Patient satisfaction: the impact of past experience. *J Health Care Mark.* 1992;12(3):56-64.
22. Stichting Perinatale Registratie Nederland. *Perinatal zorg in Nederland 2014. (Perinatal care in the Netherlands 2014).* Utrecht: Stichting Perinatale Registratie Nederland;2014.

23. Britton JR. The assessment of satisfaction with care in the perinatal period. *J Psychosom Obstet Gynaecol.* 2012;33(2):37-44.
24. Quintana JM, Gonzalez N, Bilbao A, et al. Predictors of patient satisfaction with hospital health care. *BMC Health Serv Res.* 2006;6:102.
25. Raleigh VS, Hussey D, Seccombe I, Hallt K. Ethnic and social inequalities in women's experience of maternity care in England: results of a national survey. *J R Soc Med.* 2010;103(5):188-198.
26. Scheerhagen M, Van Stel HF, Birnie E, Franx A, Bonsel GJ. *Het discriminerend vermogen van de ReproQuestionnaire (the discriminative power of the ReproQuestionnaire).* Rotterdam, The Netherlands: Erasmus Medical Center;2015.
27. Goodman P, Mackey MC, Tavakoli AS. Factors related to childbirth satisfaction. *J Adv Nurs.* 2004;46(2):212-219.
28. Mei JY, Afshar Y, Gregory KD, Kilpatrick SJ, Esakoff TF. Birth Plans: What Matters for Birth Experience Satisfaction. *Birth.* 2016;43(2):144-150.
29. Nilsson L, Thorsell T, Hertfelt Wahn E, Ekstrom A. Factors influencing positive birth experiences of first-time mothers. *Nurs Res Pract.* 2013;2013:349124.
30. Hardin AM, Buckner EB. Characteristics of a positive experience for women who have unmedicated childbirth. *J Perinat Educ.* 2004;13(4):10-16.
31. de Jonge A, Stuijt R, Eijke I, Westerman MJ. Continuity of care: what matters to women when they are referred from primary to secondary care during labour? a qualitative interview study in the Netherlands. *BMC pregnancy and childbirth.* 2014;14:103.
32. Price RA, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Medical Care Research and Review.* 2014;71(5):522-554.
33. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC pregnancy and childbirth.* 2009;9:18.
34. Pohl RF, Bender M, Lachmann G. Hindsight bias around the world. *Exp Psychol.* 2002;49(4):270-282.
35. Ruoss M. [Pain patients show a higher hindsight bias]. *Z Exp Psychol.* 1997;44(4):561-588.

5

The Discriminative Power of the ReproQ: A Client Experience Questionnaire in Maternity Care

M. Scheerhagen, H.F. van Stel, A. Franx, E. Birnie, G.J. Bonsel

Submitted

ABSTRACT

Background. Aim of the ReproQuestionnaire (ReproQ) is to measure the client's experience with maternity care, following the WHO's responsiveness model. To support quality improvement, ReproQ should be able to discriminate best from worst organisational units.

Methods and findings. We sent questionnaires to 27,487 third-trimester pregnant women (response 32%) and to 37,230 women 6 weeks after childbirth (response 32%). First, we summarized the ReproQ domain scores into three summary scores: total score (all eight domains), personal score (four personal domains), and setting score (four setting domains). Second, we estimated which proportion of variance between perinatal units is due to 'true' difference across perinatal unit, using intraclass correlation coefficients (ICCs). Third, we assessed the ability of ReproQ to discriminate between perinatal units based on both a statistical approach using multilevel regression analyses, and a relevance approach based on the minimally important difference (MID). Finally, we compared the domain scores of the best and underperforming units.

ICCs ranged between 0.004 and 0.025 for the summary scores, and between 0.002 and 0.125 for the individual domains. ReproQ was able to identify the best and worst performing units with both the statistical and relevance approach. The statistical approach was able to identify four underperforming units during childbirth (total score), while the relevance approach identified ten underperforming units.

Conclusion. ReproQ, a valid and efficient measure of client experiences in maternity care, has the ability to discriminate well across perinatal units, and is suitable for benchmarking under routine conditions.

INTRODUCTION

The performance of health care systems is primarily judged by health outcomes, such as morbidity, mortality, health status, or burden of disease. System performance differs across and within countries, partly caused by differences in the provision of care[1-5]. To highlight the role of provision of care in health system performance, the World Health Organization (WHO) introduced the measurement of client experiences with service provision and service quality as a cornerstone in health care provider evaluations[6,7].

Client experiences with care provision matter for at least two reasons. First, these may guide the client's choice of health care provider, particularly when the health outcomes across providers are about similar[7]. Second, better client experiences may contribute to improved health outcomes[6-11]. For example, clients who understand their caregiver's explanations are more likely to comply to treatment or lifestyle changes.

To cover a broad spectrum of client experiences, independent from specific system characteristics and relevant to all medical professionals and settings, the WHO developed the so-called Responsiveness concept. Responsiveness is defined as the way a client is treated by the professional and the environment in which the client is treated during a health care encounter[7,12,13]. It is based on actual performance in health practice, rather than on organisational features with claimed benefit.

Performance-based quality improvement universally proceeds in two stages. In the first stage, through aggregated client results, health care providers are ranked in terms of performance. In the second stage, each organisation digs into the differences responsible for the deviant result by disaggregation of summary scores into domain scores and/or item scores. Also, the varying performance of subgroups, e.g. deprived clients, is advocated[14-17].

A major challenge in performance measurement in maternity care, with clients being predominantly healthy young women, is the discriminative power of a measure to quantify client experiences. Poor outcomes are infrequent, and specific low performance (for clients and units) into one direction easily averages out into other directions. Moreover, the variation in performance scores across units can be the result of variation in performance at the unit level of essentially homogeneous clients, or true variation in individual performance at the client level. I.e., the so-called nested or hierarchical nature of performance data requires a statistical approach to expose the 'true' perinatal unit performance, but above all a measure with should excellent measurement properties without becoming to lengthy.

The study aim was to evaluate the discriminative power of ReproQ at the perinatal unit level (a hospital with its associated community midwife practices). ReproQ is a validated questionnaire to measure client experiences with maternity care based on WHO's Responsiveness concept[18]. We use two approaches to determine discriminative ability. The first, conventional, approach is to identify poor performing perinatal units on the basis of a statistically significant difference from the average performance score of all perinatal units, taking the nested nature of the data into account. In the second approach, we identified a

perinatal unit as poor performer if its aggregated score deviated by a minimally important or 'meaningful' difference (MID) from the best performing units. We assume that ReproQ has sufficient discriminative power for national implementation in the Dutch maternity care system if ReproQ shows discriminatory power in both approaches, and if targets for improvement can be identified.

METHODS

ReproQuestionnaire

ReproQ (33 items) consists of two complementary versions; the antepartum questionnaire addresses women's experiences in the first and second half of pregnancy, while the postpartum questionnaire addresses women's experiences the childbirth and the subsequent postpartum week. Previous results indicate that ratings of the 1st half of pregnancy remain are highly associated (ICC=0.80) with the 2nd half of pregnancy.

The eight-domain WHO Responsiveness concept was used as the conceptual base[6,7]. The four domains on personal interactions between the client and health professional are: Dignity, Autonomy, Confidentiality and Communication. The four domains regarding experiences with the organizational setting are: Prompt attention, Access to family and Community support, Quality of Basic amenities, and Choice and continuity of care. The client's responses can be summarized as a) eight separate domain scores, b) the personal summary score (covering the four 'personal' domains) and the setting summary score (covering the four 'setting'-related domains), and c) the total score (covering all eight domains); a higher score implies better performance. Each score can be presented for each of the four reference periods. The summary scores of clients can be subsequently aggregated by health care provider, organisational unit, or region.

Psychometric analyses support the content and construct validity as well as the test-retest reliability of the questionnaire[18,19].

ReproQ data collection

ReproQ data were collected digitally from clients of three large maternity care organizations (that deliver postnatal care at home from childbirth onwards during 7 – 10 days), from several ongoing regional observational studies, and from clients of 10 perinatal units interested in quality improvement. A total of 60 of 85 Dutch perinatal units participated. Further details have been reported elsewhere[19].

All women could participate. All participants gave informed consent. Invitations to fill out the antepartum ReproQ were sent around 34th weeks' gestational age. The postpartum ReproQ was sent 6 weeks after the expected date of childbirth. Non-responding clients received a reminder two weeks after the initial invitation. The study protocol and procedures were approved by the Medical Research Ethics Review Board, Erasmus Medical Center, Rotterdam, the Netherlands (MEC-2013-455).

Unit of analysis: Perinatal unit

The perinatal unit was the unit of analysis. Since each hospital can be assigned uniquely to one perinatal unit, clients were allocated to the hospital's perinatal unit in case of a hospital birth. In case of an out-of-hospital birth, clients were allocated to the hospital (and perinatal unit) that was closest to the client's home address.

Descriptive characteristics of perinatal units were obtained from various public sources[20,21,22].

Excluded data

Excluded from analysis were: 1) ReproQ data of clients who could not be allocated to one perinatal unit; 2) ReproQ data with >50% missing answers in two or more ReproQ domains; and 3) data of perinatal units who included less than 50 clients.

Analytical framework: multilevel analysis

Crude differences in summary ReproQ scores across perinatal units, the dependent variable, may originate from three sources: 1) 'true' differences across perinatal units, 2) differences in client characteristics across perinatal units, 3) residual variance. Given the hierarchical data structure (perinatal units, and clients within perinatal units), existing differences in client characteristics across perinatal units may obscure the estimation of the 'true' difference across units. In that case, multi-level analysis is the appropriate method to decompose total data variance into variance attributable to perinatal units (source 1) and variance attributable to other sources (sources 2+3), in particular variance related to client characteristics[23]. Estimation of the 'true' difference between perinatal units (source 1) requires the domain and summary scores to be corrected for the other variance components (typically client characteristics), as systematic client diversity may bias and limit the comparison of perinatal units; i.e. the case mix correction. Client characteristics included in the case mix correction were client's age, educational level, and self-rated health.

The Technical Supplement shows further details on the multilevel analysis and software used.

Discriminative ability: Two approaches

We used two complementary approaches to determine discriminative power.

Approach 1. Multilevel testing of the deviation of unit means from overall (grand) mean

Multi-level analyses were used to examine to what degree ReproQ is able to identify units that significantly perform above and below average (averaged over perinatal units), producing three parts of information:

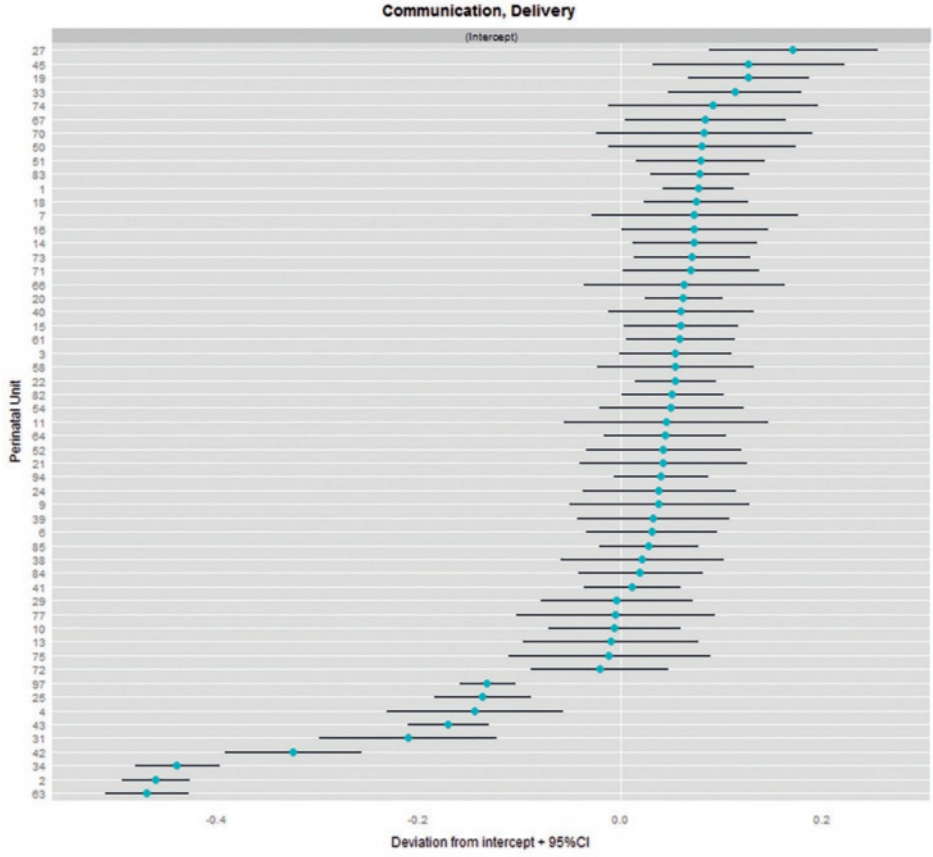
1) Estimated variance components and ICCs. The ICC of interest is the ratio of the variance in perinatal unit and the variance in client's characteristics in that unit[24,25]. An ICC close to zero implies that the client's experience is unrelated to the perinatal unit in which one receives care; an ICC close to 1 means that the perinatal unit is of decisive importance. Best and poor performing units are identified by the deviation of the 95% CI of each individual perinatal unit from the grand mean of all perinatal units.
2) Estimated G-coefficients, which represent the proportion of variance in the unit-level mean scores attributable to 'true' variation among perinatal units. A G-coefficient of one implies that all variance in domain and summary scores across perinatal units is attributed to the perinatal unit, and no variance can be attributed to other sources.
3) Estimation of the minimal number of clients needed per perinatal unit to achieve sufficient reliability (D-Study), in our study defined as 0.80. Small numbers of clients and large heterogeneity in client experiences produce wide confidence intervals, but only the numbers of clients can be influenced.

The conventional mode of presentation is the so-called caterpillar-plot; see Figure 1.

Approach 2. Relevant deviation based on MID

This approach judges ReproQ's discriminative power on the basis of the ability to demonstrate a relevant difference in domain or summary scores at the perinatal unit level, the so-called minimally important difference (MID). Underperforming units are identified as those units with domain or summary scores below a certain threshold that equals the mean score domain or summary score of the 10% best performing units minus the MID.We previously determined, at the individual level, the minimally important difference (MID) for both the summary and domain scores of the childbirth phase[19]. For this paper, we derived MIDs for the late pregnancy and postnatal period in a similar way; see Table 2. Note that a 1 unit of MID difference at the perinatal unit level reflects a large difference: it means that *all* clients cared for in that unit, on average differ 1 MID from a reference value, either being much better (best practice) or much worse (poor practice). We also presented results for a more conservative 0.5 MID.

Figure 1. Caterpillar-plot: Ranking of perinatal units for the domain Communication during childbirth (N_{PU}=55).



RESULTS

We invited 27,487 pregnant women to participate in the antepartum ReproQ (response: 8,567 (31%)) and 37,230 women who recently had given birth to respond to the postpartum ReproQ (response: 12,477 (39%)). Excluded from analysis were 1,419 pregnant women and 1,751 women who recently had given birth, for having >50% missing answers. Additionally, we excluded 761 pregnant women and 1,080 women who recently had given birth, for whom the perinatal unit code was missing or being a perinatal unit with less than 50 responses.

Table 1. Characteristics of the participating women ($n_{antepartum}$=6,387; $n_{postpartum}$=9,646) and perinatal units ($n_{antepartum}$=42; $n_{postpartum}$=55).

	Antepartum questionnaire		Postpartum questionnaire	
	N	%	N	%
Clients				
Age (years)*	385	6	500	5
≤24				
25 – 29	2,018	32	2,730	29
30 – 34	2,600	42	4,084	43
≥35	1,263	20	2,197	23
Ethnic background				
Western	5,735	92	8,711	93
Non – Western	478	8	696	7
Educational level**				
Low	399	6	754	8
Middle	2,026	33	3,280	35
High	3,783	61	5,356	57
Marital status				
Married/living together	5,974	96	9,052	96
Not living together or no relationship	226	4	339	4
Parity				
Primiparous	3,210	50	4,872	51
Multiparous	3,153	50	4,735	49
Self-reported health status				
Poor / moderate	300	5	332	4
Good	2,173	36	3,153	33
Very good	2,390	39	3,684	38
Excellent	1,244	20	2,428	25
Perinatal units				
Number of respondents#				
50 – 99	16	38	19	35
100 – 149	10	24	14	25
150 – 199	6	14	7	13
≥200	10	24	15	27
Urbanization				
Urban – 4 largest cities	10	24	14	25
Urban – 10 largest cities, except 4 largest	6	14	6	11
Rural	26	62	35	64

Table 1. *Continued*

	Antepartum questionnaire		Postpartum questionnaire	
	N	%	N	%
Hospital type in the perinatal unit				
University hospital	5	12	6	11
Teaching hospital	17	40	20	36
Non – teaching hospital	20	48	29	53
Hospital size				
<750 deliveries per year	5	12	6	11
750 – 1499 deliveries per year	20	48	26	47
≥1500 deliveries per year	17	40	23	42

* Mean age was 30.1 years (SD=4.5)
** Educational level; low 0 – 6 years; middle 6 – 12 years; high >12 years.
# Mean number of respondents per perinatal unit was 152 (range: 54 – 363) for the antenatal period, and 175 (range: 50 – 812) for the postnatal period.

Table 1 shows the characteristics of participating clients and perinatal units. Differences between the antenatal and postnatal client and unit characteristics were minimal, and about representative for the national pregnancy population.

Table 2 presents the results of the multi-level analysis of the corrected model which includes perinatal unit as random effect and client characteristics for case mix correction. Each row represents a separate analysis for the experience measure shown (total score, domain score) in each of the three periods (pregnancy, childbirth, postnatally). Columns 2 – 4 provide the estimated ICCs, or the ratio of variance assigned to perinatal units (column 2) and variance assigned to client characteristics (column 3). For example, the first row shows the results for the antenatal total ReproQ score. It shows that little variance on the client level can be assigned to the perinatal units in general (0.0001), and somewhat more to the variance of client characteristics (0.064). The ICC is 0.011, indicating that the client's experience is to a limited extent related to the perinatal unit in which one received care. The 5th column shows that the G-coefficients (or proportion of variance in the mean scores of perinatal units that can be attributed to the "true" variation among perinatal units; the higher the better) of the total score during pregnancy is 0.63. Finally, the 6th column shows that at least 272 respondents per unit are needed to achieve a reliability (G-coefficient) of 0.80.

Table 2. Results of corrected multi-level analysis model and G-study for ReproQ summary scores and domain scores during pregnancy (n=6,387), and childbirth and postnatal period (n=9,646).

	Variance of perinatal units	Variance of client characteristics	ICC	G-coefficient*	Number of respondents needed for G-coefficient of 0.8
Pregnancy					
Total score	0.001	0.064	0.011	0.63	272
Personal score	0.000	0.071	0.004	0.45	580
Setting score	0.001	0.082	0.014	0.66	243
Dignity	0.000	0.073	0.004	0.38	745
Autonomy	0.001	0.190	0.006	0.45	555
Confidentiality	0.001	0.204	0.005	0.38	755
Communication	0.000	0.115	0.004	0.35	875
Prompt attention	0.002	0.104	0.021	0.74	165
Social considerations	0.001	0.154	0.004	0.35	825
Basic amenities	0.002	0.111	0.019	0.70	202
Choice and continuity	0.001	0.271	0.003	0.42	630
Childbirth					
Total score	0.001	0.075	0.009	0.50	432
Personal score	0.002	0.101	0.019	0.71	176
Setting score	0.001	0.072	0.008	0.49	465
Dignity	0.000	0.115	0.002	0.19	1,910
Autonomy	0.007	0.333	0.020	0.66	210
Confidentiality	0.001	0.214	0.002	0.23	1,465
Communication	0.021	0.147	0.125	0.96	18
Prompt attention	0.001	0.123	0.009	0.46	523
Social considerations	0.000	0.105	0.003	0.23	1,480
Basic amenities	0.002	0.080	0.023	0.73	165
Choice and continuity	0.001	0.260	0.002	0.23	1,440

Table 2. *Continued*

	Variance of perinatal units	Variance of client characteristics	ICC	G-coefficient*	Number of respondents needed for G-coefficient of 0.8
Postnatal period					
Total score	0.001	0.080	0.015	0.63	258
Personal score	0.001	0.100	0.012	0.58	320
Setting score	0.002	0.081	0.025	0.75	145
Dignity	0.001	0.122	0.009	0.51	425
Autonomy	0.001	0.287	0.004	0.30	1,015
Confidentiality	0.001	0.212	0.007	0.42	600
Communication	0.011	0.145	0.073	0.93	35
Prompt attention	0.002	0.096	0.018	0.66	221
Social considerations	0.001	0.126	0.008	0.51	418
Basic amenities	0.004	0.108	0.035	0.78	123
Choice and continuity	0.003	0.268	0.013	0.64	249

*Mean valid response per perinatal unit was 116 antenatally, and 109 postnatally.

As Table 2 shows, all ICCs of the three summary scores for all three reference periods were lower than 0.03. Moreover, the ICCs of the case mix corrected models range from 0.002 (the domains Dignity, Confidentiality and Choice and continuity during childbirth) to 0.125 (Communication during childbirth). Moreover, the ICCs for the individual domains showed more variability than the summary scores, with Communication showing the highest ICC.

The G-coefficients of the summary scores ranged from 0.45 (personal score during pregnancy) to 0.75 (setting score during postnatal period). The G-coefficients of the domain scores ranged between 0.19 (Dignity during childbirth) to 0.93 and 0.96 (Communication during childbirth and postnatal period). For the antenatal period, Prompt attention (0.74) and Basic amenities (0.70), both part of the setting score, were the domains with highest G-coefficients.

The number of respondents needed to achieve a G-coefficient of 0.8 ranged from 18 (Communication during Childbirth) to 1,910 (Dignity during Childbirth). The total scores of the ReproQ would achieve excellent reliability (G-coefficient of 0.8) when all perinatal units would have included 272 (antenatal), 432 (childbirth), and 258 (postnatal period) valid responses.

Figure 1 shows the caterpillar-plot for the Communication domain during childbirth. Depicted are the corrected means (and 95% CIs) of all 55 perinatal units, which allows for comparison with the grand mean of all perinatal units. The varying CI widths point to heterogeneity (after case mix correction) and different sample sizes per unit. For example, unit 22 performs only moderately better, and does so significantly, due to small dispersion.

Table 3. Discriminative power of the ReproQ based on statistical power (significance perspective) and the ability to detect 1.0 and 0.5 MID difference (relevance perspective) for all ReproQ outcomes during pregnancy (n_{pu}=42), childbirth and postnatal period (n_{pu}=55).

	Overall mean	Discriminative power based on statistics			Discriminative power based on relevance			
		Best-practices	Average	Under performers	Mean best practices (ΔP90 –99 P100)	MID	Under performers (ΔP90 1MID)	Under performers (ΔP90 0.5MID)
Pregnancy								
Total score	3.73	3	38	1	3.80	0.11	7	22
Personal score	3.75	0	42	0	3.81	0.09	7	28
Setting score	3.72	3	35	4	3.79	0.12	9	23
Dignity	3.87	0	42	0	3.91	0.07	10	25
Autonomy	3.65	1	40	1	3.75	0.11	21	34
Confidentiality	3.73	0	41	1	3.81	0.09	12	36
Communication	3.75	0	42	0	3.80	0.11	6	21
Prompt attention	3.67	5	32	5	3.77	0.11	19	30
Social considerations	3.76	0	42	0	3.84	0.09	16	30
Basic amenities	3.81	4	34	4	3.89	0.08	22	29
Choice and continuity	3.63	0	42	0	3.74	0.19	7	26
Childbirth								
Total score	3.73	1	50	4	3.80	0.10	10	40
Personal score	3.66	5	46	4	3.75	0.11	13	39
Setting score	3.79	4	50	1	3.86	0.08	21	44
Dignity	3.83	1	54	0	3.88	0.09	9	35
Autonomy	3.44	7	42	6	3.61	0.17	25	43
Confidentiality	3.77	0	55	0	3.84	0.08	32	49

Table 3. *Continued*

	Overall mean	Discriminative power based on statistics			Discriminative power based on relevance			
		Best-practices	Average	Under performers	Mean best practices (ΔP90 –99 P100)	MID	Under performers (ΔP90 1MID)	Under performers (ΔP90 0.5MID)
Communication	3.62	15	32	9	3.80	0.11	17	44
Prompt attention	3.77	1	51	3	3.84	0.10	18	42
Social considerations	3.86	0	55	0	3.92	0.04	42	50
Basic amenities	3.88	7	39	9	3.94	0.05	36	48
Choice and continuity	3.66	1	54	0	3.77	0.13	23	44
Postnatal period								
Total score	3.74	5	45	5	3.83	0.10	18	45
Personal score	3.71	4	47	4	3.80	0.10	19	46
Setting score	3.78	8	39	8	3.86	0.10	29	42
Dignity	3.81	4	48	3	3.90	0.10	21	45
Autonomy	3.71	3	52	0	3.82	0.12	31	47
Confidentiality	3.74	3	50	2	3.82	0.11	23	41
Communication	3.59	13	32	10	3.75	0.09	33	47
Prompt attention	3.80	6	42	7	3.87	0.09	21	41
Social considerations	3.81	5	49	1	3.89	0.07	33	48
Basic amenities	3.86	7	40	8	3.95	0.08	27	44
Choice and continuity	3.65	5	49	1	3.79	0.14	34	48

* The mean best practices is the pooled average of the 10% best performing units.

Table 3 presents the discriminative power according to the statistical and relevance-based approaches. Using the total score during pregnancy (1st row) discriminative power according to the statistical approach would imply that 3 perinatal units showed a significantly better total score compared to the grand mean (3rd column), 38 where about average (4th column), and 1 unit showed a below-average score (5th column). For the total score, discriminative power using the statistics-based approach was largest for the postnatal period (10/55 units being deviant), followed by childbirth (5/55 being deviant) and antenatal period (4/38 being deviant). Of the summary scores, only the personal score in the antenatal period did not statistically discriminate. Overall, the domains Communication and Basic amenities during childbirth (with 23/55 and 16/55 units, respectively, being deviant) and during postnatal period (23/55 and 15/55, respectively) were the domains that discriminated best, due its high reliability (see Table 2).

Table 3 columns 6 – 9 reveal ReproQ's discriminative power based on the MID. For the total score during pregnancy (1st row), the 10% best performing units have a mean total score of 3.80 (reference value). The corresponding MID is 0.11. Applying this MID of 0.11 implies that 7 perinatal units with their CI perform below this reference. For the summary scores, the number of perinatal units that differed more than 1.0 unit of MID compared to the reference value ranged from 7 (both the total and personal scores in pregnancy) to 29 (setting score in postnatal period). The domains with most discriminating power differed for the three reference periods: Autonomy and Basic amenities during pregnancy, Basic amenities and Social considerations during childbirth, and Communication, Social considerations and Choice and continuity during postnatal period (see column 8). Applying a 0.5 unit of MID considerably increased the number of units that relevantly deviated from reference for all scores and periods.

Appendix A describes how these results can be used for the second stage of the benchmark.

DISCUSSION

The discriminative ability of ReproQ, an instrument for measuring service childbirth in maternity care, showed the ability of ReproQ to discriminate well across perinatal units using two complementary approaches; a multilevel significance-based approach, and a relevance-based analysis (MID). It did so despite four conditions that could limit discriminative performance: a predominantly healthy and relatively homogenous population, standardized care procedures, a naturalistic study design, and the use of aggregated means. Using the total score during childbirth, the significance approach identified 4 underperforming units, whereas the MID-based approach identified 10 underperforming units using a 10% best-practice norm and 1.0 MID as cut-off. Once the underperforming units are identified, the unit results can be disaggregated into domain and item information as input for quality improvement.

A study strength is that sample size was large and data were collected in routine practice, covering about 2/3 of perinatal units. Clients and practices covered the full range of relevant characteristics, adding to generalizability.

Secondly, the two approaches yielded consistent results, with the MID-based approach displaying considerably more sensitivity. The significance-based approach distinguishes between the observed outcome distribution (here: unit means) and simply tests if 'unit' has a significant impact on outcomes. Significance relies on the number of respondents, case mix correction, and details of the multilevel analysis, with the intrinsic risk that homogeneity of clients and units can lead to significant results, even when these differences are not meaningful. The reverse is more common: client heterogeneity within units and measurement error can obscure relevant unit differences. The relevance-based approach inevitably relies on the chosen reference (i.e. 10% best practices) and magnitude of the MID. One should be aware that all our choices in both approaches were rather conservative, and that an *average* difference of 1.0 MID at the unit level expresses a rather extreme difference. Other studies on client experiences only explored the discriminative power in statistical terms[26-29]. We believe the MID-based approach is a necessary complement to the significance-based approach.

A third strength is that we avoided overfitting and overcorrection by limiting case mix correction to predefined candidate factors with an accepted established effect[30].

Two limitations merit discussion. Firstly, while the respondents were largely representative for the pregnancy population, non-Western women were somewhat underrepresented (8% vs. national average 14%[22]). Since non-Western women tend to report more negative experiences than Western women[31], increased participation of non-Western women would probably lower the average summary scores but not affect the ranking of perinatal units, as our case mix analysis did not reveal a significant role of ethnicity.

Secondly, we did not include the individual professional as additional level in the analysis. One may assume an effect of individual professional's behaviour on the personal domains rather than the setting domains, and its impact is probably larger than the variation across units[29,32]. While the primary focus of quality improvement is the unit, one should be aware of the professional's role in in quality improvement cycles. Inclusion of the professional in the analysis would require a highly detailed; perhaps unfeasible registration of all caregivers involved the care process.

Three technical remarks can be made. First, a study of performance at the domain level requires about 450 respondents per unit, which is considerably higher than the minimum of 10 respondents per unit adopted in similar studies[26,27]. The view that 10 respondents are representative for a unit's performance is highly questionable, given the variability in respondent characteristics, in experiences within a unit, and in the care provided. A sample size of 450 clients is well below the average unit size of 2,000 clients, implying that a sampling approach instead of all-client measurement should suffice.

Secondly, although the estimated ICCs appeared low, they are comparable to the ICCs of other accepted client experience questionnaires[26-29]. ICCs are low because the denominator essentially is the number of client questionnaires. The impact of each unit on each individual questionnaire is small. Small effects at the respondent level may build up as large effects at the unit level.

Finally, the systematic effect of perinatal unit was stronger during childbirth and postnatally than antenatally. The likely explanation is that antenatal care is highly standardized in terms of procedures and professionals involved whereas different processes, adverse events and outcomes, do emerge during childbirth and postnatally, where unit quality is challenged. This observation emphasizes that the very assumption that quality differs across units may not be true when care is highly standardized. In that case, differences across units truly are small, causing low ICCs and lack of discrimination. We believe that favorable antenatal performance should be interpreted primarily as uniform performance rather than good performance[18,19]. This phenomenon has been described with other client experience questionnaires as well[33-36].

For the future, we recommend the use of ReproQ in maternity care. Once underperforming units have been identified, profiling of its items (which are described in terms of activities) may guide interventions to improve quality (an example is provided in Appendix A). Qualitative interviews and client involvement may further support the interpretation of domain and item scores. This may fit well in outcome-based strategies like those initiated by ICHOM[37], that includes both medical outcomes and quality of care.

CONCLUSION

ReproQ, a valid and efficient measure of client experiences in maternity care, has the ability to discriminate well across perinatal units, and is suitable for benchmarking under routine conditions.

ACKNOWLEGDEMENTS

We are grateful to Koen Ilja Nijenhuijs for his assistance with the caterpillar plots.

FUNDING

This study was funded by Stichting Miletus, www.stichtingmiletus.nl. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. de Graaf JP, Ravelli AC, de Haan MA, Steegers EA, Bonsel GJ. Living in deprived urban districts increases perinatal health inequalities. *J Matern Fetal Neonatal Med.* 2013;26(5):473-481.
2. de Graaf JP, Schutte JM, Poeran JJ, van Roosmalen J, Bonsel GJ, Steegers EA. Regional differences in Dutch maternal mortality. *Bjog.* 2012;119(5):582-588.
3. Mohangoo AD, Buitendijk SE, Szamotulska K, et al. Gestational age patterns of fetal and neonatal mortality in Europe: results from the Euro-Peristat project. *PLoS One.* 2011;6(11):e24727.
4. Zeitlin J, Mohangoo AD, Delnord M, Cuttini M, Committee E-PS. The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *J Epidemiol Community Health.* 2013;67(12):983-985.
5. Zeitlin J, Szamotulska K, Drewniak N, et al. Preterm birth time trends in Europe: a study of 19 countries. *Bjog.* 2013;120(11):1356-1365.
6. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
7. Valentine NB, De Silva A, Kawabata K, Darby C, Murray CJ, Evans BE. Health system responsiveness: concepts, domains and measurement. In: Murray CJ, Evans BE, eds. *Health systems performance assessment*. Geneva: World Health Organization; 2003:573-596.
8. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Soc Sci Med.* 2000;51(11):1611-1625.
9. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997;45(12):1829-1843.
10. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med.* 1998;47(10):1573-1588.
11. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med.* 1994;38(4):509-516.
12. de Silva A. A framework for measuring responsiveness. *Global Programme on Evidence for Health Policy Discussion Paper 32* 2000.
13. Gostin L. The domains of health responsiveness: a human rights assessment. 2002.
14. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
15. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
16. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
17. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
18. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring Client Experiences in Maternity Care under Change: Development of a Questionnaire Based on the WHO Responsiveness Model. *PLoS One.* 2015;10(2):e0117031.
19. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ.* 2016;4:e2092.
20. Posthumus AG, Scholmerich VL, Steegers EA, Kawachi I, Denktas S. The association of ethnic minority density with late entry into antenatal care in the Netherlands. *PLoS One.* 2015;10(4):e0122720.
21. STZ foundation. http://www.stz.nl/1087/over-ons/stz-ziekenhuizen. 2014. Accessed October, 24th, 2015.
22. PRN Foundation. Netherlands Perinatal Registry. *2004-2013.* 2013.
23. Twisk JWR. *Applied Multilevel Analysis.* Cambridge University Press; 2014.
24. Streiner DL, Norman GR. *Health measurement scales.* Oxford: University Press; 2008.
25. Streiner DL, Norman GR, Cairney J. *Health measurement scales.* Oxford: University Press; 2014.

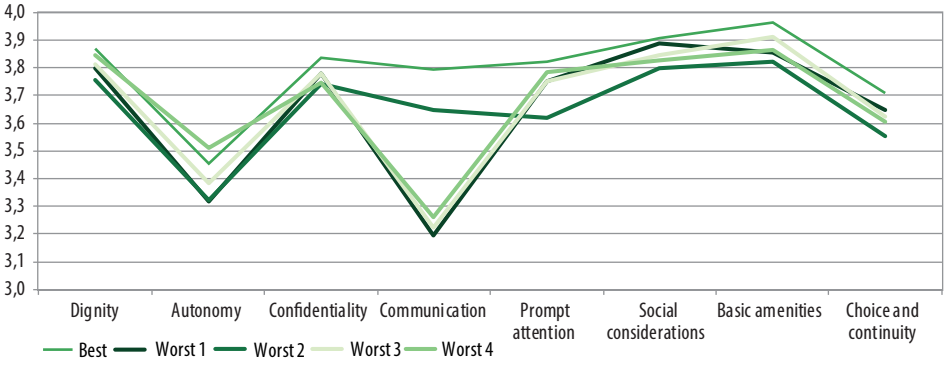
26. Bos N, Sturms LM, Stellato RK, Schrijvers AJ, van Stel HF. The Consumer Quality Index in an accident and emergency department: internal consistency, validity and discriminative capacity. *Health Expect.* 2015;18(5):1426-1438.
27. Stubbe JH, Brouwer W, Delnoij DM. Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire. *BMC Ophthalmol.* 2007;7:14.
28. de Boer D, Delnoij D, Rademakers J. The discriminative power of patient experience surveys. *BMC health services research.* 2011;11:332.
29. Krol MW, De Boer D, Sixma H, Van Der Hoek L, Rademakers JJ, Delnoij DM. Patient experiences of inpatient hospital care: a department matter and a hospital matter. *Int J Qual Health Care.* 2015;27(1):17-25.
30. Sixma H, Hendriks M, De Boer D, Delnoij D. Handboek CQI ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument [Manual CQI development: guidelines and regulations for the development of a CQI measurement]. 2008; http://www.nivel.nl/sites/default/files/bestanden/Handboek-CQI-Ontwikkeling.pdf.
31. Scheerhagen M, Van Stel HF, Birnie E, Franx A, Bonsel GJ. *Het discriminerend vermogen van de ReproQuestionnaire (the discriminative power of the ReproQuestionnaire).* Rotterdam, The Netherlands: Erasmus Medical Center;2015.
32. Roberts MJ, Campbell JL, Abel GA, et al. Understanding high and low patient experience scores in primary care: analysis of patients' survey data for general practices and individual doctors. *Bmj.* 2014;349:g6034.
33. Peterson WE, Charles C, DiCenso A, Sword W. The Newcastle Satisfaction with Nursing Scales: a valid measure of maternal satisfaction with inpatient postpartum nursing care. *J Adv Nurs.* 2005;52(6):672-681.
34. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* Oxford, United Kingdom: National Perintal Epidemiology Unit;2010.
35. Smith LF. Development of a multidimensional labour satisfaction questionnaire: dimensions, validity, and internal reliability. *Qual Health Care.* 2001;10(1):17-22.
36. Smith LF. Postnatal care: development of a psychometric multidimensional satisfaction questionnaire (the WOMBPNSQ) to assess women's views. *Br J Gen Pract.* 2011;61(591):e628-637.
37. ICHOM. Pregnancy and childbirth. 2016; http://www.ichom.org/medical-conditions/pregnancy-and-childbirth/. Accessed july, 1st 2018.

APPENDIX A. PROFILING UNDERPERFORMING UNITS AND QUALITY IMPROVEMENT

Clients' experiences in maternity care are routinely measured in several countries, but the results from the benchmark are hardly used in stage 2 of the two-stage quality cycle. In this Appendix, we illustrate how the ranking of units from stage 1 can be used to achieve quality improvement in stage 2. After the identification of the underperforming units each organisation digs into the data, to find the sources of their deviant result. Summary scores are then disaggregated into domain scores and/or item scores (stage 2); so-called 'profiling' of underperforming perinatal units. Also, the varying performance of client subgroups, e.g. deprived clients, may be taken into account[1-4].

For the 'profiling', we compared the domain scores (for the childbirth period) of the statistically best performing perinatal units with the scores of the significantly underperforming perinatal units (approach 1). A further detailed comparison of the items within domains provides clues to the precise activities underlying underperformance. We illustrate this process using the Communication-domain. We used the same dataset as used in this paper.

Figure 1. ReproQ domain scores of the single best practice and the four worst performing units during childbirth.



Figures 1 and 2 illustrate the profiling of poor performers. Figure 1 displays the domain scores of the best performing and four underperforming perinatal units, using the total score during delivery. The scores of underperforming units were consistently lower than those of the best performing units in all domains, except for the Autonomy-domain. In three of the four worst performing units the major cause of low total performance was a low score on the Communication domain.

Figure 2. ReproQ item scores of the Communication-domain of the single best practice and the four worst performing units during childbirth.

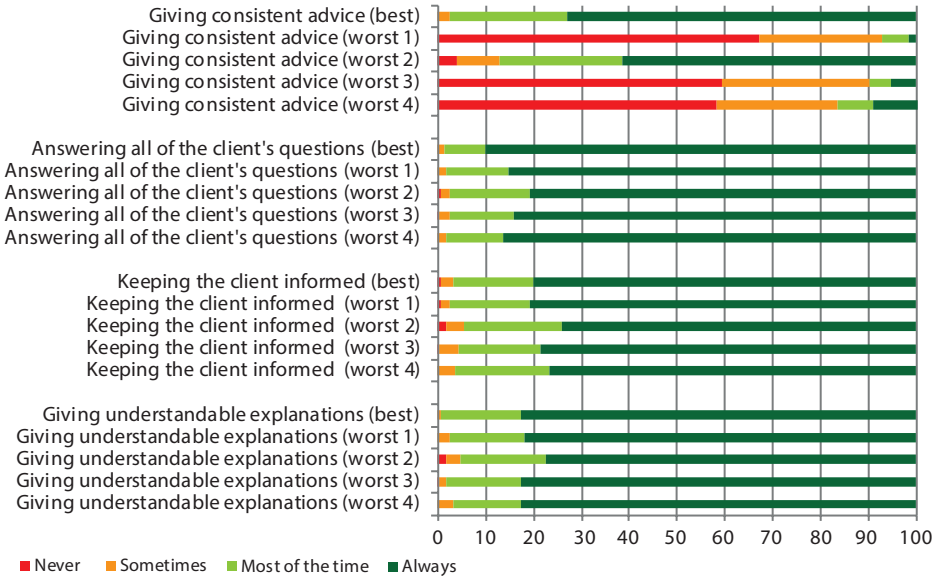


Figure 2 profiles the items scores of the identified low performing units, by disaggregating the Communication domain into its item scores. The low Communication domain score was predominantly associated with one specific item: 'giving consistent advice'. Next, interventions should be developed to improve clients' experiences with 'giving consistent advice', which may start with getting more information on the professional approaches and protocols which are into place. This example shows that ReproQ is not only suitable for benchmarking, but is also able to provide useful disaggregated information for quality improvement[5-7].

Stage 2 of the quality cycle as illustrated here can only be successful if the following conditions are met. Firstly, best and poor performing units should be willing to share information. Best practices may provide others a competitive edge[8]. Incentives must stimulate the eagerness of perinatal units to learn from others and adapt routines. Secondly, while measurement costs are low, administrative and financial support should be supplied. Finally, notwithstanding the relevance of client experiences, the performance in terms of medical outcomes should also be taken into account, which may influence priority setting of improvement measures[9,10].

REFERENCES

1. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
2. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
3. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
4. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
5. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring Client Experiences in Maternity Care under Change: Development of a Questionnaire Based on the WHO Responsiveness Model. *PLoS One.* 2015;10(2):e0117031.
6. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ.* 2016;4:e2092.
7. Scheerhagen M, Birnie E, Franx A, Van Stel HF, Bonsel GJ. What determines women's birth experiences? Applications for a benchmark. 2017.
8. Collins-Fulea C, Mohr JJ, Tillett J. Improving midwifery practice: the American College of Nurse-Midwives' benchmarking project. *J Midwifery Womens Health.* 2005;50(6):461-471.
9. Darby C, Valentine NB, Murray CJL, de Silva A. *Wolrd health organization (WH): strategy on measuring responsiveness.* Geneva: World Health Organization,;2000.
10. Murray CJL, Frenk L. Summary measures of population health in the context of the WHO framework for health system performance assessment. In: Murray CJL, Evans DB, eds. *Health Systems Performance Assessment*. Geneva, Swiss: World Health Organisation,; 2002.

TECHNICAL SUPPLEMENT

Differences in summary ReproQ scores across perinatal units, the dependent variable, may originate from three sources: 1) 'true' differences across perinatal units, 2) differences in client characteristics across perinatal units, 3) residual variance. Given the hierarchical data structure (perinatal units and clients within perinatal units), existing differences in client characteristics across perinatal units may obscure the estimation of the 'true' difference across units. When data are hierarchically ordered, multilevel analysis rather than conventional ordinary linear regression analysis is the appropriate analytical framework[1]. Multi-level analysis decomposes total data variance into variance attributable to perinatal units (source 1) and variance attributable to other sources (sources 2+3), in particular variance related to client characteristics. Estimation of the 'true' difference between perinatal units (source 1) requires that the domain and summary scores should be corrected for the other variance components (typically client characteristics) that bias and limit the comparison of perinatal units (i.e. the case mix correction). The need for case mix correction may be judged from the extent to which the clients' characteristics are related to the domain and summary scores of the perinatal unit they received care in, using Intraclass Correlation Coefficients (ICCs). The comprehensive analysis of variance components and ICCs is called a generalizability or G-study[2]. The methodology has been described and applied in other client experiences measurement studies as well[3,4]. A key result of multilevel analysis is the point estimate (and 95% confidence interval [CIs]) of the 'true' perinatal unit effect (one estimate for each perinatal unit) corrected for case mix differences. Note that the size of the 95%CIs differs across perinatal units, reflecting a) different numbers of clients across perinatal units, and b) different degrees of heterogeneity of client characteristics across perinatal units.

Case mix correction

The conventional procedure to determine the need for *case mix correction* is to compare the results of two multilevel regression models: 1) an 'empty' model with the ReproQ domain or summary scores as dependent variable, and solely a random intercept for each perinatal unit; and 2) an corrected model, with the ReproQ domain or summary score as dependent variable, a random intercept for each perinatal unit, and client characteristics included as explanatory variables[3,4].

We avoided overfitting and overcorrection, by limiting correction to predefined candidate factors with a demonstrated effect. Of the available client characteristics (age, educational level, ethnicity, parity, and self-rated health) only age, educational level, and self-rated health contributed significantly ($p<0.05$) to all domain and summary scores, and were therefore were included in the case mix correction. We also tested for random slopes, but none of these were significant and therefore remain unreported.

Discriminative power: Two approaches

In our study design, groups of individual clients' scores are associated with (nested within) perinatal units. Although some of the variance in clients' scores can be attributed to individual experiences, some of the variance in clients' scores is likely to be attributable to perinatal unit, with some perinatal units performing better than others [Roberts, 2014]. In our analyses we corrected for this nested structure.

Next, we used two complementary approaches to determine discriminative power.

Approach 1. Multilevel testing of the deviation of unit means from overall mean

Multi-level analyses were used to examine to what degree ReproQ is able to identify units that significantly perform above and below average (averaged over perinatal units), producing three parts of information: 1) the estimation of variance components and ICCs; 2) the estimation of the G-coefficient; and 3) the resulting estimation of the number of clients needed per perinatal unit to achieve sufficient reliability (D-Study). The conventional mode of presentation is the so-called caterpillar-plot.

In the first part, the output of the multilevel analysis in terms of variance components (explanatory variables) is used, and intra-class correlation coefficients (ICCs) are computed using these quantities. Here, the ICC of interest is the ratio of the variance in perinatal unit and the variance in client's characteristics in that unit[2,5]. An ICC close to zero implies that the client's experience is unrelated to the perinatal unit in which one receives care. In that case benchmarking of perinatal units with ReproQ would be of no value. In contrast when the ICC is close to 1, the perinatal unit is of decisive importance. The ICC-calculations were performed on the empty model as well as the corrected model (see before). The same output is used to obtain uncorrected and corrected means per perinatal unit, with their 95% confidence intervals. Best and poor practices are identified by the deviation of the 95% CI of each individual perinatal unit from the grand mean of all perinatal units (above or below the grand mean). Unit mean, their CIs, and the grand mean are conventionally plotted as caterpillar plots, ranking the perinatal units according to each unit's deviation from the grand mean domain or summary score (see Figure 1 for an example). Note that the grand mean is a common, yet arbitrary reference or norm.

In the second part, the reliability of the perinatal unit effect is established, given the realized numbers of clients; this is called the G-coefficient[2]. The G-coefficient expresses the proportion of variance in perinatal unit-level mean scores attributable to 'true' variation among perinatal units. The G refers to 'generalizability theory'[2,6]. Generalizability theory is a conventional base to study nested psychological data, like here. It assumes that any measurement is subject to multiple sources of error variance, such as client characteristics. A G-coefficient of one (the highest possible score) implies that all variance in domain and summary scores across perinatal units is attributed to the perinatal unit, and no variance can be attributed to other sources.

The third part, calculated for the average perinatal unit, and for a predefined level of reliability (more certainty requires more respondents) the minimum number of clients per perinatal unit to achieve significance. Small numbers of clients and large heterogeneity of client experiences produce wide confidence intervals, where only the first can be influenced. We deliberately selected units with at least 50 clients, but sample size still differed considerably across units as did the heterogeneity of their experiences. The aim of a D-Study (D from 'decision') is to estimate the number of included clients that is needed to achieve for any perinatal unit a predefined level of reliability, usually 0.80[2,6]. Stated otherwise, a D-Study estimates the minimal number of clients needed to create sufficient discriminative power for benchmark purposes, by using the variances derived from the G-study.

Approach 2. Relevance deviation based on MID

This approach judges ReproQ's discriminative power on the basis of the ability to proof relevant differences in domain or summary scores at the perinatal unit level; we introduced the MID for this purpose. We were aware that a MID at the individual level is not the same as the MID at the unit level (see below).

We previously determined, at the individual level, the minimally important difference (MID) for both the summary score and the individual domain scores, for the childbirth phase only[7]. The distribution-based and anchor point based MID had comparable size. For this paper, we derived MIDs for the late pregnancy and postnatal period in a similar way, see Table 2.

While a MID of 1 unit by definition is relevant if individual client's scores are compared with a norm or with each other, 1 MID difference at the perinatal unit level reflects a large difference: it means that *all* clients cared for in that unit, on average differ 1 MID from a reference value, either being much better (best practice) or much worse (poor practice). Compare birth weight, where a 200 gram difference of an individual from the gestational age norm is judged trivial, while an average 200 gram difference represents an important difference at the group level. We therefore added as a second criterion a difference of 0.5 MID unit, which still represents a stringent relevance criterion at the organisational level. As a norm to compare unit averages (summary score, domain scores) with in terms of the MID, we used the pooled average domain and summary scores of the 10% best performing units.

Statistical software

We computed all multi-level analyses and ICCs with R version 3.2.3[8], using the ggPlot program for graphical display of output, in particular an adapted version of the ggCaterpillar function for creating the caterpillar plots[9]. G-coefficients and D-sample sizes were estimated using G_string_IV, using the so-called 'unbalanced, nested one facet-design' (design 1.3) with perinatal unit as cluster [nesting variable] and client as facet[10]. (Details can be obtained from the authors.) For other statistical analyses we used SPSS 20.0.

REFERENCES

1. Twisk JWR. *Applied Multilevel Analysis.* Cambridge University Press; 2014.
2. Streiner DL, Norman GR, Cairney J. *Health measurement scales.* Oxford: University Press; 2014.
3. Bos N, Sturms LM, Stellato RK, Schrijvers AJ, van Stel HF. The Consumer Quality Index in an accident and emergency department: internal consistency, validity and discriminative capacity. *Health Expect.* 2015;18(5):1426-1438.
4. Stubbe JH, Brouwer W, Delnoij DM. Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire. *BMC Ophthalmol.* 2007;7:14.
5. Streiner DL, Norman GR. *Health measurement scales.* Oxford: University Press; 2008.
6. Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. *American Psychologist.* 1989;44(6):922-932.
7. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ.* 2016;4:e2092.
8. R Core Team. R: A Language and Environment for Statistical Computing. 2015. http://www.R-project.org.
9. StackOverflow. In R plotting random effects. 2013; https://stackoverflow.com/questions/13847936/in-r-plotting-random-effects-from-lmer-lme4-package-using-qqmath-or-dotplot.
10. *G String IV* [computer program]. Hamilto, Canada2015.

Part 3

Implementation and application

6

Measuring Clients' Experiences with Antenatal Care Before or After Childbirth: it Matters

M. Scheerhagen, E. Birnie, A. Franx, H.F. van Stel, G.J. Bonsel

PeerJ. 2018; 6, e5851

ABSTRACT

Background. When clients' experiences with maternity care are measured for quality improvement, surveys are administered once, usually six weeks or more after childbirth. Most surveys conveniently cover pregnancy, childbirth and postnatal care all in one. However, the validity of measuring the experiences during pregnancy (antenatal experiences) after childbirth is unknown. We explored the relation between the measurement of antenatal experiences late in pregnancy but prior to childbirth ('test' or gold standard) and its retrospective measurement after childbirth (retrospective test). Additionally, we explored the role of modifying determinants that explained the gap between these two measurements.

Methods and findings. Client's experiences were measured by the ReproQuestionnaire that consists of an antenatal and postnatal version, and covers the eight WHO Responsiveness domains. 462 clients responded to the antenatal and postnatal questionnaire, and additionally filled out the repeated survey on antenatal experiences after childbirth. First, we determined the association between the test and retrospective test using three scoring models: mean score, equal or above the median score and having a negative experience. The association was moderate for having any negative experience (absolute agreement=68%), for the median (absolute agreement=69%) and for the mean score (ICC=0.59). Multiple linear and logistic regression analysis for all three scoring models revealed systematic modifiers. The gap between antenatal and postnatal measurement was (partly) associated with clients' experiences during childbirth and postnatal care and by professional discontinuity during childbirth but unrelated to the perceived health outcome.

Conclusions. The antenatal experiences should be measured before and not after childbirth, as the association between the antenatal experiences measured before and after childbirth is moderate.

INTRODUCTION

Clients' experiences with care are considered to be an important independent indicator of health care performance[1,2]. Being relevant for its own sake, clients' experiences could also affect health outcome through several pathways[3-6]. For example, clients who truly understand the explanation of their caregiver are more likely to comply to treatment or lifestyle change.

As clients' experiences are an independent indicator of performance, clients' experiences are systematically measured using surveys, usually held after the care-episode. Such measurements could help to identify areas for improvement[7,8]. Targets of quality improvement are found by identifying health care organizations or areas with below average scores or single negative outliers on questions representing the characteristics of service delivery, e.g. communication and prompt access to services. Next, the organization develops and implements a plan to meet these goals, and verifies if the goals are met[9-12].

Clients' experiences in maternity care are routinely measured in several countries. Data on clients' experiences are usually collected through surveys administered six weeks or more after childbirth. Most surveys cover pregnancy, childbirth and postnatal care in one measurement[13-17]. As these surveys cover almost about 9 months of care, with different health care professionals, settings and possibly events, measurement of client's experiences bears the risk of being vulnerable to memory failure and/or changes in perception due to modifying intercurrent events that happened since the antenatal experiences. Assuming the antenatal measurement of such experiences to be the gold standard, the question is whether the response on the postnatal survey shows random and/or systematic error. Stated otherwise, when the clients' experiences are measured before childbirth and repeated after childbirth, does this lead to the same clients' experience scores? Ideally, valid measurement of antenatal experiences postnatally should not be systematically affected by the care process, experiences or outcomes that occur *after* antenatal measurement. Despite the widespread practice of a one-stage postnatal measurement, to our knowledge this question has never been explored. If random error is considerable or systematic shifts are present, the convenient one-stage measurement perhaps should be replaced by a two-stage measurement procedure, that includes the measurement of clients' experiences not only after childbirth but also antenatally.

We explored the presence of memory effects in the measurement of clients' experiences in maternity care using the ReproQuestionnaire (ReproQ). ReproQ is the national survey for client experience measurement in childbirth care in the Netherlands. It was especially designed for a two-stage measurement procedure, consisting of antenatal and postnatal versions. ReproQ was extensively validated (n>18,000)[18,19] and is currently regarded as one of the national maternity care indicators[20].

METHODS

ReproQuestionnaire

The ReproQ consists of two versions, each covering the experiences of two reference periods. The antenatal version covers the experiences during early and late pregnancy; the postnatal version covers the experiences during childbirth and postnatal care. Both versions are identical, in the sense that the same type of experiences is asked for, but items (questions) are contextually adapted. Altogether, a client is invited to judge a typical item for four consecutive periods.

The conceptual basis of the ReproQ was the WHO responsiveness model[1,2]. The WHO developed this universally applicable concept that consists of four domains on the interactions of the client with the health professional (Dignity, Autonomy, Confidentiality, and Communication), and of four domains on the client orientation of the organizational setting (Prompt attention, Access to family and community support, Quality of Basic amenities, and Choice and continuity of care)[1,2]. The response mode of all the experience items uniformly consists of four categories: "never", "sometimes", "often", and "always", with a numerical range of 1 (worst) to 4 (best).

Additional sections of the ReproQ address the client's socio-demographic characteristics, details about the care process during pregnancy and childbirth, and maternal and infant health outcomes in non-medical terms as perceived by the mother. We also added a relevance question on which two out of eight domains were most important to the client.

Previous psychometric analyses showed that content and construct validity were good, as was the test-retest reliability of the experience during childbirth. Full details of the development and the psychometric properties of the questionnaire are described elsewhere[18,19].

Design, ReproQ scoring models, outcomes

The Medical Ethical Review Board, Erasmus Medical Center, Rotterdam, the Netherlands, approved the study protocol (study number MEC-2013-455).

The study was designed as a cohort study with three measurements. First, women received an invitation to fill out the antenatal ReproQ around a gestational age of 34 weeks. This is called 'test'. Second, women received an invitation to fill out the postnatal ReproQ six weeks after the expected date of childbirth. Non-responding women received a reminder two weeks after invitation to the antenatal and postnatal questionnaire. Third, we invited women who responded to the antenatal and postnatal ReproQ again to fill out the antenatal experiences after childbirth. This is called the 'retrospective test'. We sent the retrospective test at least 14 days after women filled out the postnatal ReproQ.

Three different scoring models exist to summarize clients' experiences and to monitor adverse outcomes at the individual or aggregate level. The three models may be applied to an individual item, to an individual domain (called domain score), to two summary scores

of the four personal and four setting domains (called personal and setting score), or to a summary score of all domains (called total score).

Table 1 displays the scoring models and their definitions. The first model creates a dichotomous variable (called 'negative score') at the client level, reflecting the presence of any so-called negative experience. As Table 1 shows, the definition of a 'negative' experience is based in part on the two domains that a client identifies as most important, thereby creating a personalized score. Since the likelihood of a negative experience partially depends on the number of items per domain, absolute percentages of negative scores cannot be compared across domains. The negative score model assumes that, for the individual client or for an organisation, a negative experience cannot be compensated by very good experiences on other items or domains. This is contrary to the mean score where good experiences can compensate poor experiences.

The second scoring model computes a continuous mean score (called 'mean score', range 1.0 – 4.0) at the client level, for each domain or group of domains separately. The total, personal and setting summary scores are not the mean of all items involved in the domains, but the unweighted mean of the mean domain scores involved in that summary measure. For the calculation of the summary scores, each domain has the same weight, even if the domains rest on a different numbers of items.

Finally, the third model creates a dichotomous variable at the client level reflecting whether her mean item, domain or summary score is equal to/above *or below* the median *of the distribution* of the respective item, domain or summary scores *of all cases* (called 'median score'). The 'median score' model was added because of the skewed distributions of clients' experience scores.

Data collection

ReproQ data were obtained from two sources: 10 perinatal units (a hospital with its associated community midwife practices) and two maternity care organizations. These organizations deliver postnatal care at home from childbirth onwards over a period of seven to 10 days. Women can register and apply for this service during pregnancy. For perinatal units, clients were invited to participate by their caregiver, who asked for consent. For maternity care organizations, all women were invited to fill out the client experience questionnaire, after consent was ticked.

Data were collected in two periods. In the first period (October 2013 to January 2015), data was collected with the antenatal ('test') and postnatal ReproQ. There were no restrictions to invite women to fill out the antenatal and postnatal ReproQ; all women could participate provided that informed consent was signed or ticked. The second period, December 2014, administered the data of the retrospective test. Women were excluded from participation of the retrospective test for the following reasons: 1) women did not respond to the antenatal

Table 1. Scoring models, outcome measures and measures of association.

Scoring model	Definition	Outcome measure	Measure of association		
			For summary and domain scores	For item scores	In regression analysis
Negative score	Ticking the category 'never' in at least one of the items of a domain (indicating a very poor experience), and/or filling out 'sometimes' in at least one of the items of the two domains that the client identified as most important	Dichotomous	AA	AA	OR
Mean score	The unweighted average score of items within a domain, treating the item response categories numerically; the total, personal and setting summary scores equal the mean of the mean domain scores involved in that summary measure	Mean (SD)	ICC	AA	β
Median score	Whether the client's mean item, domain or summary score is equal to/above or below the median of the distribution of the respective item, domain or summary scores of all cases	Dichotomous	AA	AA	OR

and postnatal questionnaires, 2) women filled out less than 50% of the antenatal and/or postnatal experience score, or 3) they filled out the questionnaires on paper. (This was done for the reasons of data management efficiency; n=166). Women were excluded from analyses if they filled out less than 50% of items of the retrospective test questionnaire, or if women filled out the retrospective test over 1.5 years after childbirth. The latter criterion excluded women who could be pregnant again.

Measures of agreement

In this study we used two dichotomous scores and one continuous score for the domain and summary scores, with two different agreement statistics. For the negative and median scores, we used the percentage absolute agreement (AA), classified as 'excellent' (90% – 100%), 'good' (75% – 89%), 'moderate' (60% – 74%), or 'poor' (<60%)[21]. For the mean score, we used the Intraclass Correlation Coefficient (ICC) as measure of agreement (two way mixed model, absolute agreement, single measure), and classified the estimated ICCs as: 'excellent' (≥.81), 'good' (.61 – .80), 'moderate' (.41 – .60), 'poor' (≤.40)[21].

For the individual items, agreement between the test and retrospective test was quantified as the percentage absolute agreement.

Data analysis

Figure 1 shows the analytic framework. All analyses were performed on the reported experience of the second half of the pregnancy, because in psychometric analysis the experiences during first and second half of pregnancy are highly associated (AA_{Neg}=91.6%; AA_{MD}=85.9%; ICC=0.83). The late antenatal experiences were chosen as comparator ('test' or gold standard), because the second half of pregnancy covers more antenatal check-ups than the first half, and therefore thought to be more representative for the entire antenatal phase. Moreover, the timespan between the second half the pregnancy and the retrospective test is smaller than the timespan between early pregnancy and the retrospective test, and therefore the risk of memory effects is probably smaller.

We used all retrospective test data collected up to 1.5 years after childbirth (range: 3.5 month to 1.5 years after birth). The wide range had limited impact on the experience scores of the retrospective test and the association between the test and retrospective test; both slightly decreased over time.

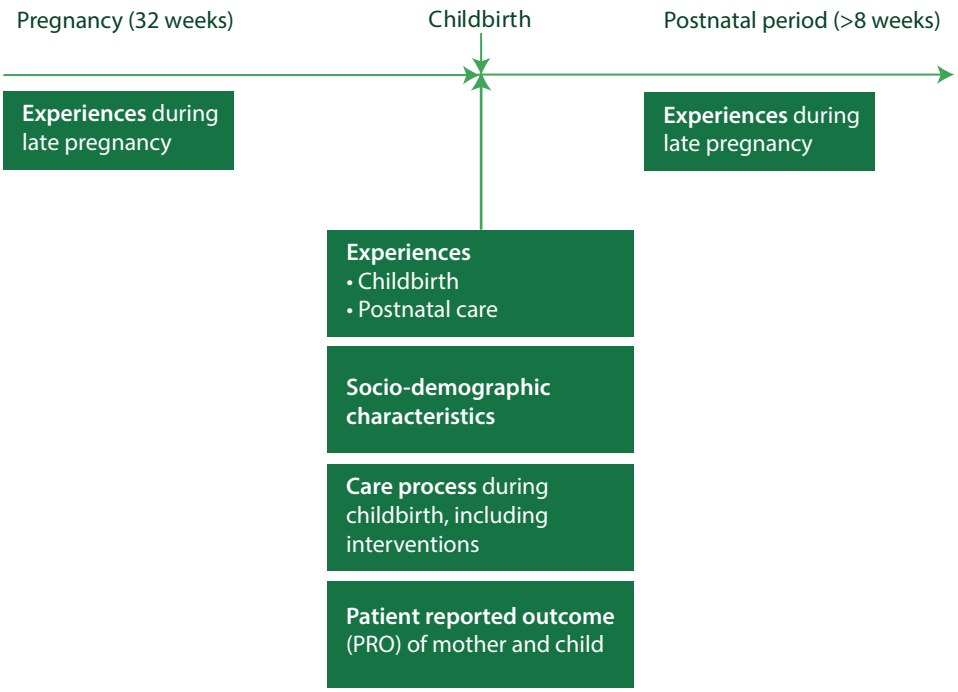
First we explored the crude agreement between the antenatal experiences measured before (test or gold standard) and after childbirth ('retrospective test'). For that purpose the three outcome measures were computed for a. the total score, b. the personal and setting summary scores, and c. the individual domain scores, and subsequently the agreement of the gold standard and retrospective test was calculated. The agreement of the individual items between the before (gold standard) and after childbirth (retrospective test) measurement was calculated. While the domain and summary measures were calculated conventionally,

for the individual item analyses, we split the 'no-agreement' category into "test better experience than retrospective test" and "test worse than retrospective test".

Second, we explored the effects of background characteristics and systematic effects of intercurrent events, as determinants of the antenatal total experience score as measured after childbirth. For the negative and median score models, we used multiple binary logistic regression analysis. For the continuous mean score model, we applied multiple linear regression analysis. Dependent variable was the antenatal total experience score as measured after childbirth; independent variables were the antenatal total experience score as measured before childbirth (gold standard score) and a set of potentially modifying factors. The following sets of determinants were included in the regression model (enter method): socio-demographic characteristics, previous experiences with care (antenatal, childbirth and postnatal care), characteristics of the care process during pregnancy and childbirth including interventions during childbirth, and perceived health outcomes of mother and child.

Considering the abundance of possible determinants and limited sample size, we included in the multivariable analyses only those that were determinants of clients' experiences

Figure 1. Framework of analyses to determine the association of the antenatal experiences measured during pregnancy and after childbirth.



during birth[22]. A determinant was overall judged as significant if the estimated adjusted beta- or OR-coefficient was statistically significant ($p<0.05$, two-sided) in at least two of these analyses, a conservative approach.

For the binary logistic regression analysis, the goodness of fit was assessed using the proportion of correct predictions. For linear regression we used the adjusted R^2.

RESULTS

Figure 2 shows the flow diagram. We invited 3,313 women for the retrospective test, of whom 1,091 women responded (33%). Of these, 629 women were excluded from analysis. The remaining 462 women were included.

Figure 2. Flow diagram of study.

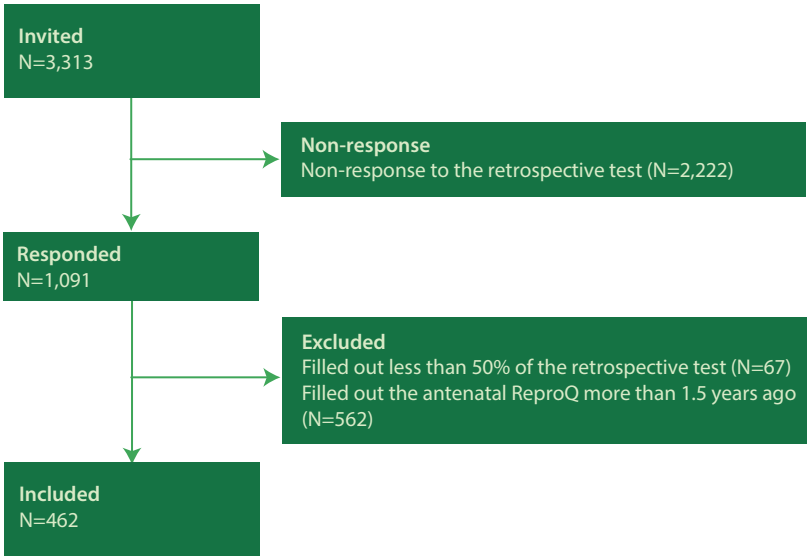


Table 2 presents the characteristics of the included women (n=462). Mean age was 32 years (SD=4.8). Half of the women gave childbirth for the first time. 26 (6%) women were of non-Western background; and 14 (3%) women reported to have a low educational level. 241 (52%) women reported not to know the health care professional who supervised their delivery. 70 (16%) women were referred to secondary care during their pregnancy and 144 (32%) were referred during parturition. 84 (18%) women reported that they felt unhealthy and that they were hospitalized after childbirth. Additionally, 59 women (13%) perceived their babies' health as unhealthy and reported that their babies were hospitalized.

Table 2. Characteristics of women who filled out both the test and retrospective test (n=462)[$].

	N	%
Socio demographic characteristics		
Age		
≤24	13	3
25 – 29	130	28
30 – 34	185	40
≥35	130	28
Parity		
Primiparous	229	50
Multiparous	233	50
Ethnic background		
Western	435	94
Non – Western	26	6
Educational level		
Low	14	3
Middle	135	29
High	312	68
Marital status		
Married/living together	447	97
Not living together or no relationship	14	3
Planned pregnancy		
Yes	421	91
No	41	9
Care process		
Professional continuity		
Yes	220	48
No	241	52
Setting continuity		
No referral	238	53
Referral to secondary care during pregnancy	70	16
Referral to secondary care during parturition	144	32
Realization of the expected place of childbirth		
Yes	263	58
No	182	40
No prior expectations	11	2

Table 2. *Continued*

	N	%
Intervention		
Induced labor		
No	355	78
Yes	103	23
Mode of childbirth		
None	270	58
Episiotomy	81	18
Vacuum or forceps extraction	46	10
Cesarean	65	14
Patient reported outcome		
Baby		
Healthy and not hospitalized	315	68
Healthy, but hospitalized	60	13
Unhealthy, but not hospitalized	28	6
Unhealthy and hospitalized	59	13
Mother		
Healthy and not hospitalized	245	53
Healthy, but hospitalized	27	6
Unhealthy, but not hospitalized	106	23
Unhealthy and hospitalized	84	18

[5] The percentage of missing data was below 3% for all characteristics.

Table 3 shows the crude agreement between the antenatal experiences measured before and after childbirth for the summary and domain scores. For the total score, 35% of the women reported one or more negative experiences filling out the 'test', and 33% when filling out the retrospective test. The absolute test-retrospective test agreement (AA) of 'having a negative experience' was 67.5% (CI: 63.0 – 71.8%). The absolute test-retrospective test agreement (AA) of 'a score above the median' was 69.6% (CI: 65.2 – 73.8%). The ICC of the total experience scores ($mean_{test}$=3.77; $mean_{retrospective\ test}$=3.69) was 0.59. The negative, median and mean score models all indicated a moderate association. The associations of the personal and setting scores were comparable for the negative and median score models, but the association for the mean personal score was weaker then for the mean setting score (ICC 0.49 vs. 0.59).

All individual domains showed a good to excellent association for having a negative experience. For the median and mean scores, all domain associations were moderate, except for Confidentiality, which had an ICC of 0.27, indicating a poor association.

Table 3. The association between the late antenatal experiences measured during pregnancy and after childbirth, expressed as having a negative experience, below the median score and mean score (n=462).

	Negative experience score #			Median experience score $			Mean experience score				
	test (%)	retrospective test (%)	Absolute agreement (AA) (%)	test (%)	retrospective test (%)	Absolute agreement (AA) (%)	Mean test	SD test	Mean retrospective test	SD retrospective test	ICC
Total score	**35.1%**	**32.5%**	**67.5%**	**60.4%**	**47.7%**	**69.6%**	**3.77**	**0.23**	**3.69**	**0.29**	**0.59**
Personal score	**22.1%**	**19.5%**	**75.8%**	**74.6%**	**59.0%**	**70.5%**	**3.81**	**0.23**	**3.72**	**0.29**	**0.49**
Setting score	**18.8%**	**19.9%**	**76.8%**	**50.2%**	**44.0%**	**69.6%**	**3.73**	**0.28**	**3.66**	**0.33**	**0.59**
Dignity	2.6%	3.5%	96.1%	74.0%	58.7%	69.9%	3.89	0.24	3.81	0.31	0.42
Autonomy	19.9%	15.6%	77.5%	75.1%	86.0%	74.0%	3.64	0.42	3.61	0.45	0.42
Confidentiality	0.4%	1.1%	98.5%	88.1%	76.6%	76.8%	3.91	0.26	3.82	0.38	0.27
Communication	1.9%	3.0%	96.8%	50.6%	42.7%	71.4%	3.78	0.29	3.69	0.38	0.41
Prompt attention	3.5%	5.4%	94.2%	53.5%	43.0%	69.2%	3.68	0.31	3.59	0.37	0.52
Social considerations	1.1%	1.5%	97.8%	67.3%	64.6%	71.9%	3.79	0.35	3.76	0.41	0.46
Basic amenities	2.2%	1.9%	96.8%	70.1%	59.1%	69.6%	3.83	0.32	3.74	0.39	0.48
Choice and continuity	13.6%	13.9%	80.7%	53.5%	47.4%	69.2%	3.61	0.54	3.54	0.58	0.49

Having a negative experience (never in an domain and/or 'sometimes' in the individually chosen two most important domains).
$ Above the median

Table 4. Level of absolute agreement between the items measured during pregnancy and after childbirth (n=462).

Itemscore	Negative experience score #			Median experience score $			Mean experience score		
	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test
Dignity									
Respecting privacy	99.6	0.4	0.0	87.4	10.6	1.9	87.1	10.7	2.2
Treating with respect	99.6	0.2	0.2	90.0	7.8	2.2	89.7	8.1	2.2
Giving personal attention	97.6	1.1	1.3	81.8	12.6	5.6	80.5	13.4	6.1
Treating with kindness	98.9	0.6	0.4	87.0	8.7	4.3	86.3	9.0	4.6
Considering your wishes and customs	97.6	1.7	0.6	77.5	16.0	6.5	74.8	18.0	7.2
Trustworthy as health professional	98.3	1.1	0.6	75.5	16.9	7.6	74.0	17.9	8.1
Autonomy									
Refuse treatment	96.5	0.6	2.8	74.2	15.8	10.0	69.9	17.4	12.7
Involved in decision-making	98.1	1.5	0.4	73.2	16.9	10.0	71.0	17.5	11.6
Consent screening	95.5	2.8	1.7	95.5	2.8	1.7	95.8	3.1	1.2
Birthplan	83.3	6.1	10.6	65.6	17.7	16.7	56.6	25.1	18.3
Confidentiality									
Handeling your medical details and records	100.0	0.0	0.0	85.5	9.1	5.4	85.1	9.6	5.3
Secured provision of medical information to others	98.9	0.9	0.2	82.0	14.1	3.9	80.9	15.2	3.9
Communication									
Responsive to client questions	99.6	0.4	0.0	83.1	12.6	4.3	82.4	13.2	4.3
Consistency of advice across professionals	97.8	1.7	0.4	68.6	20.6	10.8	62.7	24.3	13.0
Comprehensibility of explanation	99.6	0.2	0.2	82.7	11.5	5.8	81.9	12.2	5.9
Provision of information while treated	98.5	0.9	0.6	74.5	16.5	9.1	72.7	17.8	9.5

Table 4. *Continued*

Itemscore	Negative experience score #			Median experience score $			Mean experience score		
	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test
Prompt attention									
Access for appointment/contact in urgent situations	100.0	0.0	0.0	87.4	8.0	4.5	83.9	9.2	7.0
Access for appointment/contact without urgency	98.5	1.1	0.4	66.9	21.4	11.7	62.5	23.7	13.8
Time from health care professional when requested	99.6	0.4	0.0	77.3	15.8	6.9	75.6	17.4	7.0
Waiting time for service	95.2	2.8	1.9	85.9	10.0	4.1	57.7	25.7	16.6
Setting within reach	99.4	0.4	0.2	82.3	11.3	6.5	81.6	11.6	6.8
Prompt phone response of health professional	99.6	0.4	0.0	76.0	16.2	7.8	74.2	17.7	8.1
Social considerations									
Involvement of the partner in care provision	98.7	0.9	0.4	77.7	13.4	8.9	74.1	15.1	10.8
Taking into account of family duties	99.4	0.2	0.4	78.6	11.9	9.5	75.3	13.3	11.4
Feeling supported by your family	99.4	0.4	0.2	87.7	6.1	6.3	85.8	7.2	7.0
Basic amenities									
Comfort of setting	97.4	2.6	0.0	71.0	19.0	10.0	66.7	22.2	11.1
Hygiene of setting	99.1	0.6	0.2	84.0	11.3	4.8	82.6	12.4	5.0
Accessibilty of setting	99.6	0.2	0.2	88.7	7.4	3.9	88.3	7.6	4.1

Table 4. *Continued*

Itemscore	Negative experience score #			Median experience score $			Mean experience score		
	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test	Test=retro-spective test	Test>retro-spective test	Test<retro-spective test
Choice and continuity									
Continuity of care provision when change of individual professional (same discipline)	99.1	0.2	0.6	69.9	19.5	10.6	67.8	20.7	11.5
Continuity of care provision when change of professional (across disciplines)	97.8	1.5	0.6	73.2	18.8	8.0	55.2	26.4	18.4
Allowance for selecting a preferred type of health professional	81.6	8.4	10.0	73.8	14.5	11.7	66.8	17.9	15.3
Being clear who was in charge of your care	97.0	1.7	1.3	79.4	12.1	8.4	72.0	16.3	11.7

Having a negative experience (never in an domain and/or 'sometimes' in the individually chosen 2 most important domains).
$ Above the median

Table 5. Impact of antenatal, childbirth and postnatal experiences with care and other determinants on the total antenatal experience score measured after childbirth, according to three scoring models (n=462).

Goodness of fit	Overall sign ^	Negative experience score #			Median experience score $			Mean experience score		
				71%			73%			70%
		OR	95% CI	p	OR	95% CI	p	β	95% CI	p
Socio demographic characteristics										
Ethnic background										
Western (ref)		1			1			0.00		
Non-Western		1.22	0.47 – 3.13	0.69	0.75	0.27 – 2.07	0.58	-0.03	-0.11 – 0.06	0.53
Educational level										
Low / middle		0.75	0.46 – 1.21	0.23	1.24	0.76 – 2.01	0.40	-0.02	-0.06 – 0.02	0.11
High (ref)		1			1			0.00		
Planned pregnancy										
Yes (ref)		1			1			0.00		
No		1.22	0.47 – 3.13	0.87	1.33	0.59 – 3.00	0.49	0.06	-0.01 – 0.13	0.12
Experiences with care										
Antenatal experience	*	3.08	1.95 – 4.88	<0.01	3.94	2.51 – 6.19	<0.01	0.62	0.52 – 0.71	<0.01
Childbirth experience	*	2.07	1.32 – 3.26	<0.01	1.89	1.16 – 3.08	0.01	0.27	0.17 – 0.38	<0.01
Postnatal experience	*	1.45	0.89 – 2.37	0.14	2.17	1.35 – 3.49	<0.01	0.14	0.05 – 0.23	<0.01
Care process										
Professional continuity										
Yes (ref)		1			1			0.00		
No	*	1.60	0.99 – 2.60	0.06	0.50	0.31 – 0.82	0.01	-0.05	-0.09 – 0.00	0.04
Setting continuity										
No referral (ref)		1			1			0.00		
Referral during pregnancy		0.91	0.47 – 1.76	0.77	1.05	0.51 – 2.14	0.89	0.00	-0.06 – 0.06	0.97
Referral during birth		1.16	0.61 – 2.23	0.65	0.86	0.43 – 1.70	0.66	-0.02	-0.08 – 0.04	0.79

Table 5. *Continued*

Goodness of fit	Overall sign ^	Negative experience score #			Median experience score $			Mean experience score		
				71%			73%			70%
		OR	95% CI	p	OR	95% CI	p	β	95% CI	p
Expected place of birth was realized										
Yes (ref)		1			1			0.00		
No / no prior expectation		0.93	0.52 – 1.64	0.79	2.09	1.15 – 3.78	0.02	0.03	-0.02 – 0.08	0.23
Intervention										
Induced labor										
No (ref)		1			1			0.00		
Yes		1.50	0.88 – 2.55	0.14	0.77	0.44 – 1.35	0.37	0.02	-0.03 – 0.07	0.33
Intervention										
No (ref)		1			1			0.00		
Yes		1.73	1.05 – 2.87	0.03	0.85	0.51 – 1.43	0.54	0.03	-0.02 – 0.07	0.58
Perceived (patient reported) outcome										
Outcome baby										
Healthy and not hospitalized (ref)		1			1			0.00		
Unhealthy and/or hospitalized		0.98	0.59 – 1.60	0.92	0.87	0.52 – 1.46	0.60	0.00	-0.05 – 0.05	0.99
Outcome mother										
Healthy and not hospitalized (ref)		1			1			0.00		
Unhealthy and/or hospitalized		0.89	0.56 – 1.43	0.63	0.81	0.51 – 1.30	0.39	-0.03	-0.08 – 0.01	0.11
Constant		0.12			0.29			-0.20		

Having a negative experience (never in an domain and/or 'sometimes' in the individually chosen 2 most important domains).
$ Above the median
^ The determinant was of significant influence for at least two of the outcome measures

The item analyses showed good to excellent associations for having a negative experience (see Table 4). For the median score, the associations varied from excellent to moderate, except for 'Influence on childbirth plan' (AA=59.7%) which was poor. For the mean score, not only this item (AA=56.6%) but also 'Waiting time for service' (AA=57.7%) and 'Continuity of care provision when change of professional' (across disciplines) (AA=55.2%), had a poor association.

Table 4 also depicts the magnitude and direction of change between the before and after birth measurements. For the negative score, agreement was very high, indicating that scores were fairly stable between the test and retrospective test, with slightly more clients reporting negative scores at the test, the 'Birthplan' item being an exception. The median and mean scores showed more variability in scores between the test and retrospective test, with the overall trend of higher scores at the test.

Table 5 shows the results of the regression analyses. The experience score of the retrospective test were not significantly influenced by any of the socio-demographic characteristics. However, the retrospective test score was significantly associated with the women's antenatal, childbirth and postnatal experiences. Of the care process determinants, only professional continuity was relevant. Finally, the perceived maternal and infant health outcome had no significant influence on the retrospective test. Despite the different analyses and scoring models, the goodness of fit was comparable for the three measures (70 – 73%).

DISCUSSION

To determine the optimal timing of the collection of data on clients' antenatal experiences, we assessed the association between the antenatal experiences measured before and after childbirth for the summary, domain and item scores. The total score showed a moderate association, irrespective of the scoring model used. For the domain scores, the associations varied with the scoring model selected, being overall excellent for the negative score, and moderate for the median and mean scores. For the domains, agreement was quite uniform within the scoring model used. Confidentiality was the only domain with a poor association for the mean score. For the individual items, associations were particularly low for 'Influence on your childbirth plan', 'Waiting time for service', and 'Continuity of care provision when change of professional (across disciplines)'. Overall, the measurement of antenatal experiences after birth results in elevated variability of experiences across clients, with the overall trend that scores after birth are somewhat lower than before birth. Additionally, the gap between antenatal and postnatal measurement is (partly) associated with clients' experiences during childbirth and postnatal care and by professional discontinuity during childbirth, but it is unrelated to the perceived health outcome.

One key result is that the antenatal experience score measured after childbirth was only moderately associated with the antenatal experiences measured before childbirth, irrespective of the scoring model applied. In contrast, the personal, setting, domain and item scores were stronger associated for having a negative experience than for the median and mean scores. One explanation for this is that a negative experience lingers better in one's memory than an equally moderate or good experience, as shown in decision and judgment theory[23-25]. An alternative explanation is of a statistical nature: changes in experiences are less easy to capture using a dichotomous measure like the negative score, producing much more agreement between the test and the retrospective test. The same argument, however, does not apply to the dichotomous median score. For the negative score, the cut-off has a fixed definition and is therefore absolute. In contrast, the cut-off for the median score equals the median of the distribution of the summary and domain scores 'as observed', and is therefore a relative position. Furthermore, the odds of having a negative experience increases with the number of items, whereas the odds of having an experience score equal or above the median is independent from the number of items.

In the ideal situation, a strong association between the antenatal experiences measured before and after childbirth is expected and desired. Furthermore, valid measurement of antenatal experiences postnatally should not be systematically affected by the care process, experiences or outcomes that occur *after* antenatal measurement. However, our results strongly suggest the opposite: women's experiences with childbirth and postnatal care had a positive and systematic impact on the antenatal experiences measured postnatally.

One possibility is that women's response scales changed after birth. It is well known from research on judgment and decision[26] and response shift[27-29], that pre-treatment judgment scales may differ systematically from post-treatment scales with, in our case, childbirth as the so-called catalyst. A change of reference frame or internal standards of comparison might result in scale recalibration[26-29]. The change comparison process may be related not only to a change of status quo, but also to the change of women's affect and mood after childbirth[26]. Another possibility is that retrospective judgment of past experiences invokes the risk of memory errors. Recall bias, i.e. 'wrong' assessment post-hoc of a former outcome[30], may have occurred under the influence of childbirth and/or postnatal events or experiences. Another form of memory error, so-called hindsight bias (i.e. the influence of outcome knowledge on memory reconstruction, increasing the predictability of the outcome) is less likely as (favorable) childbirth and postnatal experiences contributed positively to the gap between antenatal and postnatal measurement instead of bridging it[31].

In the ideal situation, the gap between antenatal and postnatal measurement should be independent from the care process and intervention determinants. Overall, effect sizes of these variables were moderate to negligible and not significant. One exception to this is professional continuity during childbirth that was of significant impact on the antenatal experiences measured after childbirth. This is probably due, at least in part, to clients'

expectations: a new professional during childbirth is never as well informed about a client's wishes and customs as her attending professional during pregnancy, and trust between the new health care professional and the client is lacking. Even though the antenatal health care professional could (and should) inform a client that a transfer during childbirth is possible, clients may not feel prepared for a change of professional.

Surprisingly, the perceived health outcome of mother and child had no impact on the antenatal experiences measured after childbirth. This is in contrast with literature, which suggests that, in retrospect, when women after childbirth recollect their antenatal experiences, these experiences could adapt in the direction of the (perceived) health outcome during childbirth; i.e. hindsight bias[31-33]. One explanation is that hindsight bias did not occur in our case. Another explanation is that clients do not perceive a relationship between the health outcomes of birth and the experiences during pregnancy, as different services are provided, often by different health care professionals and often in different settings.

Another surprise is that none of the included socio-demographic determinants were significantly associated with the gap between the test and the retrospective test. This is contrary to the results of research on judgment and decision[26] and response shift[27-29]. Several explanations can be put forward. Firstly, contrary to Sprangers & Schwartz, a change of antenatal and postnatal scales (recalibration, with childbirth as the so-called catalyst) did not occur or the change was small or undetectable. Secondly, several studies suggest that the agreement between the test and retrospective test is similar between subgroups, even though the experiences are different[22,34-36]. Stated otherwise, the effect may have been cancelled within patients or even be unrelated to patient characteristics. Thirdly, the socio-demographic characteristics do not directly affect the experience scores but only exert an indirect effect, through influencing the clients' mechanisms to accommodate the change in her situation (here: childbirth)[27-29]. Consequently, the impact of socio-demographics may already be incorporated in the impact of previous experiences. Fourthly, our sample was too small to detect any impact of socio-economic status and ethnicity on the antenatal experiences measured after childbirth. However, that argument did not apply for marital status, maternal age and parity, which are socio-demographic characteristics that did not qualify for the multivariable analyses. Finally, we may have omitted relevant variables, e.g. personality traits or affect and mood[26,37].

Our study in maternity care is a specific case of a general problem - as such it provides a warning for similar studies. Measurement problems may occur when experiences with care are evaluated but adjacent care episodes are different in terms of disease course or severity or care provided (e.g., in terms of professionals involved, locations) and separated by a critical event which could serve as 'catalyst' (e.g., intervention, hospitalization, complication). A possible change of patient's pre- and post 'catalyst' response scales and the risk of memory errors when patient's experiences are measured afterwards may result

in reduced validity and/or reliability of measurements. To avoid these risks, we recommend that patient experiences with care to be measured within its own care episode.

Strengths & Limitations

One strength of this study is that, to our knowledge, this is the first study exploring the validity of clients' antenatal experiences measured after childbirth. Nevertheless, several limitations merit discussion. Firstly, women with a low educational level, non-Western women, women <24 years of age, and setting continuity (referral to secondary care) were slightly underrepresented compared to the national pregnancy population[38], despite considerable efforts to adapt the questionnaire and other measures taken to further the participation of these groups. Our results suggest, however, that these variables are all unrelated to the gap between the antenatal and postnatal measurements. Parity, induced labour, mode of delivery, and maternal and neonatal admission rates were comparable to the national average. National data on professional continuity are lacking, but data are comparable to one of our other studies (n=3,479 women[22]). Secondly, we did not register whether the clients' situation changed during the interval between test and retrospective test other than the events, experiences and perceptions during childbirth and postnatal care. It is possible that omitted variables could further modify the gap between test and retrospective test.

CONCLUSION

Clients' experiences during pregnancy, childbirth and postnatal care are often measured for quality improvement cycles. We recommend measuring the antenatal experiences in late pregnancy instead of after childbirth, as the agreement between the antenatal experiences measured before and after childbirth is overall moderate for the summary scores.

The gap between antenatal and postnatal measurement is (partly) associated with clients' experiences during childbirth and postnatal care and by professional discontinuity during childbirth. Furthermore, measuring the antenatal experiences during pregnancy is the golden standard from a psychometric point of view. From an efficiency point of view, one could also argue to measure the antenatal experiences after birth and adjust the data to meet the experiences of the golden standard.

ACKNOWLEGDEMENTS

We are grateful to Bas Henk Peter Brummelhuis for his assistance with the data collection.

REFERENCES

1. Valentine N, de Silva A, Kawabata K, Darby C, Murray C, Evans B. Health system responsiveness: concepts, domains and measurement. In: Murray C, Evans B, eds. Geneva, Switzerland: World Health Organization; 2003:573-596.
2. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
3. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Soc Sci Med.* 2000;51(11):1611-1625.
4. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997;45(12):1829-1843.
5. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med.* 1998;47(10):1573-1588.
6. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med.* 1994;38(4):509-516.
7. Haugum M, Danielsen K, Iversen HH, Bjertnaes O. The use of data from national and other large-scale user experience surveys in local quality work: a systematic review. *Int J Qual Health Care.* 2014;26(6):592-605.
8. Weinick RM, Quigley DD, Mayer LA, Sellers CD. Use of CAHPS patient experience surveys to assess the impact of health care innovations. *Joint Commission journal on quality and patient safety.* 2014;40(9):418-427.
9. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
10. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
11. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
12. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
13. Dzakpasu S, Kaczorowski J, Chalmers B, Heaman M, Duggan J, Neusy E. The Canadian maternity experiences survey: design and methods. *Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC.* 2008;30(3):207-216.
14. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483
15. Redshaw M, Heikkila K. Delivered with care: A national survey of women's experience of maternity care. 2010; https://www.npeu.ox.ac.uk/delivered-with-care.
16. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg.
17. Wiegers TA, Keirse MJ, Berghs GA, van der Zee J. An approach to measuring quality of midwifery care. *Journal of clinical epidemiology.* 1996;49(3):319-325.
18. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring Client Experiences in Maternity Care under Change: Development of a Questionnaire Based on the WHO Responsiveness Model. *PLoS One.* 2015;10(2):e0117031.
19. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ.* 2016;4:e2092.
20. CPZ. Uniform gebruik cliëntervaringslijst ReproQ in de Geboortezorg [Uniform use client experience list ReproQ in maternity care]. 2015; http://www.goedgeboren.nl/netwerk/h/215/0/3596/rubriek--9/Uniform-gebruik-cli-ntervaringslijst-ReproQ-in-de-Geboortezorg, 2015.
21. Singh AS, Vik FN, Chinapaw MJ, et al. Test-retest reliability and construct validity of the ENERGY-child questionnaire on energy balance-related behaviours and their potential determinants: the ENERGY-project. *Int J Behav Nutr Phys Act.* 2011;8:136.

22. Scheerhagen M, Brinie E, Franx A, Van Stel HF, Bonsel GJ. What determines women's birth experiences? Applications for a benchmark. 2017;Submitted.
23. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica* 1979;4:263-291.
24. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain.* 1996;66(1):3-8.
25. Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. Cognitive and emotional perspectives. *Jama.* 1993;270(1):72-76.
26. Stiggelbout AM, de Vogel-Voogt E. Health state utilities: a framework for studying the gap between the imagined and the real. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2008;11(1):76-87.
27. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health Qual Life Outcomes.* 2004;2:14.
28. Schwartz CE, Andresen EM, Nosek MA, Krahn GL. Response shift theory: important implications for measuring quality of life in people with disability. *Archives of physical medicine and rehabilitation.* 2007;88(4):529-536.
29. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med.* 1999;48(11):1507-1515.
30. Blome C, Augustin M. Measuring change in quality of life: bias in prospective and retrospective evaluation. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2015;18(1):110-115.
31. Fischhoff B. Hindsight not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. 1975. *Quality & safety in health care.* 2003;12(4):304-311; discussion 311-302.
32. Pohl RF, Bender M, Lachmann G. Hindsight bias around the world. *Experimental psychology.* 2002;49(4):270-282.
33. Ruoss M. [Pain patients show a higher hindsight bias]. *Zeitschrift fur experimentelle Psychologie : Organ der Deutschen Gesellschaft fur Psychologie.* 1997;44(4):561-588.
34. Britton JR. The assessment of satisfaction with care in the perinatal period. *Journal of psychosomatic obstetrics and gynaecology.* 2012;33(2):37-44.
35. Quintana JM, Gonzalez N, Bilbao A, et al. Predictors of patient satisfaction with hospital health care. *BMC health services research.* 2006;6:102.
36. Raleigh VS, Hussey D, Seccombe I, Hallt K. Ethnic and social inequalities in women's experience of maternity care in England: results of a national survey. *Journal of the Royal Society of Medicine.* 2010;103(5):188-198.
37. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. *BMC medical informatics and decision making.* 2016;16(1):138.
38. PRN Foundation. Netherlands Perinatal Registry. *2004-2013*2013.

7

Shared Agenda Making for Quality Improvement; Towards More Synergy in Maternity Care

C.J.M. Groenen, N.T.L. van Duijnhoven, J.A.M. Kremer, M. Scheerhagen, F.P.H.A. Vandenbussche, M.J. Faber

Eur J Obstet Gynecol Reprod Biol. 2017; 219:15-19

ABSTRACT

Objectives. Professionals in maternity care have started working in a network approach. To further enhance the efficacy of this multidisciplinary maternity network, the identification of priorities for improvement is warranted. The aim of this study was to create key recommendations for the improvement agenda, in co-production with patients and professionals.

Study design. We conducted a Delphi study to inventory (round 1), prioritize (round 2) and eventually approve (round 3) the improvement agenda for the maternity network. Both patients and professionals joined this study.

Initial input for the study consisted of experiences from 397 patients, collected using the ReproQ questionnaire. In round 1, the expert panel, gave improvement recommendations, based on the ReproQ results. This resulted in 11 recommendations. In the second round, the expert panel prioritised these recommendations. In the consensus meeting then finally the concrete improvement agenda was composed.

Results. Priority scores differed considerably between patients and professionals in seven items, while four items received similar priority scores from both groups. The four most important improvement activities were: Realise more single bedrooms in hospitals; Create more opportunities for the continued presence of the community midwife during labour; Initiate a digital patient record view system for the network with a view function for patients; and Introduce a case manager for pregnant woman.

Conclusion. Based on patient experience and the active involvement of patients and professionals, we were able to compose the shared agenda for quality improvement in maternity care.

INTRODUCTION

Patient centered care implies the involvement of patients in the improvement agenda of health care. Patient involvement has shown benefits for shared decision making, research partnership, changes to service delivery and patient outcomes[1-5]. However, patient involvement in quality improvement is still limited[6], mainly due to uncertainties about the why and how[7]. Despite these uncertainties, Ocloo and Matthews shared principles that help to underpin practice in a collaborative framework with patients[1]. They recommend involving a diversity of patients, a clearly articulated purpose and a process that is co-designed or co-produced with patients.

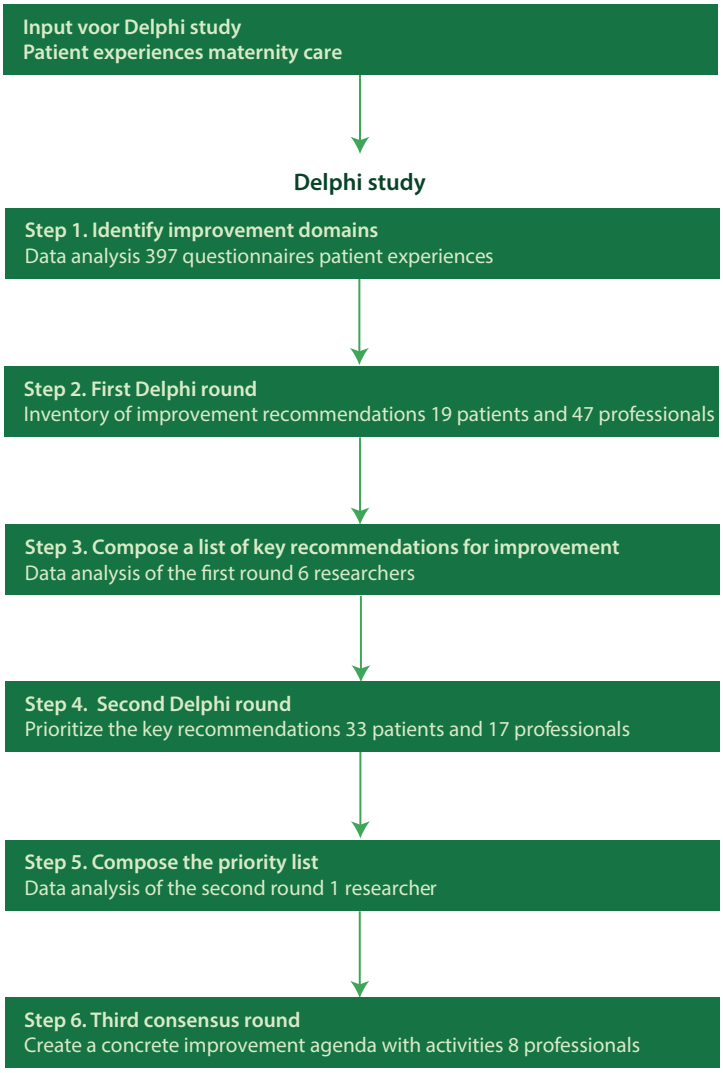
In order to realise patient centered care, maternity care professionals in the Netherlands started working in a network[8]. It is our strong belief that a central position for the patient, rather than the organisation, leads to better maternity care. In doing so, it is the patient who connects the professionals from different organisations. We believe that for providing direction in a new maternity network, a patient included improvement agenda is most valuable. Based on the aforementioned principles we therefore designed a study in which patients were actively engaged in creating and prioritising the improvement activities for the maternity network. We developed a Delphi study involving both patients and professionals as experts. The goal of this study was to achieve an improvement agenda for the multidisciplinary maternity network in co-production with patients.

METHODS

Setting

The study was performed in one multidisciplinary maternity network in the area of Nijmegen, region in the Netherlands with an average of 3.800 births a year and over 330 health professionals involved in maternity care. The different professionals working in maternity care were: community based midwives active in eleven independent practices, hospital based midwives, obstetricians (in training), and paediatricians working in two different hospitals (one providing secondary care and one providing secondary and tertiary care). The maternity care assistants gathered in one organisation and youth health doctors and nurses were positioned in 14 offices, which were coordinated by one organisation. Most pregnant women had received care from both the community based midwives, maternity care assistants, and youth health doctors and nurses. Besides, 59% of all pregnant women also received care from professionals working in a hospital[9].

Figure 1. The step-wise Delphi method to develop key recommendations for the improvement agenda.



Design

We used the Delphi method in our study. In the Delphi method an expert panel participates to gain consensus about a topic[10]. The expert panel participates anonymously. The Delphi study consists of rounds of questionnaires that are sent to the experts to gather and synthesise information. Our expert panel consisted of both patients and professionals. The patients were women who gave birth in the month before plotting the questionnaires, so

they had experience with maternal care. The professionals formed a representative diversity of the professions active in the different organizations from the multidisciplinary network, to enhance the acceptance of the key recommendations in the whole multidisciplinary network.

Creating the improvement agenda

Figure 1 shows the step by step Delphi method to develop key recommendations for the improvement agenda: 1) data analysis of the Repro Questionnaires of patient experience with maternity care, 2) first round Delphi questionnaire, 3) data analysis of the first round, 4) second round Delphi questionnaire, 5) data analysis of the second round, and 6) setting up the improvement agenda in a consensus group. These six steps includes the three Delphi rounds.

Step 1: data analysis of the questionnaires of patient experiences

Experiences from patients with the maternity care provided by the multidisciplinary network were measured by the Repro Questionnaire[11]. This validated questionnaire was developed to evaluate prenatal, natal and postnatal care, regardless of where the care is given[11-13]. Development of this self-report questionnaire was based on the 8-domain WHO Responsiveness model, including the following domains: Dignity, Autonomy, Confidentiality, Communication, Prompt attention, Social considerations, Basic amenities, Choice and continuity[14]. This questionnaire consisted of 32 questions, divided between the 8 domains. Examples of concrete questions were: treating with respect and giving personal attention, involving patient in decision-making, secured provision of medical information, information while treated and continuity of care provision when change of professional. For further detailed information about the questionnaire see 'Measuring client experiences in maternity care under change: development of a questionnaire based on the WHO Responsiveness model[11].

On a 4-point scale, women could evaluate their experience with maternity care, with '1' being the lowest score and '4' being the highest. The Repro Questionnaire was sent anonymously six weeks after childbirth to 812 women who gave birth in April and May 2013. The response rate was 49% and 397 Repro Questionnaires were analysed, serving as the input for the Delphi study and our starting point to create the improvement agenda. Due to the fact that all domains received mostly high scores, we decided to measure the negative scores for each domain, *i.e.* how often women scored '1' or '2'.

Step 2: first round Delphi questionnaire

The goal of the first round was to inventory concrete possibilities to improve the maternity care on the WHO Responsiveness domains in the ReproQ. The four WHO domains with the highest negative scores of patient' experience were presented to the expert panel in a

Delphi questionnaire. We explained the Repro Questionnaire and displayed the results and the accompanying ReproQ questions to the expert panel. Subsequently, the experts were asked to answer questions how to improve care in each of the four domains, i.e. Choice and continuity, Autonomy, Dignity, and Communication in the Nijmegen area. The Delphi questionnaires were sent to 50 professionals and given to 50 women at a postnatal check up appointment.

Step 3: data analysis of the first round

The responses were collected in an Excel file. The six researchers from the study group analysed the answers and categorised them by grouping corresponding ideas. This resulted in a list of 11 ideas to improve maternity care.

Step 4: second round Delphi questionnaire

The goal of the second round was to prioritise the improvement ideas. The list of 11 ideas was for that purpose sent to the expert panel with the request to assign 10 points in total to the ideas. Responders had the choice to select one idea and award it 10 points or to split the points between several ideas. Questionnaires were sent to 30 professionals and given to 50 women at a postnatal check up appointment.

Step 5: data analysis of the second round

The results of the second round in step 4 were analysed by the study group, thus composing a topic list for the consensus group meeting. In this topic list the for patients and professionals were shown separately.

Step 6: setting up the improvement agenda in the third consensus round

The goal of the third and last Delphi round was to make a concrete improvement agenda, that was applicable and achievable so that implementation of the activities would be feasible. The face-to-face consensus group was composed of one chairman, eight health professionals and one researcher to take notes. The results of the second round of the Delphi study were shared and the professionals were asked to formulate concrete improvement activities, of which regional implementation should be feasible within six months. The group members had to gain consensus of opinion on each improvement activity.

Data analysis

We used SPSS (version 20.0 for Windows: SPSS Inc., Chicago, IL, USA) to analyse ReproQ data. The study group jointly analysed the results of rounds 1 and 2 by coding and discussing the codes.

RESULTS

Table 1 shows the results of step 2, step 4 and step 6, respectively Delphi round 1, 2 and 3.

Step 1: data analysis of the questionnaires of patient experiences with maternity care

Results of 'negative scores' revealed four WHO Responsiveness domains with a score above 5%, which were selected for the Delphi study: Choice and continuity (16%), Autonomy (7%), Dignity (7%) and Communication (6%).

Step 2: first round Delphi questionnaire

We received 66 responses (19 patients=38% response rate and 47 professionals=94% response rate) with 65 different improvement ideas.

Step 3: data analysis of the first round

From these 65 different improvement ideas, several were comparable or complementary. For example, one suggestion was 'one website with the same information', another was 'similar information leaflets for pregnancy from the different organisations' and a third was 'multidisciplinary information meetings for pregnant women'. We combined these into one idea: 'uniform and unambiguous information from all professionals for pregnant women by using one website and similar information leaflets and meetings.' Some ideas were named differently, but the purport was similar. For example, professionals used the term case manager ('every pregnant woman will get one case manager'), whereas patients described the functions of a case manager ('one professional who is the sole contact person', 'one professional who guides me through my pregnancy').

Thus, we composed a list of 11 key recommendations.

Step 4: second Delphi questionnaire round

All 50 responders (33 patients=66% response and 17 professionals=57% response) filled in the list as requested. The method of scoring differed considerably between the responders: from 10 points for one idea to ten times 1 point for 10 ideas.

Step 5: data analysis of the second round

Seven recommendations were prioritised differently between patients and health professionals. Patients scored the highest on the ideas 'a single room for every woman' (22.1%), 'continuing of care during labour' (15.2%) and 'a case manager for every woman' (12.7%), while professionals scored respectively 2.9, 3.5 and 5.9%. Three patients awarded 10 points to one particular idea: 'shared decision making is part of the care', 'enable digital patient information transfer between professionals' and 'a single bedroom for every woman in case hospitalisation is needed'.

Table 1. Results Delphi study.

	First round	Second round		Third round
		Score (%)*		
	Improvement ideas	33 patients	17 professionals	Improvement agenda
Patients priority	A single bedroom for every woman if hospitalisation is needed.	**73 (22.1)**	5 (2.9)	1. Discuss with the hospital managers the wish and added value of single bedrooms for women in hospitals and try to realise this.
	Continuing of care during labour. If a transfer to secondary care is needed the community midwife is present for support.	**50 (15.2)**	6 (3.5)	2. Create more opportunities to continue presence of the community midwife during hospital labour.
	Enable digital patient information transfer between professionals .	**48 (14.6)**	23 (13.5)	3. Start of a digital patient record view system between organisations with view function for patients, so all required information is multidisciplinary available.
	Every pregnant woman has one case manager.	**42 (12.7)**	10 (5.9)	4. Introduction of a case manager for every pregnant woman.
Professionals priority	Online insight in patient records for patients and involved professionals from other organisations.	23 (7.0)	**36 (21.2)**	Included in 3.
	Uniform information from all professionals for pregnant women by using one website and similar information leaflets and meetings.	5 (1.5)	**33 (19.4)**	5. Create a multidisciplinary team of professionals to realise similar information provision for patients.
	Mandatory multidisciplinary training and (casuistry) meetings for more coordinated patient centered care.	15 (4.5)	**16 (9.4)**	6. Organisation of more communication trainings for professionals to realise patient centered care. Therefore using i.e. role models, more (patient) feedback, discussing in meetings. Including training the professionals to use the protocols more personalised rather than rigid.
Shared scores	Shared decision making is part of standard care.	33 (10.0)	14 (8.3)	Included in 6.
	A birth plan for every woman with wishes, decisions and a personal care path.	19 (5.8)	9 (5.3)	7. Implementation of a uniform birth plan for all pregnant women in the whole maternity network, with wishes, decisions and a personal care path . This birth plan must be known by all involved professionals.
	More attention or training by professionals for listening to pregnant women.	10 (3.0)	10 (5.9)	Included in 6.
	One telephone number for questions or needed help.	12 (3.6)	8 (4.7)	*Not used for the improvement agenda, considered not feasible.*

*Percentage of maximal amount of points

Professionals scored the highest on 'insight in patient records' (21.2%), 'same information' (19.4%), 'multidisciplinary training' (9.4) and 'training in listening' (5.9%), while patients scored respectively 7.0, 1.5, 4.5 and 3.0%.

Four of the recommendations received comparable priority scores from patients and professionals. From these recommendations, 'shared decision making is part of the care' scored 10.0% among patients and 8.3% among professionals. The other three recommendations revealed the lowest overall priority scores.

Step 6: setting up the improvement agenda in a consensus group

The results of the priority list were shared in the face-to-face consensus group. Due to the multidisciplinary character of the consensus group, members learned from each other. Sometimes, a proposed concrete idea was already implemented by one of the organisations. Therefore, the first group objective was to learn from each other or to implement something multidisciplinary instead of only in one organisation. The group concluded that 10 ideas out of the 11 required concrete implementation activities, because they could improve maternity care. The idea 'one telephone number for the whole maternity network' was considered unfeasible in the region, and because it was prioritised the least (4%), it was decided not to include this idea in the improvement agenda.

Some of the initial ideas in Delphi round 1 were already concrete (for example 'a single room for every woman' and 'the introduction of a birth plan'), but most were more thematic (for example 'shared decision making' and 'continuing of care at birth'). The latter received the most attention in the consensus group, because of the need to formulate concrete improvement activities. For example, 'shared decision making' was translated into the more concrete 'train the professionals in using the protocols in a more personalised manner instead of rigid'. The group concluded the session with seven concrete improvement activities on which they reached agreement.

DISCUSSION

In the present study, a total of 449 patients and 47 professionals contributed to the creation of the shared improvement agenda for the multidisciplinary maternity network. Therefore, this improvement agenda is a well developed co-production between health professionals and patients, in conformity with the principles of Ocloo and Matthews[1]. In this, the patients formed the link between the professionals of the various organisations. We consider this to be essential in creating a patient centered network and a shared agenda for quality improvement. Utilizing the Delphi study method, we had the possibility to (anonymously) involve a large number of patients and professionals in the decision making process. The prioritising scores given by patients and professionals showed substantial differences and without the inclusion of patients our improvement agenda would obviously have been very

different. Previous studies also show differences between patients' experiences and the perception of health professionals' of their patients' experiences[15,16]. A discrepancy between the perspectives of experts and patients was even the most frequently reported barrier in patient and public involvement programs to develop and implement clinical practice guidelines[17]. These findings give more importance to involving patients in the improvement of health care and to realise patient centered care

To our knowledge, this is the first maternity improvement agenda based on both patient and professional experiences. Also, this time different professionals from different organisations worked together to create this agenda and were connected through the patient in order to improve patient centered care. This combined approach is novel now that the most common way is still to develop improvement activities solely on a health organisation level. We expect this original and broad approach of the shared agenda will lead to more patient satisfaction regarding maternity care, even if we believe this care to already be of high quality, such as in the Nijmegen area. With the current improvement agenda, professionals and managers of the multidisciplinary network have a concrete guide to further improve their care. This agenda has already been offered to the network professionals and the implementation of different activities has already started.

The main strength of this study is the inclusion of the large number of patients from a broad perspective of society. Furthermore, all multidisciplinary professions were involved in the study and the Delphi study resulted in a concrete useful shared agenda for quality improvement. This study does have some shortcomings, however. In the third round we invited only professionals and no patients into the consensus group. We also opted for two different groups of patient expert members in round 1 and 2, because we distributed the questionnaires anonymously at postnatal checkups. As such, we did not apply the exact Delphi study design. Because a multidisciplinary network approach in maternity care is one of the main goals for the Dutch Government[18], and because the networks are looking how to offer more patient centered care, this study is of great value to Dutch maternity care. We believe that the Delphi design with patients included is a helpful instrument in the development of a concrete and useful improvement agenda and this will lead towards more synergy in maternity care. This study could be implemented as a network activity in any care network, and can be repeated periodically. It can therefore be used as a Plan-Do-Check-Act cycle. In doing so, networks continuously work on improvement and on patient centered care.

ETHICAL APPROVAL

The medical ethical committee of the Radboud University Medical Centre has awarded full ethical approval for this project (CMO No. 2011/381). The study is registered at the Dutch Trial Register (NTR, TC=4063).

REFERENCES

1. Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf.* 2016;25(8):626-632.
2. de Labrusse C, Ramelet AS, Humphrey T, Maclennan SJ. Patient-centered Care in Maternity Services: A Critical Appraisal and Synthesis of the Literature. *Womens Health Issues.* 2016;26(1):100-109.
3. Ekman I, Swedberg K, Taft C, et al. Person-centered care--ready for prime time. *Eur J Cardiovasc Nurs.* 2011;10(4):248-251.
4. Lewin SA, Skea ZC, Entwistle V, Zwarenstein M, Dick J. Interventions for providers to promote a patient-centred approach in clinical consultations. *Cochrane Database Syst Rev.* 2001(4):Cd003267.
5. Van Royen P, Beyer M, Chevallier P, et al. The research agenda for general practice/family medicine and primary health care in Europe. Part 3. Results: person centred care, comprehensive and holistic approach. *Eur J Gen Pract.* 2010;16(2):113-119.
6. Armstrong N, Herbert G, Aveling EL, Dixon-Woods M, Martin G. Optimizing patient involvement in quality improvement. *Health Expect.* 2013;16(3):e36-47.
7. Boivin A, Lehoux P, Burgers J, Grol R. What are the key ingredients for effective public involvement in health care improvement and policy decisions? A randomized trial process evaluation. *The Milbank quarterly.* 2014;92(2):319-350.
8. Cunningham FC, Ranmuthugala G, Plumb J, Georgiou A, Westbrook JI, Braithwaite J. Health professional networks as a vector for improving healthcare quality and safety: a systematic review. *BMJ Qual Saf.* 2012;21(3):239-249.
9. Perined. *Perinatal care in The Netherlands 2013 (Perinatale zorg in Nederland 2013).* Utrecht: Perined;2015.
10. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One.* 2011;6(6):e20476.
11. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring Client Experiences in Maternity Care under Change: Development of a Questionnaire Based on the WHO Responsiveness Model. *PLoS One.* 2015;10(2):e0117031.
12. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the dutch obstetric care system. *BMC Health Services Research.* 2014;14(622).
13. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ.* 2016;4:e2092.
14. Gostin L, Hodge J, G, Valentine N, Nygren-Krug H. *The Domains of Health Responsiveness – A Human Rights Analysis.* Geneva, Switzerland: World Health Organization;2003.
15. Aarts JW, Faber MJ, van Empel IW, Scheenjes E, Nelen WL, Kremer JA. Professionals' perceptions of their patients' experiences with fertility care. *Human reproduction (Oxford, England).* 2011;26(5):1119-1127.
16. van Empel IW, Aarts JW, Cohlen BJ, et al. Measuring patient-centredness, the neglected outcome in fertility care: a random multicentre validation study. *Human reproduction (Oxford, England).* 2010;25(10):2516-2526.
17. Legare F, Boivin A, van der Weijden T, et al. Patient and public involvement in clinical practice guidelines: a knowledge synthesis of existing programs. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2011;31(6):E45-74.
18. Nederland Z, Zorg CP. *Zorgstandaard Integrale Geboortezorg.* 2016.

8

Experiences of Women who Planned Birth in a Birth Centre Compared to Alternative Planned Places of Birth. Results of the Dutch Birth Centre Study

M.F. Hitzert, M.A. Hermus, M. Scheerhagen, I.C. Boesveld, T.A. Wiegers, M.E. van den Akker-van Marle, P. van Dommelen, K.M. van der Pal-de Bruin, J.P. de Graaf

Midwifery. 2016; 40:70-78

ABSTRACT

Objective. To assess the experiences with maternity care of women who planned birth in a birth centre and to compare them to alternative planned places of birth, by using the responsiveness concept of the World Health Organization.

Design. This study is a cross-sectional study using the ReproQ questionnaire filled out eight to ten weeks after birth. The primary outcome was responsiveness of birth care. Secondary outcomes included overall grades for birth care and experiences with the birth centre services. Regression analyses were performed to compare experiences among the planned places of birth. The study is part of the Dutch Birth Centre Study.

Setting. The women were recruited by 82 midwifery practices in the Netherlands, within the study period 1 August 2013 and 31 December 2013.

Participants. A total of 2162 women gave written consent to receive the questionnaire and 1181 (54.6%) women completed the questionnaire.

Measurements and findings. Women who planned to give birth at a birth centre:

1) had similar experiences as the women who planned to give birth in a hospital receiving care of a community midwife.
2) had significantly less favourable experiences than the women who planned to give birth at home. Differences during birth were seen on the domains Dignity (OR=1.58, 95% CI=1.09 – 2.27) and Autonomy (OR=1.77, 95% CI=1.25 – 2.51), during the postpartum period on the domains Social considerations (OR=1.54, 95% CI=1.06 – 2.25) and Choice and continuity (OR=1.43, 95% CI=1.00 – 2.03).
3) had significantly better experiences than the women who planned to give birth in a hospital under supervision of an obstetrician. Differences during birth were seen on the domains Dignity (OR=0.51, 95% CI=0.31 – 0.81), Autonomy (OR=0.59, 95% CI=0.35 – 1.00), Confidentiality (OR=0.57, 95% CI=0.36 – 0.92) and Social considerations (OR=0.47, 95% CI=0.28 – 0.79). During the postpartum period differences were seen on the domains Dignity (OR=0.61, 95% CI=0.38 – 0.98), Autonomy (OR=0.52, 95% CI=0.31 – 0.85) and Basic amenities (OR=0.52, 95% CI=0.30 – 0.88). More than 80% of the women who received care in a birth centre rated the facilities, the moment of arrival/departure and the continuity in the birth centre as good.

Key conclusions and implications for practice. In the last decades, many birth centres have been established in different countries, including the United Kingdom, Australia, Sweden and the Netherlands. For women who do not want to give birth at home a birth centre is a good choice: it leads to similar experiences as a planned hospital birth. Emphasis should be placed on ways to improve Autonomy and Prompt attention for women who plan to give birth in a birth centre as well as on the improvement of care in case of a referral.

INTRODUCTION

Traditionally, the quality of maternity care is described in terms of perinatal morbidity and mortality outcomes. Currently, other aspects of health care, such as client experiences, are important as well, also in terms of their potential to affect clinical outcomes[1-4]. The Dutch maternity care system is often set as an example to learn from, because of its high home birth rate, its low number of obstetric interventions and a consequence, low cost and yet high assumed health outcomes[5-9]. In the Netherlands, the quality of care experienced by women during the maternity care process in general is high[10].

The Dutch maternity care system is based on primary care provided by independent community midwives caring for women with a 'normal', uncomplicated, or low-risk pregnancy. Obstetricians provide in-hospital secondary care for women with a complicated, or high-risk pregnancy or birth. When a complication occurs or the risk of a complication increases substantially during pregnancy or during labour, or when pharmacological pain relief is requested, a woman will be referred from primary to secondary care. For women who were referred to secondary care before the 36th week of pregnancy, their planned place of birth will by necessity be in a hospital, under supervision of an obstetrician. Low-risk women can choose where they want to give birth: in a birth centre, in hospital or at home, all receiving care from a community midwife. Dutch birth centres have been established in the last decade to accommodate the growing number of low-risk women who do not want to give birth at home. A birth centre is a setting where women with uncomplicated pregnancies can give birth in a home-like environment[11].

Several international studies have explored the influences of the birth settings on the experience of women. A randomized, controlled trial in Sweden showed that low-risk women giving birth in a birth centre expressed greater satisfaction with care than women who gave birth in a hospital[12]. A study in Australia showed that a birth centre setting ensured that women received personalised, genuine care that transcended the entire childbearing continuum[13]. Differences in philosophy between hospital and birth centre settings is seen as an important component of care experiences[14]. It is also known that women who have given birth in a specific birth centre were less satisfied than those who have given birth at home[15]. In Australia, women giving birth at home rated their midwives higher than women giving birth at a hospital, with women giving birth in a birth centre generally scoring between the other two groups[16].

Currently we know very little of how women who planned to give birth in a birth centre experienced their care in the Netherlands. There is no study available that compares the experiences in birth centres with other birth settings in the Netherlands. Therefore, the aim of this study was to assess the experiences with maternity care of women who planned birth in a birth centre and to compare them to alternative planned places of birth, by using the responsiveness concept of the World Health Organization. The World Health Organization introduced the concept of responsiveness as one of the available approaches to address

service quality in an internationally comparable way[17]. The concept offers the opportunity to capture client's experiences on eight predefined domains. Responsiveness is defined as aspects of the way individuals are treated and the environment in which they are treated during health system interactions[18,19]. The concept has been applied in the Dutch maternity care a few times before[20,21].

This research is part of the Dutch Birth Centre Study[22]. This national project evaluates the effect of Dutch birth centres on aspects such as client and partner experiences, process and outcome variables, costs and professional experiences.

METHODS

Setting

The study was designed as a cross-sectional study. A minimum of three midwifery practices working in the area of each of the 23 birth centres included in the Dutch Birth Centre Study, were randomly recruited. This resulted in the participation of 82 midwifery practices. During the study period from 1 August to 31 December 2013 these 82 midwifery practices recruited women for participation. The midwifery practices varied in size and were located all over the country.

Data collection

Almost all women in the Netherlands, including women who gave birth under responsibility of an obstetrician, receive postpartum care from community midwives. During the data collection period, the community midwives of the 82 practices asked the women who received postpartum care for permission to send them a questionnaire. In this way, data were obtained from women with different planned places of birth: at a birth centre, in a hospital, or at home and under care of a midwife or an obstetrician. Excluded were women who could not read or speak Dutch and women with no specific preference for a place of birth.

A total of 2,162 women gave written consent either to receive the questionnaire through e-mail, as a hard-copy or to have an interview by phone. We explicitly tried to include women from different backgrounds, by giving the choice of an interview by phone. The women completed the questionnaire around eight to ten weeks after birth. A reminder was sent two weeks later, when needed.

Questionnaire

The ReproQ is a two-part questionnaire (part 1 prenatal, part 2 postnatal) and was developed to assess the responsiveness of the maternity care system in the Netherlands by evaluating client experiences. Responsiveness is defined as 'aspects of the way individuals are treated and the environment in which they are treated during health system interactions'[21]. The

postnatal part of the ReproQ was used in this study and includes two reference periods: the event of labour and birth and the subsequent postpartum week. The questionnaire consists of the following components: 1) questions about the process of care, including referral or emergency situations, 2) a question about the grade of overall experience during birth and the postpartum period, 3) questions about the eight domains of the WHO concept of responsiveness, 4) questions including experienced health outcomes, 5) the individual ranking of the various domains of responsiveness according to their importance and 6) the respondent's socio-demographic characteristics. For this study, questions about facilities (e.g. homelike environment, hotel service and bath) and transfers (e.g. change of caregiver and change of room) were included for women who received care in a birth centre.

The responsiveness concept is described to consist of eight domains: 1) Dignity, 2) Autonomy, 3) Confidentiality, 4) Communication, 5) Prompt attention, 6) Social consideration, 7) Basic amenities and 8) Choice and continuity. Each domain consists of several items, see Table 1.

The questions could be answered on a four-point scale with the values: always (4), mostly (3), sometimes (2) and never (1)[17]. An average score per domain was computed this way. The questionnaire avoids any implicit or explicit preference towards the providers or the organizational structures, leaving room to compare different organizational structures and different levels of care[21].

Data handling

Questionnaires were excluded if more than 50% of the answers were missing in two or more domains. The client experiences were compared according to the women's planned place of birth. The information was based on the place of birth as it was planned one month before the birth, as recorded in the questionnaire. Subgroup analyses were performed for women referred to secondary care during birth and women who were not referred.

Data analysis

The basic characteristics of our respondents were compared with the characteristics of all the women receiving postpartum care of a participating midwife, the reference group. Therefore, data of all births occurring in the midwifery practices that participated in our study between August 2013 and December 2013 were derived from the Netherlands Perinatal Registry (PRN-foundation). This PRN-foundation is a joint effort of four professions (midwives, general practitioners, obstetricians and paediatricians) that provide perinatal care in the Netherlands. All these professions have their own volunteer-based medical registries, which are linked to one combined PRN-registry[23].

Table 1. Items covered by the eight responsiveness domains.

Domain	Item
Dignity	Respecting privacy
	Treating with respect
	Giving personal attention
	Treating with kindness
	Considering personal wishes regarding birth
	Trustworthy as health professional
Autonomy	Involving client in decision-making
	Acceptance of treatment refusal
	Involving client in decision-making on pain relief
	Involving client in decision-making on setting of birth
Confidentiality	Providing medical information to family members after consent
	Discussing the medical situation without others hearing it
	Secured provision of medical information to others
Communication	Responsive to client questions
	Consistency of advice across professionals
	Comprehensibility of explanation
	Provision of information while treated
Prompt attention	Access for contact in urgent situations
	Access for contact without urgency
	Waiting time for service
	Availability of maternity care assistance
	Physical accessibility of setting
	Prompt phone response of health professional
Social considerations	Involvement of the partner or family in care provision
	Attention for family and household
	Support from partner or family
Basic amenities	Comfort of setting
	Hygiene of setting
	Physical accessibility of places (e.g. room and bathroom)
Choice and continuity	Continuity of care provision when change of individual professional (same discipline)
	Continuity of care provision when change professional (across disciplines)
	Allowance for selecting a preferred type of health professional
	Being explicit on which health professional is actual in charge

Univariate analyses were carried out using the chi-square test and the Fisher's exact test for categorical factors and a one-way analysis of variance was carried out for continuous-characteristics. The mean and median grade (on a 10-point scale), including the 25th and 75th percentile, of the experience of overall care were calculated according to the planned place of birth.

Logistic regression analyses were performed with the responsiveness outcomes as dependent variables (optimal=4 and non-optimal<4) and with the planned place of birth as independent variable. We adjusted for the basic characteristics that differed among the groups: parity, education and ethnicity. The birth centre group was used as reference. P values less than 0.05 (two-sided) were considered statistically significant.

Descriptive analyses were performed on the additional questions about the birth centre services. The questions were filled out only by women who received care in a birth centre. The analyses were performed with SPSS 21.0[24].

Ethical considerations

The design and planning of the study were presented to the Medical Ethics Committee of the University Medical Centre Utrecht. They confirmed that this study agrees with the Dutch legal regulations in terms of the methods used in this study and, therefore, an official ethical approval is not required[25]. To invite the clients for participation in this study, permission from the midwifery practices was obtained. Informative letters to the clients were given by the midwifery practices directly. The letter clearly explained that if a client did not want to participate, she was not obligated to do so and this would not affect her care process. By signing the letter, clients consented either to receive the questionnaire digitally, as a hard-copy or to have an interview by phone.

RESULTS

Study population

A total of 2162 women gave permission to receive the questionnaire; 1654 (76.5%) by e-mail, 464 (21.5%) by post and 44 (2.0%) women wanted to be interviewed by phone. We received 1181 completed questionnaires (including interviews by phone), with a total response rate of 54.6%. Forty-seven questionnaires were excluded, leading to 1134 questionnaires available for the analysis: 263 with a planned birth centre birth, 350 with a planned home birth, 262 with a planned hospital birth under care of a community midwife and 115 with a planned hospital birth under supervision of an obstetrician.

Table 2 shows the characteristics of the participants and the reference group. No differences were found in parity and referral during birth between the respondents and the total group of women who gave birth in one of the participating midwifery practices. However, the respondents were significantly older, had a higher SES score, were more often

of Dutch origin, were more often under supervision of the midwife at the start of labour and the respondents received less often an intervention during birth, compared to the reference group.

Table 2. Characteristics of the respondents and the reference group.

	Participants	Reference group
	(n=1081)	(n=61169)
	No. (%)	No. (%)
Characteristics		
Age*		
≤25	56 (5.6)	9204 (15.1)
26 – 35	736 (73.2)	42516 (69.6)
≥36	213 (21.2)	9322 (15.3)
Parity		
Primiparous	490 (47.9)	28160 (46.1)
Multiparous	532 (52.1)	32971 (53.9)
SES*		
Low	70 (6.5)	10342 (16.9)
Middle	807 (74.7)	41395 (67.7)
High	204 (18.9)	9432 (15.4)
Ethnicity*		
Dutch	921 (91.7)	46280 (78.1)
non-Dutch	83 (8.3)	12981 (21.9)
Start birth*		
Midwife supervision	880 (82.1)	35288 (57.7)
Obstetrician supervision	192 (17.9)	25881 (42.3)
Referral during birth		
No	815 (76.6)	46258 (75.6)
Yes	249 (23.4)	14903 (24.4)
Interventions*		
No vacuum/forceps or section caesarean	928 (86.0)	47144 (77.1)
Vacuum extraction/forceps	98 (9.1)	4852 (7.9)
Section caesarean	53 (4.9)	9173 (15.0)

* p-value <0.05 (chi-square test)

Table 3. Respondent's characteristics according to planned place of birth.

	Community midwife			Obstetrician	
	Birth centre	Hospital	Home	Hospital	Total
	(n=263)¥	(n=262)¥	(n=350)¥	(n=115)¥	(n=990)¥
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
Age					
≤ 25	12 (4.6)	14 (5.8)	21 (6.5)	3 (2.7)	50 (5.3)
26 – 35	195 (75.0)	174 (72.5)	238 (73.2)	76 (69.1)	683 (73.1)
≥ 36	53 (20.4)	52 (21.7)	66 (20.3)	31 (28.2)	202 (21.6)
Parity*					
Primiparous	154 (58.8)	113 (46.5)	126 (38.0)	47 (42.3)	440 (46.4)
Multiparous	108 (41.2)	130 (53.5)	206 (62.0)	64 (57.7)	508 (53.6)
Education*					
Low	16 (6.1)	14 (6.0)	26 (8.0)	10 (9.4)	66 (7.1)
Middle	64 (24.4)	72 (30.9)	120 (36.9)	35 (33.0)	291 (31.4)
High	182 (69.5)	147 (63.1)	179 (55.1)	61 (57.5)	569 (61.4)
Ethnicity*					
Dutch	247 (93.9)	203 (84.6)	312 (96.3)	93 (85.3)	855 (91.3)
Non-Dutch	16 (6.1)	37 (15.4)	12 (3.7)	16 (14.7)	81 (8.7)
Actual place of birth**					
Birth centre	128 (48.7)	6 (2.3)	4 (1.1)	0 (0.0)	138 (13.9)
Home	18 (6.8)	26 (9.9)	232 (66.3)	0 (0.0)	276 (27.9)
Hospital, under care of a midwife	7 (2.7)	137 (52.3)	20 (5.7)	0 (0.0)	164 (16.6)
Hospital, under supervision of an obstetrician	107 (40.7)	91 (34.7)	90 (25.7)	114 (99.1)	402 (40.7)
Unknown	3 (1.1)	2 (0.8)	4 (1.1)	1 (0.9)	10 (1.0)
Experienced health mother in general					
Poor/moderate	9 (3.4)	6 (2.5)	6 (1.8)	8 (7.2)	29 (3.1)
Good	76 (28.9)	67 (27.5)	84 (25.5)	42 (37.8)	269 (28.4)
Very good	100 (38.0)	101 (41.4)	138 (41.8)	35 (31.5)	374 (39.5)
Excellent	78 (29.7)	70 (28.7)	102 (30.9)	26 (23.4)	276 (29.1)
Experienced health mother after birth					
Healthy	172 (65.4)	182 (69.7)	254 (72.8)	67 (58.3)	675 (68.3)
Small problems	77 (29.3)	67 (25.7)	82 (23.5)	39 (33.9)	265 (26.8)
Big problems/problems, impact unclear	14 (5.3)	12 (4.5)	13 (3.8)	9 (7.8)	48 (4.9)

Table 3. *Continued*

	Community midwife			Obstetrician	
	Birth centre	Hospital	Home	Hospital	Total
	(n=263)¥	(n=262)¥	(n=350)¥	(n=115)¥	(n=990)¥
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
Experienced health baby after birth					
Healthy	229 (87.4)	229 (87.4)	318 (91.4)	93 (80.9)	869 (88.0)
Small problems	29 (11.1)	25 (9.5)	22 (6.3)	20 (17.4)	96 (9.7)
Big problems/problems, impact unclear	4 (1.6)	8 (3.1)	8 (2.3)	2 (1.8)	22 (2.2)
Hospital admission of the child after birth*					
No	188 (72.3)	196 (74.8)	304 (87.4)	58 (50.9)	746 (75.8)
Yes, at the maternity ward	63 (24.2)	58 (22.1)	38 (10.9)	41 (36.0)	200 (20.3)
Yes, high care	9 (3.5)	8 (3.1)	6 (1.7)	15 (13.2)	38 (3.9)

* p-value <0.05 (chi-square test/Fisher's test)
** p-value <0.05 (statistical test are performed on expected place is equal to the final place of birth; hospital births under supervision of an obstetrician and unknown groups are excluded)
¥ numbers are varying between characteristics due to missing data

Table 3 shows the characteristics of the respondents according to their planned place of birth. The women who planned to give birth at a birth centre were more often primiparous and highly educated compared to the women who planned to give birth under care of a community midwife in a hospital, at home or under supervision of an obstetrician in a hospital. The women who planned to give birth at a birth centre or at home were more often of Dutch origin compared to the women who planned to give birth in a hospital (under care of a community midwife or of an obstetrician).

Grades for experiences during birth and the postpartum period

In general, the mean and median grades of experiences during birth and the postpartum period (adjusted for parity, education and ethnicity) were quite similar within each planned places of birth. The mean grades for the planned place of birth were 8.4 (sd=1.3) at a birth centre, 8.4 (sd=1.3) in a hospital under care of a community midwife, 8.7 (sd=1.3) at home and 8.0 (sd=1.6) in a hospital under supervision of an obstetrician. The mean grade for the planned place of birth at a birth centre was significantly ($p<0.05$) higher than the mean grade for the planned place of birth in a hospital under supervision of an obstetrician. The median grades were respectively 9, 8, 9 and 8.

Table 4. Responsiveness outcomes according to planned place of birth.

Under care of a community midwife												
	Birth centre (REF)		Hospital					Home				
	(n=263)		(n=262)					(n=350)				
	No. (%)		No. (%)					No. (%)				
Responsiveness during birth	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**CRUDE OR**	**Adj OR**	**95% CI**	**optimal**	**non-optimal**	**CRUDE OR**	**Adj OR**	**95% CI**
Dignity	163 (62.0)	100 (38.0)	165 (63.0)	97 (37.0)	1.04	0.94	0.65 – 1.37	265 (75.7)	85 (24.3)	1.91	1.58*	1.09 – 2.27
Autonomy	92 (36.2)	162 (63.8)	104 (41.3)	148 (58.7)	1.24	1.11	0.76 – 1.61	182 (53.5)	158 (46.5)	2.03	1.77***	1.25 – 2.51
Confidentiality	180 (69.8)	78 (30.2)	170 (67.7)	81 (32.3)	0.91	0.84	0.57 – 1.25	244 (71.3)	98 (28.7)	1.08	1.08	0.75 – 1.57
Communication	145 (55.3)	117 (44.7)	131 (52.0)	121 (48.0)	0.87	0.79	0.55 – 1.14	200 (58.8)	140 (41.2)	1.15	1.05	0.75 – 1.48
Prompt attention	145 (55.1)	118 (44.9)	139 (55.4)	112 (44.6)	1.01	0.99	0.69 – 1.42	218 (65.1)	117 (34.9)	1.52	1.37	0.97 – 1.93
Social considerations	212 (80.6)	51 (19.4)	187 (74.8)	63 (25.2)	0.71	0.70	0.45 – 1.08	276 (82.9)	57 (17.1)	1.17	1.16	0.76 – 1.79
Basic amenities	215 (82.1)	47 (17.9)	189 (76.2)	59 (23.8)	0.70	0.68	0.44 – 1.07	278 (84.5)	51 (15.5)	1.19	1.21	0.77 – 1.90
Choice and continuity	159 (60.7)	103 (39.3)	157 (64.1)	88 (35.9)	1.16	1.08	0.74 – 1.57	221 (67.8)	105 (32.2)	1.36	1.16	0.81 – 1.64
Responsiveness postpartum	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**CRUDE OR**	**Adj OR**	**95% CI**	**optimal**	**non-optimal**	**CRUDE OR**	**Adj OR**	**95% CI**
Dignity	169 (64.3)	94 (35.7)	165 (63.0)	97 (37.0)	0.95	0.93	0.64 – 1.35	254 (73.0)	94 (27.0)	1.50	1.37	0.95 – 1.97
Autonomy	196 (76.6)	60 (23.4)	176 (70.4)	74 (29.6)	0.73	0.71	0.47 – 1.07	270 (80.6)	65 (19.4)	1.27	1.20	0.80 – 1.82
Confidentiality	174 (67.4)	84 (32.6)	154 (61.1)	98 (38.9)	0.76	0.76	0.53 – 1.11	239 (69.3)	106 (30.7)	1.09	1.09	0.76 – 1.56
Communication	96 (36.6)	166 (63.4)	108 (42.9)	144 (57.1)	1.30	1.19	0.83 – 1.73	155 (45.5)	186 (54.5)	1.44	1.28	0.91 – 1.80
Prompt attention	158 (60.1)	105 (39.9)	137 (54.6)	114 (45.4)	0.80	0.81	0.56 – 1.16	223 (66.6)	112 (33.4)	1.32	1.22	0.86 – 1.73
Social considerations	179 (68.1)	84 (31.9)	162 (65.1)	87 (34.9)	0.87	0.83	0.57 – 1.22	253 (76.0)	80 (24.0)	1.48	1.54*	1.06 – 2.25
Basic amenities	208 (80.6)	50 (19.4)	197 (81.1)	46 (18.9)	1.03	1.02	0.65 – 1.63	267 (81.9)	59 (18.1)	1.09	1.02	0.66 – 1.58
Choice and continuity	156 (59.5)	106 (40.5)	156 (63.7)	89 (36.3)	1.19	1.19	0.82 – 1.72	226 (69.3)	100 (30.7)	1.54	1.43*	1.00 – 2.03

Table 4. *Continued*

	Under supervision of an obstetrician				
	Hospital				
	(n=115)				
	No. (%)				
Responsiveness during birth	**optimal**	**non-optimal**	**CRUDE OR**	**Adj OR**	**95% CI**
Dignity	56 (48.7)	59 (51.3)	0.58	0.51**	0.32 – 0.81
Autonomy	30 (28.6)	75 (71.4)	0.70	0.59*	0.35 – 1.00
Confidentiality	65 (58.6)	46 (41.4)	0.61	0.57*	0.36 – 0.92
Communication	55 (49.1)	57 (50.9)	0.78	0.71	0.45 – 1.13
Prompt attention	55 (49.1)	57 (50.9)	0.79	0.70	0.44 – 1.11
Social considerations	76 (67.3)	37 (32.7)	0.49	0.47**	0.28 – 0.79
Basic amenities	83 (73.5)	30 (26.5)	0.61	0.60	0.35 – 1.04
Choice and continuity	59 (52.7)	53 (47.3)	0.72	0.65	0.41 – 1.04
Responsiveness postpartum	**optimal**	**non-optimal**	**CRUDE OR**	**Adj OR**	**95% CI**
Dignity	61 (53.0)	54 (47.0)	0.63	0.61*	0.38 – 0.98
Autonomy	72 (64.3)	40 (35.7)	0.55	0.52**	0.31 – 0.85
Confidentiality	71 (63.4)	41 (36.6)	0.84	0.82	0.51 – 1.32
Communication	49 (43.4)	64 (56.6)	1.32	1.24	0.78 – 1.98
Prompt attention	57 (50.4)	56 (49.6)	0.68	0.65	0.41 – 1.03
Social considerations	73 (64.6)	40 (35.4)	0.86	0.88	0.54 – 1.43
Basic amenities	78 (69.6)	34 (30.4)	0.55	0.52*	0.30 – 0.88
Choice and continuity	57 (50.9)	55 (49.1)	0.70	0.72	0.46 – 1.15

Birth centre as reference and adjusted for parity, education and ethnicity
* p<0.05, ** p<0.01, *** p<0.001

Table 5. Responsiveness outcomes according to planned place of birth for referred and non-referred women.

	Under care of a community midwife											
	Birth centre						Hospital					
	non-referral (REF)		referral				non-referral (REF)		referral			
	(n=177)		(n=83)				(n=196)		(n=60)			
	No. (%)		No. (%)				No. (%)		No. (%)			
Responsiveness during birth	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**Adj OR**	**95% CI**	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**Adj OR**	**95% CI**
Dignity	125 (70.6)	52 (29.4)	132 (67.3)	64 (32.7)	0.33***	0.19 – 0.58	132 (67.3)	64 (32.7)	29 (48.3)	31 (51.7)	0.51*	0.27 – 0.97
Autonomy	74 (43.8)	95 (56.2)	86 (45.7)	102 (54.3)	0.38**	0.20 – 0.71	86 (45.7)	102 (54.3)	15 (25.9)	43 (74.1)	0.45*	0.22 – 0.94
Confidentiality	126 (72.8)	47 (27.2)	138 (73.4)	50 (26.6)	0.66	0.37 – 1.17	138 (73.4)	50 (26.6)	28 (49.1)	29 (50.9)	0.41**	0.21 – 0.78
Communication	108 (61.4)	68 (38.6)	108 (56.8)	82 (43.2)	0.52*	0.30 – 0.91	108 (56.8)	82 (43.2)	19 (33.9)	37 (66.1)	0.48*	0.25 – 0.93
Prompt attention	108 (61.0)	69 (39.0)	117 (61.9)	72 (38.1)	0.51*	0.29 – 0.88	117 (61.9)	72 (38.1)	19 (33.9)	37 (66.1)	0.32***	0.16 – 0.62
Social considerations	151 (85.3)	26 (14.7)	145 (77.1)	43 (22.9)	0.39**	0.20 – 0.75	145 (77.1)	43 (22.9)	36 (64.3)	20 (35.7)	0.49*	0.25 – 0.97
Basic amenities	152 (86.4)	24 (13.6)	142 (76.3)	44 (23.7)	0.44*	0.22 – 0.86	142 (76.3)	44 (23.7)	43 (76.8)	13 (23.2)	0.90	0.43 – 1.87
Choice and continuity	125 (71.0)	51 (29.0)	132 (71.7)	52 (28.3)	0.26***	0.15 – 0.45	132 (71.7)	52 (28.3)	21 (38.2)	34 (61.8)	0.25***	0.13 – 0.48
Responsiveness postpartum	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**Adj OR**	**95% CI**	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**Adj OR**	**95% CI**
Dignity	122 (68.9)	55 (31.1)	131 (66.8)	65 (33.2)	0.48*	0.28 – 0.84	131 (66.8)	65 (33.2)	32 (53.3)	28 (46.7)	0.71	0.37 – 1.35
Autonomy	133 (78.2)	37 (21.8)	141 (75.0)	47 (25.0)	0.78	0.42 – 1.46	141 (75.0)	47 (25.0)	31 (55.4)	25 (44.6)	0.40**	0.20 – 0.80
Confidentiality	119 (68.8)	54 (31.2)	121 (64.0)	68 (36.0)	0.80	0.45 – 1.41	121 (64.0)	68 (36.0)	28 (49.1)	29 (50.9)	0.56	0.29 – 1.06
Communication	70 (39.8)	106 (60.2)	82 (43.2)	108 (56.8)	0.73	0.41 – 1.30	82 (43.2)	108 (56.8)	22 (39.3)	34 (60.7)	0.92	0.48 – 1.77
Prompt attention	118 (66.7)	59 (33.3)	111 (58.7)	78 (41.3)	0.39***	0.22 – 0.68	111 (58.7)	78 (41.3)	23 (41.1)	33 (58.9)	0.54	0.28 – 1.02
Social considerations	119 (67.2)	58 (32.8)	123 (65.8)	64 (34.2)	0.88	0.49 – 1.58	123 (65.8)	64 (34.2)	35 (62.5)	21 (37.5)	0.83	0.43 – 1.60
Basic amenities	144 (83.7)	28 (16.3)	154 (84.2)	29 (15.8)	0.49*	0.25 – 0.95	154 (84.2)	29 (15.8)	39 (72.2)	15 (27.8)	0.42*	0.20 – 0.90
Choice and continuity	108 (61.4)	68 (38.6)	123 (66.8)	61 (33.2)	0.80	0.46 – 1.39	123 (66.8)	61 (33.2)	29 (52.7)	26 (47.3)	0.56	0.30 – 1.07

Table 5. *Continued*

	Under care of a community midwife					
	Home					
	non-referral (REF)		referral			
	(n=196)		(n=60)			
	No. (%)		No. (%)			
Responsiveness during birth	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**Adj OR**	**95% CI**
Dignity	230 (81.6)	52 (18.4)	31 (49.2)	32 (50.8)	0.20***	0.11 – 0.38
Autonomy	158 (57.2)	118 (42.8)	22 (36.7)	38 (63.3)	0.48*	0.26 – 0.90
Confidentiality	207 (74.5)	71 (25.5)	35 (57.4)	26 (42.6)	0.43**	0.23 – 0.79
Communication	176 (63.8)	100 (36.2)	22 (36.1)	39 (63.9)	0.34***	0.19 – 0.63
Prompt attention	189 (69.7)	82 (30.3)	26 (43.3)	34 (56.7)	0.32***	0.17 – 0.58
Social considerations	232 (85.9)	38 (14.1)	40 (67.8)	19 (32.2)	0.30***	0.15 – 0.58
Basic amenities	229 (86.4)	36 (13.6)	46 (78.0)	13 (22.0)	0.55	0.26 – 1.16
Choice and continuity	195 (73.9)	69 (26.1)	23 (39.0)	36 (61.0)	0.23***	0.12 – 0.42
Responsiveness postpartum	**optimal**	**non-optimal**	**optimal**	**non-optimal**	**Adj OR**	**95% CI**
Dignity	212 (75.4)	69 (24.6)	38 (61.3)	24 (38.7)	0.51*	0.28 – 0.95
Autonomy	220 (81.2)	51 (18.8)	47 (79.7)	12 (20.3)	0.97	0.46 – 2.06
Confidentiality	200 (71.7)	79 (28.3)	36 (59.0)	25 (41.0)	0.59	0.32 – 1.08
Communication	132 (47.8)	144 (52.2)	22 (36.7)	38 (63.3)	0.79	0.43 – 1.45
Prompt attention	186 (68.6)	85 (31.4)	34 (57.6)	25 (42.4)	0.63	0.34 – 1.15
Social considerations	207 (76.7)	63 (23.3)	42 (72.4)	16 (27.6)	0.62	0.32 – 1.20
Basic amenities	214 (81.4)	49 (18.6)	49 (84.5)	9 (15.5)	1.29	0.56 – 2.96
Choice and continuity	186 (70.7)	77 (29.3)	35 (60.3)	23 (39.7)	0.70	0.38 – 1.28

Non-referral as reference and adjusted for parity, education and ethnicity

* p<0.05, ** p<0.01, *** p<0.001

Responsiveness outcomes

Table 4 shows the crude and adjusted odds ratios (ORs) for each domain of responsiveness during birth and the postpartum period, according to the planned place of birth. We adjusted for parity, education and ethnicity, with the birth centre group as reference.

Among all the domains, the domains social considerations and basic amenities performed the best, followed by the domains Dignity, Confidentiality and Choice and continuity. The last domains were the domains Autonomy, Communication and Prompt attention.

No significant differences were found between the birth centre group and the hospital group under care of a community midwife.

The women who planned to give birth at a birth centre scored significantly lower on responsiveness than the women who planned to give birth at home.

A significantly higher score on the domains Dignity ($p<0.05$) and Autonomy ($p<0.001$) during birth was found for the women who planned to give birth at home. They also reported a significantly higher score on the domains Social consideration ($p<0.05$) and Choice and continuity ($p<0.05$) during the postpartum period, compared to the birth centre group.

The women who planned to give birth at a birth centre reported a significantly higher score on Dignity ($p<0.01$), Autonomy ($p<0.05$), Confidentiality ($p<0.05$) and Social considerations ($p<0.01$) during birth compared to the hospital group under supervision of an obstetrician. They also reported a significantly higher score on Dignity ($p<0.05$), Autonomy ($p<0.01$) and Basic amenities ($p<0.05$) in the postpartum period.

Referrals

Table 5 shows the adjusted odds ratios of the referred and non-referred group for each domain of responsiveness during birth and the postpartum period. The reported scores were higher for the women who were not referred. The women who planned to give birth at a birth centre and who were not referred reported a significantly higher score during birth on all the domains except for Confidentiality, compared to the referred women in this group. The non-referred women reported also a significantly higher score on Dignity ($p<0.05$), Prompt attention ($p<0.001$) and Basic amenities ($p<0.05$) in the postpartum period.

The women who planned to give birth under care of a community midwife in a hospital and were not referred reported a significantly higher score on all domains during birth except Basic amenities, compared to the referred women in this group. Their score during the postpartum period was also significantly higher on the domains Autonomy ($p<0.01$) and Basic amenities ($p<0.05$) compared to the referred women in this group.

The women who planned to give birth at home and were not referred reported a significantly higher score on all the domains except Basic amenities during birth and only on Dignity ($p<0.05$) in the postpartum period, compared to the referred women.

For the women who planned to give birth in a hospital under supervision of an obstetrician no distinction between referred or not referred can be made, because they all have been referred during pregnancy.

Birth centre services

Table 6 shows the experiences of the respondents with the birth centre services. Most of the women who received care in a birth centre assessed the homelike environment (81.3%), hotel service (84.2%) and bath (94.8%) as good. More than 40% of the women reported that they did not use wireless internet although it was available.

Almost all the women (93.0%) reported that the birth centre experiences met their expectations. 84.9% of the women arrived and 84.7% of the women left the birth centre on their preferred time. However, 13.6% of the women preferred to arrive earlier. Most of the women who were referred from a birth centre to the obstetric unit did not evaluate the change of room (81.5%) or caregiver (81.8%) as a problem. None of the women who stayed postpartum in the same room as during birth found it a problem. As few as 8.6% of the women evaluated the postpartum stay in a different room as a small problem.

Table 6. Experiences with birth centre services.

Facilities	**good**	**sufficient**	**insufficient**
Homelike environment	156 (81.3)	32 (16.7)	4 (2.1)
Hotel service	123 (84.2)	20 (13.7)	3 (2.1)
Bath	91 (94.8)	4 (4.2)	1 (1.0)
Expectations	**good**	**sufficient**	**insufficient**
Met	185 (93.0)	13 (6.5)	1 (0.5)
Moment	on time	too late	too early
Arrival	169 (84.9)	27 (13.6)	3 (1.5)
Departure	166 (84.7)	13 (8.7)	17 (6.6)
Continuity	**no problem**	**small problem**	**big problem**
Change of room in case of referral	44 (81.5)	9 (16.7)	1 (1.9)
Change of caregiver in case of referral	18 (81.8)	4 (18.2)	0 (0.0)
Postpartum stay in the same room as birth	32 (100.0)	0 (0.0)	0 (0.0)
Postpartum stay in different room as birth	32 (91.4)	3 (8.6)	0 (0.0)

DISCUSSION

The aim of this study was to assess the experiences with maternity care of the women who planned birth in a birth centre compared to alternative planned places for childbirth, by using the responsiveness concept of the World Health Organization.

The women had, in general, good experiences during birth and the postpartum period. Women who planned to give birth at a birth centre reported similar experiences as those who planned to give birth at a hospital under care of a community midwife. Women who planned to give birth at home were most positive about their experiences and scored highest on the domains Autonomy and Prompt attention. A referral to secondary care had a negative effect on the experiences of women in all settings. Women who received care in a birth centre highly valued the facilities, moment of arrival/departure and continuity in a birth centre. In case of referral, the physical travel from the birth centre to the obstetric unit was not a problem for most of the women.

Strengths and limitations

This is the first study comparing the experiences of women who planned to give birth at a birth centre with that of women who planned to give birth in the three other settings in the Netherlands: under care of a community midwife in a hospital, at home and under supervision of an obstetrician in a hospital. The used questionnaire avoids any implicit or explicit preference towards the providers or organizational structures, captures the client's actual experience and is unique in the coverage of the eight responsiveness domains. Therefore, we were able to evaluate the maternity care as a whole, with its different services, professionals and time windows. The experiences (positive and negative) are allocated to the entire maternity chain and not to a specific profession or person. In addition, the present study includes a nationwide approach and high coverage of Dutch birth centres.

The analyses were performed according to the women's planned place of birth. Our information was based on the place of birth which was planned one month before the birth. For women who were referred to secondary care before the 36th week of pregnancy, their planned place of birth will by necessity be in a hospital, under supervision of an obstetrician. In general, around 15% of the women are referred during pregnancy to the second echelon after the 36th week[23]. In addition, some women are referred immediately at the onset of labour from home to the second echelon. Therefore, some of the women who planned to give birth under care of a community midwife at a birth centre or in a hospital have not actually been in these places or experienced these conditions. According to the 'intention to treat'-principle however, they should not be excluded from the analyses.

The women were asked to participate in the study by their own community midwife. Although we asked the midwife to invite every woman receiving postpartum care for participation, we have no information if this was done. Our response rate was 54.6%, which is a good response in itself but a selection bias might have occurred. We, therefore,

compared the characteristics of the respondents with those of all the women who received postpartum care from the included midwifery practices. It appeared that the respondents have characteristics (older, higher educated, more often of Dutch origin and having less interventions during birth) that are associated with a more optimal birth experience, which may have positively influenced the results[20,26,27].

Interpreting the results

The women have, in general, good experiences during birth and the postpartum period. Another Dutch study showed that the quality of care experienced by low-risk women during the entire maternity care process is high[10]. The few significant differences between the settings during birth are especially associated with the personal related domains (Dignity, Autonomy and Confidentiality). In the postpartum period, the differences are more related to the setting related domains (Social consideration, Basic amenities and Choice and continuity). Although most differences were not significant, the women in the birth centre group have on most of the domains slightly better experiences compared to the women in the hospital group under care of a community midwife. More than 80% of the women who received care in a birth centre highly valued the facilities, the moment (on time) of arrival and departure and the continuity in the birth centre. This is in line with what several other international studies have found[12-14].

The women who planned to give birth at home have significantly better experiences than the group of women who planned to give birth at a birth centre. This is in line with what other international studies have found and can possibly be explained by the positive influence of the familiar environment at home[16,28]. Another study which compared the experiences of women giving birth in a birth centre and at home, did not find differences on overall satisfaction[15]. That study included only one specific birth centre. We found that the women in the birth centre group have significantly better experiences than the group of women who planned to give birth under supervision of an obstetrician in a hospital. This is not surprising, since it is known that women who perceive no health problems for themselves or their baby have better experiences. The women giving birth in a hospital under supervision of an obstetrician are high-risk women and, therefore, probably more anxious or worried about their own or their baby's health[21].

Being referred during labour/birth has a negative influence on the experiences. This is in line with a study that found a significantly negative association between referral and the birth experience 10 days postpartum[29]. Another study found referral as a significant risk factor for a negative recall of birth experience in women 3 years postpartum[30]. And a cross-national study showed the negative influence of a referral as well[31]. However, there is also a Dutch study which found no association between the referral and the experience of birth three weeks postpartum[32]. Moreover, a physical transfer from the birth centre to the obstetric unit has shown not to be a problem for most of the women in this study.

Implications for practice

In the last decades, many birth centres have been established in different countries, including the United Kingdom, Australia, Sweden and the Netherlands. Although no significant differences were found between the experiences of women in the birth centre group and those in the hospital group under care of a community midwife, the following trend can be seen: the women in the birth centre group have on some domains slightly better experiences. Additionally, women highly valued the birth centre services. This should be considered in the further development of birth centres in the different countries. Given the result that the women who planned to give birth at home have better experiences than the women who planned to give birth at a birth centre, more emphasis may be put on the home-like environment in the birth centres.

Being referred to secondary care has a negative effect on the experiences in all settings. Referrals cannot always be prevented, but one possible solution might be that the community midwife or her colleague, who are familiar with the woman, continues accompanying the client. In general, priority must be given to 1) Autonomy (more specific: including the client in decision-making on pain-relief/setting of birth, acceptance of treatment refusal) and 2) Prompt attention (more specific: access for contact in all situations, waiting time for service, physical accessibility of the setting, prompt phone response).

CONCLUSION

The women had, in general, good experiences during birth and the postpartum period. The domains Social considerations and Basic amenities performed the best. The domains Autonomy, Communication and Prompt attention scored relatively lower. So, one should focus more on the latter domains.

Although no significant differences were found between the birth centre group and the hospital group under care of a community midwife, the following trend can be seen: the birth centre group report on some domains slightly better experiences. The women who planned to give birth at a birth centre reported less positive experiences than the women who planned to give birth at home. Most of the women who received care in a birth centre highly valued the services. For women who do not want to give birth at home a birth centre is a good choice, it leads to slightly better, but not significantly, experiences as a planned hospital birth.

REFERENCES

1. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Social science & medicine.* 1997;45(12):1829-1843.
2. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Social science & medicine.* 1998;47(10):1573-1588.
3. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Social science & medicine.* 2000;51(11):1611-1625.
4. Valentine N, Darby C, Bonsel GJ. Which aspects of non-clinical quality of care are most important? Results from WHO's general population surveys of "health systems responsiveness" in 41 countries. *Social science & medicine.* 2008;66(9):1939-1950.
5. Oppenheimer C. Organising midwifery led care in The Netherlands. *BMJ.* 1993;307(6916):1400-1402.
6. Mander R. The relevance of the Dutch system of maternity care to the United Kingdom. *Journal of advanced nursing.* 1995;22(6):1023-1026.
7. Bradley PJ, Bray KH. The Netherlands' Maternal-Child Health Program: Implications for the United States. *Journal of Obstetric, Gynecologic, & Neonatal Nursing.* 1996;25(6):471-475.
8. Johnson TR, Callister LC, Freeborn DS, Beckstrand RL, Huender K. Dutch women's perceptions of childbirth in the Netherlands. *MCN: The American Journal of Maternal/Child Nursing.* 2007;32(3):170-177.
9. De Vries R, Wiegers TA, Smulders B, van Teijlingen E. The Dutch Obstetrical System. *Birth models that work.* 2009:31.
10. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC pregnancy and childbirth.* 2009;9(1):18.
11. Wiegers T, de Graaf H, van der Pal K. De opkomst van geboortecentra en hun rol in de zorg. *Tijdschrift voor gezondheidswetenschappen.* 2012;90(8):475-478.
12. Waldenström U, Nilsson CA. Women's satisfaction with birth center care: a randomized, controlled study. *Birth.* 1993;20(1):3-13.
13. Coyle KL, Hauck Y, Percival P, Kristjanson LJ. Ongoing relationships with a personal focus: mothers' perceptions of birth centre versus hospital care. *Midwifery.* 2001;17(3):171-181.
14. Coyle KL, Hauck Y, Percival P, Kristjanson LJ. Normality and collaboration: mothers' perceptions of birth centre versus hospital care. *Midwifery.* 2001;17(3):182-193.
15. Borquez HA, Wiegers TA. A comparison of labour and birth experiences of women delivering in a birthing centre and at home in the Netherlands. *Midwifery.* 2006;22(4):339-347.
16. Cunningham JD. Experiences of Australian mothers who gave birth either at home, at a birth centre, or in hospital labour wards. *Social Science & Medicine.* 1993;36(4):475-483.
17. Valentine NB, de Silva A, Kawabata K, et al. Health system responsiveness: concepts, domains and operationalization. *Health systems performance assessment: debates, methods and empiricism.* 2003:573-596.
18. Murray CJL, Frenk J. A framework for assessing the performance of health systems. *Bulletin of the world Health Organization.* 2000;78(6):717-731.
19. Valentine NB, de Silva A, Kawabata K, et al. Health system responsiveness: concepts, domains and operationalization. *Health systems performance assessment: debates, methods and empiricism.* 2003:573-596.
20. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the dutch obstetric care system. *BMC health services research.* 2014;14(1):622.

21. Scheerhagen M, van Stel, H.F., Birnie, E., Franx, A., Bonsel, G.J. Measuring client experiences in maternity care under change: development of a questionnaire based on the WHO responsiveness model. *Plos one.* 2015.
22. Hermus MAA, Wiegers TA, Hitzert MF, et al. The Dutch Birth Centre Study: study design of a programmatic evaluation of the effect of birth centre care in the Netherlands. *BMC pregnancy and childbirth.* 2015;15(1):148.
23. The Netherlands Perinatal Registry. *Perinatal care in the Netherlands 2013 (in Dutch: Perinatale zorg in Nederland 2013).* Utrecht2014.
24. Spss I. IBM SPSS statistics version 21. *Boston, Mass: International Business Machines Corp.* 2012.
25. Central Committee on Research Involving Human Subjects. http://www.ccmo.nl/en/ccmo-directives. Accessed June 13th 2015, 2015.
26. Raleigh VS, Hussey D, Seccombe I, Hallt K. Ethnic and social inequalities in women's experience of maternity care in England: results of a national survey. *Journal of the Royal Society of Medicine.* 2010;103(5):188-198.
27. Liabsuetrakul T, Petmanee P, Sanguanchua S, Oumudee N. Health system responsiveness for delivery care in Southern Thailand. *International Journal for quality in health care.* 2012;24(2):169-175.
28. Lock LR, Gibb HJ. The power of place. *Midwifery.* 2003;19(2):132-139.
29. Kleiverda G, Steen AM, Andersen I, Treffers PE, Everaerd W. Confinement in nulliparous women in The Netherlands: subjective experiences related to actual events and to post-partum well-being. *Journal of Reproductive and Infant Psychology.* 1991;9(4):195-213.
30. Rijnders M, Baston H, Schönbeck Y, et al. Perinatal factors related to negative or positive recall of birth experience in women 3 years postpartum in the Netherlands. *Birth.* 2008;35(2):107-116.
31. Christiaens W, Gouwy A, Bracke P. Does a referral from home to hospital affect satisfaction with childbirth? A cross-national comparison. *BMC health services research.* 2007;7(1):109.
32. Wiegers TA, Zee Jvd, Keirse MJNC. Transfer from home to hospital: what is its effect on the experience of childbirth? *Birth.* 1998;25(1):19-24.

9

General Discussion

AIM OF THIS THESIS

This thesis focuses on the development and validation, the determination of discriminative power, and the implementation and application for quality improvement of ReproQ, a questionnaire of client experiences in maternity care. The ReproQ evaluates the responsiveness of the Dutch maternity care system. According to the World Health Organization's (WHO) definition, responsiveness is defined as the way a client is treated by the professional and the environment in which the client is treated from the client's perspective[1-3].

The development of ReproQ was initiated in the context of increasing awareness of the unsatisfactory performance of the Dutch maternity care system. Based on these findings, the National Committee on Perinatal Care established by the Ministry of Health recommended in 2010 the implementation of several changes in the maternity care system[4]. One of the recommendations was to set-up perinatal units (in Dutch: VSVs, verloskundige samenwerkingsverbanden) with a focus on integrated care[4-6]. This and other organizational innovations, e.g. birth centers, necessitated the evaluation of the maternity care system in terms of health outcomes and responsiveness[7].

Central to this thesis is the measurement of responsiveness of the Dutch maternity care system. However, accepted and validated instruments to evaluate responsiveness did not exist[8-11]. For this reason, the ReproQ was developed. The anticipated use of ReproQ was primarily for monitoring and improving responsiveness, but also to support research on the effectiveness and inequalities of the system.

MAIN FINDINGS

Content and construct validity

Clients and health care professionals supported the relevance and content validity of the Responsiveness-domains. The item coverage was judged satisfactory. Moreover, our analysis supported the assumed domain structure. Since the instrument by design avoids any preference toward a specific health care professional or specific organizational structure, it appeared highly suitable to evaluate the current transition of two-tier Dutch maternity care system towards integrated care.

Feasibility: response rate and fill-out time

The response rate was moderate in absolute terms, but comparable to conventional routine measurement of client experiences: 31% of the invited pregnant women responded and 39% of the women who had recently given birth. The socio-demographic characteristics of our sample were overall representative, except for women under 24 years of age[12], and for women with a low educational level[13] and women with non-Western background[12] who all were slightly underrepresented. The following care characteristics were representative

compared to all Dutch pregnant women: setting continuity (i.e. being referred during childbirth), proportion of interventions (induced labor, cesarean section rates) and the perceived health outcome of mother and baby[12]. In our sample, women who did not use pain medication during childbirth were slightly overrepresented. Average time to fill-out the postnatal ReproQ (including the additional modules) was 14 minutes (95% CI: 11 – 17min) based on 433 respondents.

Test-retest reliability of the postnatal ReproQ

The test-retest reliability of the measurement of experiences during birth and postnatal care was good (ICCs of 0.78 and 0.74 respectively; partly unpublished data). Test-retest reliabilities of experiences during early and late pregnancy were not evaluated, but the correlation between the early and late pregnancy was high (ICC=0.80).

Minimally Important Difference (MID) of the experiences during birth

We explored differences between units and subgroups of clients using two methods: the first method defines 'difference' in terms of a statistically significant difference, which is the dominant approach in literature and practice. The second method defines 'difference' in terms of a so-called relevant difference to the client. The relevant difference was estimated using the Minimally Important Difference (MID), which was 9% for having at least one negative experience and 0.10 for the total mean experience during birth. These MIDs varied between the summary scores and the domain scores. Surprisingly, the magnitudes of the anchor-based MID and the distribution-based MID of the mean summary scores were comparable.

Determinants in case mix correction, and determinants of a low client experience score

A fair ranking of perinatal units requires that case mix correction should be applied if a determinant influences the ReproQ experience scores, if the same determinant is distributed unequally across units, and if it is beyond the health care professionals' influence[14].

Determinants of the ReproQ domain and summary scores that qualify for case mix correction are the client's socio-demographic characteristics, and the client's perceived health outcome of mother and child.

Our findings also imply that a low experience score can be assigned to variation in specific subgroups of clients, i.e., women with low antenatal experiences, women who did not experience professional and setting continuity, and women who did not have expectations towards the place of birth. As these determinants could be influenced by the health care organizations, these determinants do not qualify for case mix correction. These determinants can, however, be used to define the subgroups that qualify for quality improvement. The domains that need improvement are discussed in 2.6.

Ability to identify best practices and underperformers in a benchmark

The ReproQ domain and summary scores were able to distinguish best performing from underperforming units, and identify differences in experiences between subgroups of clients, using both the statistically significance approach and the MID-based relevance approach. This result supports the good discriminative power of ReproQ. However, the identification of the underperforming units depends not only on the approach chosen, but also on the selected reference point and criterion.

Considering stage 2 of the benchmark, the ReproQ was able to identity differences in domain and item scores among units. Therefore, ReproQ is suitable for benchmarking under routine conditions.

Measurement of antenatal experiences after birth

The association between the antenatal experiences measured before and after birth was moderate. The measurement after birth was systematically influenced by adverse previous experiences during childbirth and postnatal care and the lack of professional continuity. This finding supports our approach that client's antenatal experiences should be measured before instead of after birth, as has been done in other evaluations[8-11].

Suitability of ReproQ for quality improvement in practice

Chapter 7 illustrates how the results of the ReproQ can help to support quality improvement. ReproQ domain and item scores below a certain threshold, indicating underperformance, were identified. Next, clients and health care professionals made recommendations to improve these underperforming domains and items, prioritized these recommendations and eventually consented on an improvement agenda. Chapter 8 illustrates that the ReproQ can be used successfully as an evaluation instrument to measure the service delivery of birth centers, one example of organizational change in perinatal care. Both studies illustrate that ReproQ is suitable in practice for quality improvement.

SCOPE OF REPROQ

We made the following design choices when developing ReproQ:

1) ReproQ is based on the universal WHO responsiveness model; and consequently should cover the experiences with health service delivery.
2) ReproQ should avoid any implicit preference toward specific providers, professionals or organizational structures; this is also a point of departure of the WHO concept;
3) ReproQ covers the experiences, during pregnancy, birth and postpartum period.

Below we will discuss the operationalization each of these choices.

WHO Responsiveness model

WHO defines 'responsiveness' as the way a client is treated by the professional and the environment in which the client is treated as seen from the client's perspective[3,15]. This definition stays close to the health care encounter of the client, the health care professional and the system. Responsiveness is clearly linked to the Human Rights Declaration: human rights can never be made subject to external goals, in the same way that essential service quality aspects may not be subjected to the maximizing medical outcomes[1-3].

The WHO operationalized responsiveness into four professional related domains, and another four setting related domains. The model has been extensively validated in WHO member countries' health systems and is able to compare client experiences with care within and between countries[3,15]. Below we address the domains and items we selected for ReproQ.

During development, clients (pregnant women and woman who had recently given birth) as well as health care professionals judged all the WHO domains as relevant for Dutch maternity care. These groups judged that neither any of the domains should be excluded; nor that the model should be extended with an additional domain. We therefore conclude that the eight ReproQ domains can be considered to have content validity.

Given the domains, the content validity of the individual items was judged as good, as most comments on the test version of ReproQ concerned the clarity of wording and the relevance of the items. Although there was some debate about the content of several items, the structure and domains of the Responsiveness model were never challenged. Exploration of the psychometric characteristics of the concept-ReproQ showed that the construct validity was overall good, except for the two-item Confidentiality domain. Two different problems were encountered: one item could not be answered by 44% of the clients; the second item was assigned to the domain Dignity due to the Exploratory Factor Analysis. This resulted in the redesign of the Confidentiality domain. After improvement, still 20% of the women could not answer the question whether or not her health professionals discussed her medical condition with her family only with permission. This suggests that at least part of the clients are unaware of possible violations of their confidentiality.

Altogether, our analyses so far support that the operationalization of the ReproQ, with the WHO Responsiveness model as concept, is valid.

Neutral with respect to specific professional and organizational structure

The design choice towards neutrality with respect to professional and organizational structure is based on two observations: 1) it is the responsiveness of the health system that is under review, rather than the responsiveness of performance of an individual provider or an individual health care organization; and 2) no organizational set-up is better or worse in terms of responsiveness than the client experiences it evokes. When system performance is suboptimal and quality improvement is required or desired, neutrality with respect to

professional and organization requires that further analyses are needed to explore the determinants of this suboptimal performance. The provider, organizational setting and client's socio-demographic characteristics are all part of these determinants.

Additionally, one should be aware that the principle of neutrality with respect to specific professional and organizational setting is thought to make ReproQ future-proof, as it fits the current transition of the two-tier Dutch maternity care system towards integrated care[4-6,16,17]. In this transition, professionals' tasks are reallocated across health care providers, and new types of birth settings and disciplines emerge (e.g. the clinical midwife) as well as new types of cooperation and integrated financial arrangements. Due to the principle of neutrality, ReproQ may be used to study maternity care system performance as it currently is, during transition. Moreover, neutrality in combination with the generic Responsiveness model also allows for international comparison with other maternity care systems. The principle of neutrality has the advantage that none of the clients is asked to judge her experiences a specific professional (person, discipline) or organizational setting. This is contrary to the precursor of the ReproQ that explicitly took different experiences with different caregivers into account in a birth center context[18]. Asking clients to report their experiences with each of these professionals separately multiplies the length of the questionnaire considerably, impacting the response and completion rates, and the costs.

The neutrality principle also acknowledges that in complex settings with many healthcare professionals involved, the individual performance of a health care professional partially depends on the performance of others. The same argument applies to specific organizational structures, when women receive care in multiple organizational settings.
In summary, we recommend to maintain the principle of neutrality with respect to professional and organizational structures.

Measuring client experiences: which reference period?

A third design choice is the reference period, which the client should take into account thinking about her experiences. ReproQ deliberately covers the client experiences during early and late pregnancy as well as during childbirth and the postpartum period. Despite the empirical challenges of such a long time span we deliberately did so for two reasons.

First, the principle argument is that in Western health systems the maternity care period starts with early antenatal care and ends with postpartum period. Preconception care and fertility care are at this stage not an accepted integral part of this entity. This is also the view of the National Committee on Birth Care[4]. We are reluctant to include the preconception phase, as it varies in duration (from unplanned pregnancy to years of trying and intervening) and type of interventions (ranging from no contact with health care professionals, seeking information on preconception risks, receiving lifestyle advices, to IVF-treatments), and does not necessarily results in an ongoing pregnancy[19-22].

Second, the current and financing system, and the anticipated integral reimbursement system in the Netherlands, cover these three reference periods. Preconception care, fertility care, and preventive child health care are excluded.

At the opposite end of the spectrum is child health care. During the 4 months after birth, care for mother and child is very different: the focus is on the child rather than the mother, and on behavioral issues rather than health, and there are no working relations with the clinical setting and activities bear a strong procedural rather than outcome orientation. Few clients will regard these GPs and professionals from well-baby clinics as equivalent to medical health care professionals even if the case of medical problems. Moreover, beyond four months, extension of ReproQ seems difficult to defend as care is weakly connected to childbirth and postnatal maternity care[23]. Perhaps one exception is the small percentage of cases that are (very) prematurely born, congenital diseased, or with severe maternal complications) that receive care beyond this period.

The next question is whether ReproQ in terms of contents and measurement schedule should differentiate between early antenatal, late antenatal, birth, and postpartum. While one survey comprising all four reference periods in one appears attractive, this in our view makes no sense; neither to the client (different professionals and setting, and varying nature of adverse events and treatments), nor to the professionals who want to improve services. As an average encounter over the phases of maternity care is impossible to judge in terms of experiences, we developed two versions of the questionnaire (early and late pregnancy, and childbirth and postnatal period). For each antenatal item a parallel postnatal item was phrased, tailored to the changed setting.

In retrospect, we propose one amendment related to the initial separation of the reference period into four phases. Despite the obvious differences between early and late antenatal care, the antenatal experiences with care during early and late pregnancy were surprisingly highly correlated: this was true for the total, summary and domain scores, for the negative score (absolute agreement of 92%) as well as for the mean score (ICC of 0.83). Weighing the information gain and added value, we recommend to remove the early and late pregnancy distinction and either join both into one reference period with one response, or use the aggregated score of the early and late responses.

A similar combination of the clients' experiences with childbirth and postnatal care is not recommended. The time frame and service delivery in both phases are quite different, and clients experiences are rather different, as the measures of association indicate: the ICC is 0.58 for the mean score, and absolute agreement is 65% for the negative score (unpublished data from Chapter 4).

In summary, the best aggregation for benchmark purposes is division into the three reference periods (antenatal, birth and postpartum period), with all the derived measures (3*8 domain scores, 3*2 summary and 3 total scores).

DATA COLLECTION

Data collection covers the definition of the reference and target populations; the process of data collection; sending the questionnaire (administration mode and timing); establishing of the response rate; and the determination of representativeness of the data.

Reference and target population

The reference population was the population of pregnant women and women who have recently given birth. Except for availability of email address and informed consent, there were no particular inclusion or exclusion criteria for women to fill out the questionnaire; all women could participate (translations in English and Polish were available). Hence, the reference and target populations are the same.

We also invited women whose baby past away around birth, women who had a miscarriage, abortion or IUVD, as these women also gave birth and often received some kind of postnatal care. Although the content of the questionnaire should not be adapted for these specific subgroups, we recommend rewriting the introduction of the questionnaire reckoning with their loss.

The reference and target populations were pregnant women and women who have recently given birth. Although we argue that the mother is the principle bearer of experiences with care, one may consider measuring the father-to-be's experiences also. We assume that the fathers are often included in decision-making at home during antenatal care; they play a role during birth, and also have to be empowered during postnatal care. In short, the father could be regarded an indirect or secondary client, whose experiences could be weighed with that of the mother. Either fathers can receive their own invitation to fill out the questionnaire or, alternatively, each couple could fill out one questionnaire. Since the frequency and intensity of interaction with healthcare provision is often lower for fathers than mothers, and each parent may have different experiences, we recommend inviting the father (to-be) and mother (to-be) with separate questionnaires. During analyses, the response of the father should be weighted regarding the frequency of his interaction with the healthcare system, and the experienced involvement of the father rated by the mother. Alternatively, the experiences of each parent could be analyzed separately. This option appears less attractive, as the experiences of both parents are not independent but probably interact.

Process of data collection

In our studies, we implemented two different strategies of data collection.

In the first strategy, used in several perinatal units, clients were approached personally by their health care professional and given the option to fill out an informed consent form. Clients who gave consent received an invitation to respond to a paper-and-pencil questionnaire or a digital questionnaire depending on the client's wishes. In the second

strategy, clients received a digital invitation by email based on a list of clients' email addresses and supporting information (e.g. expected date of childbirth) provided by the health care organizations (in our case maternity care organizations).

In the first strategy, only 35% of the clients were initially invited to participate, as the number of consent forms returned was considerably lower than the number of clients who received care in these health care organizations. In the second strategy, all clients with an email address were invited (93%).

The low response rate of our first strategy was mainly due to two reasons. First, from the viewpoint of the health care professional, the implementation of the personal approach proved rather demanding as the optimal timing could not be connected to a routine care procedure and/or existing IT-facilities. Secondly, health care professionals later disclosed to the researchers informally that they avoided inviting specific client subgroups, among others clients who they thought were vulnerable, clients with (presumed) limited understanding of the Dutch language, teenage clients or clients with a complex medical situation. The health care professionals argued that these subgroups are reluctant towards care, and consequently that they are more likely to drop out of care when invited to participate. Considering that the proportions of clients reached were considerably larger in the second than the first approach (client list: 93% vs. personal approach: 35%), and the response rate of both strategies approaches was comparable eventually (client list: 30% vs. personal approach: 28%), we conclude that only the second strategy is the most viable option from a logistic as well as scientific perspective.

In an ideal case of integrated care, the perinatal unit should be able provide the aforementioned lists of clients, with in this context email addresses (or lists of addresses if paper and pencil is considered). The main barrier to achieve this is that perinatal units currently do not have one overall combined registry or digital administration of the clients of all participating midwifery practices, hospitals and maternity care organizations, and are legally not allowed to have that information as long as the member organizations of the perinatal unit are not fully integrated. Unfortunately, none of the individual organizations can reach all clients: all clients receive antenatal care, but not all clients receive care from a midwife (85% at the start of the antenatal check-ups), a gynecologist or a hospital (70% during birth)[12]. The most promising candidates are the maternity care organizations that include about 95% of all clients[24]. Since most clients visit multiple caregivers and health care organizations antenatally and/or postnatally, the overlap of a client entered in multiple digital registries or administrations should be avoided. Technology to achieve this is already available.

There are several reasons why maternity care organizations overall appear to be the best starting point for data collection. First, maternity care organizations use well-developed digital quality systems, which are already used for routine quality measurements. These routine quality measurements rely on similar logistics and data processing as needed for

ReproQ. Secondly, almost all pregnant women antenatally book for maternity care, and more than 95% of women who delivered actually receive maternity care. Thirdly, merging the client lists of the 20 largest maternity care organizations (market share about 85%) is relatively easy, compared to the merge of the client lists of over 700 midwifery practices and about 90 hospitals. Finally, the risk of duplicates is limited, as most clients register with and receive care from only one maternity care organization. Legal barriers are absent. Maternity care organizations can invite all clients to fill out the questionnaire without consent, as long as the data are only for (internal) quality improvement. Consent is, however, required when the data are shared with associated health care organizations or within a perinatal unit, or when the data are used for other purposes then quality improvement.

Sending the questionnaire

Administration mode

For future use, we recommend only sending a digitally questionnaire for two reasons. First, almost all clients can be reached digitally since about 95% of the women have an e-mail account (personal communication H.E. Ernst-Smelt, Careyn Kraamzorg); the internet availability at home is 93% in The Netherlands[25]. Second, a digital data collection saves manpower and time compared to other forms of data collection. A digital approach may be less optimal for women who do not fully understand the Dutch language and illiterate women. An interview study on alternative questionnaire modes showed that about 55% of Non-Dutch women and/or women with a low educational level preferred a face-to-face interview, and almost 20% a self-report paper questionnaire. For these groups intended participation was 5 – 10% higher for paper and oral modes compared to the digital mode. A telephone interview was preferred in only 5 – 10% of cases[26].

We recommend translating the questionnaire in several foreign languages, as well as voice recording the questionnaire in Dutch. A multi-mode approach for non-Dutch and/or low SES groups may be considered.

Timing of the questionnaire

E-mail lists of clients are composed antenatally. Consequently, the moments of distributing the antenatal and postnatal questionnaires are based on the expected date of birth. For the postnatal questionnaire, this can result in sending the questionnaire too early (when the delivery occurs after than the expected date of birth; max. 2 weeks too early) or too late (when the true delivery occurs before the expected date of birth; max. 16 weeks too late). In the future, we prefer to have a separate client list for the postnatal questionnaire, so sending the postnatal questionnaire can be based on the actual date of birth.

The optimal moment to send the postnatal questionnaire is unknown. Sending the questionnaire later than six weeks after birth could result in recall bias due to exposure to other influences (e.g. women return to work, assuming their usual habits and patterns),

and/or in non-response because sharing one's birth experiences at that time may seem less relevant. Sending the questionnaire earlier than six weeks after is not necessarily a better option. It may result in better recollection of the experiences but the risk of mood swings and hormonal disturbances might affect responses and response rates. Practically, sending the questionnaire two weeks after birth concurs with the current quality improvement cycles of the maternity care organizations, which increases their support to participate. A recent (unpublished) study that used the ReproQ at 1 – 2 weeks after childbirth showed higher response rates (45 – 50%) but also lower ReproQ summary scores. This suggests that timing may affect response rates as well as outcomes.

Establishing the response rate

Apart from representativeness (see 4.4 below), our overall response rate of 39% at 6 – 8 weeks after childbirth is comparable to other client experiences surveys used in scientific studies or by government organizations[9,27-30], but lower than the 60 – 70% response rate of routine client evaluation surveys held by maternity care organizations at 1 – 2 weeks after childbirth (personal communication J. Dorscheidt, Kraamzorg De Waarden). Response rates are difficult to compare, however, due to the different timing and lengths of the questionnaires.

The following measures could be taken to increase the response rate. Firstly, considering the perceived relevance, health care professionals and organizations can stimulate the client to fill out the questionnaire through a personal approach[31-33]. E.g., health care professionals could 'announce' the questionnaire during consultation hours or at the end of the postnatal care, emphasize importance and relevance, and motivate clients to fill out the survey. Related to this, they survey should be sent by an organization which is known personally to the client, instead of an 'anonymous' research group or governmental organization. Secondly, reminder strategies may help: more reminders could positively impact the response rate[26]. Thirdly, multi-mode approach (see 4.2.1) and translation of ReproQ in own language should be considered. Fourthly, the response rate could benefit from a reduction of the many of medical and non-medical surveys that women receive after childbirth. Finally, to increase the number of clients that complete the questionnaires, the burden of filling in the additional questions (e.g. socio-demographics, and information of the care process) could be reduced once that information can be retrieved from other sources.

Representativeness

A high response rate is not an aim in itself, but it enhances representativeness. For a valid quality improvement cycle the sample should also be representative. If response is low, but non-selective, power or reliability is impaired but the validity is not. Even if particular subgroups are under- or overrepresented this only hampers the representativeness when

the under- or overrepresented groups also have substantially different ReproQ scores. The impact of under- or overrepresentation can be statistically adjusted.

A representative sample can be realized by one of the following strategies: First, a perinatal unit can invite all suitable clients and weight the results of the subgroups that are under- or overrepresented in retrospect. A second possibility is to weight the proportions of invited clients with the response rates in the under- or overrepresented subgroups. This requires that data on the degree of under- or overrepresentation is available for each subgroup.

Representativeness in our studies was checked by the comparison of our sample with the national average[12]. Regarding clients' socio-demographic characteristics the following subgroups were underrepresented compared to the national average: clients younger than 24 years of age (5% vs. 11%)[12], women with a non-Western background (7% vs. 14%)[12], and women who have a low educational level (8% vs. 18%)[34].

It is most likely that the underrepresentation is caused by the lack of perceived control of these women, which is reflected in reluctance to participate: they do not believe that participation or responding matters[35], and a language barrier[36,37]. Sample variation resulting in a low prevalence in the case-mix of the participating units is less likely, as our data collection covered over 2/3 of the perinatal units. Equally unlikely is that these women did not complete all ReproQ items, as only 15% of the clients prematurely dropped out. Our analyses showed, however, that the experiences of the underrepresented subgroups were similar to the average ReproQ scores. In short, our sample appears representative but may be slightly underrepresented for low age women, and women with non-Western background and/or low educational level.

Representativeness was further checked regarding the process of care and outcome. The following characteristics were representative for all Dutch pregnant women: setting continuity (i.e. being referred during childbirth), proportion of interventions (induced labor, cesarean section rates) and the perceived health outcome of mother and baby (compared with postnatal hospitalization rates in the National Perinatal Registry[38]). In our sample, women who did not use pain medication during childbirth were slightly overrepresented (61% vs. the national average of 51%)[12]. As the use of pain medication has no influence on the clients' experiences during birth[12], we believe that the overrepresentation has not biased the average ReproQ scores. Reference data regarding other care characteristics are absent: professional continuity, degree to which expected place of birth is realized, day and time of birth, and presence of adverse outcome in a previous pregnancy.

Our recommendation is to strive at representative data for the care process and perceived health outcome, since these characteristics have a relatively high impact on the ReproQ scores. Although our data appear representative at the national level, this does not guarantee that the data collected by individual perinatal units is also representative.

BENCHMARKING

Units that aim to improve care can participate in a benchmark. The first stage of a benchmark is to rank perinatal units according to responsiveness and to identify underperformers. Identification of underperforming units is based on two choices: 1) the reference chosen, for which two options are available: mean performance score of all units or mean performance score of a subset of best performing units; and 2) the difference or threshold criterion that separates 'poor' from 'at least satisfactory' performance, for which also two options are available: 1) one based on a relevant difference and 2) one based on a statistically significant difference. These choices are discussed in section 5.1.

Section 5.2 addresses a third choice: the outcome measure of interest. We contrast the mean score and the so-called negative experience score. The remaining sections address the process to arrive at improvement measures (section 5.3) and the requirements to achieve a full benchmark (section 5.4.).

Identifying underperformance: reference and criterion

The usual approach in a benchmark is to rank units relative to the grand mean of all units as reference[39-44]. In Chapter 5 we presented an alternative reference: the pooled average of the 10% best performing units. A disadvantage of both approaches is that it does not challenge the average and best performing units to improve their care; these units need an alternative source of information to define their goals for quality improvement. Disadvantages of the alternative approach, a group of best practices as reference, are that the definition of 'best' performing units inevitably is somewhat arbitrary, and that the subgroup of 'best units' should not be too small (say about 5 units) in order to produce a stable subgroup means. The confidence interval of the mean of this subgroup partially depends on sample size and therefore on the number of units.

The next step is to define the criterion to identify the poor from the average and best performing units. The conventional criterion is based on the 95% confidence interval (CI) of the reference. In practice, the upper limit of the 95% CI of a unit's mean performance is tested against the lower limit of the 95% CI of the average reference. If the upper limit of a unit does not overlap with the reference point, a unit is categorized as underperforming unit, assuming higher scores to indicate better performance. Conversely, when the lower 95% CI limit of a unit does not overlap with the reference point, the unit is categorized as best practice[39-41,43,44]. The ability to identify underperforming units according to this criterion is strongly influenced by the 95% CI of each unit, which consequently depends on the unit's sample size (primarily a researcher's choice) and the true variance of the experience scores in this unit (i.e. response heterogeneity) which is a unit characteristic rather than a choice or error).

In Chapter 5 we estimated that an average unit requires at least 432 completed ReproQs (childbirth phase) to achieve reliable estimates for the total ReproQ score; the summary

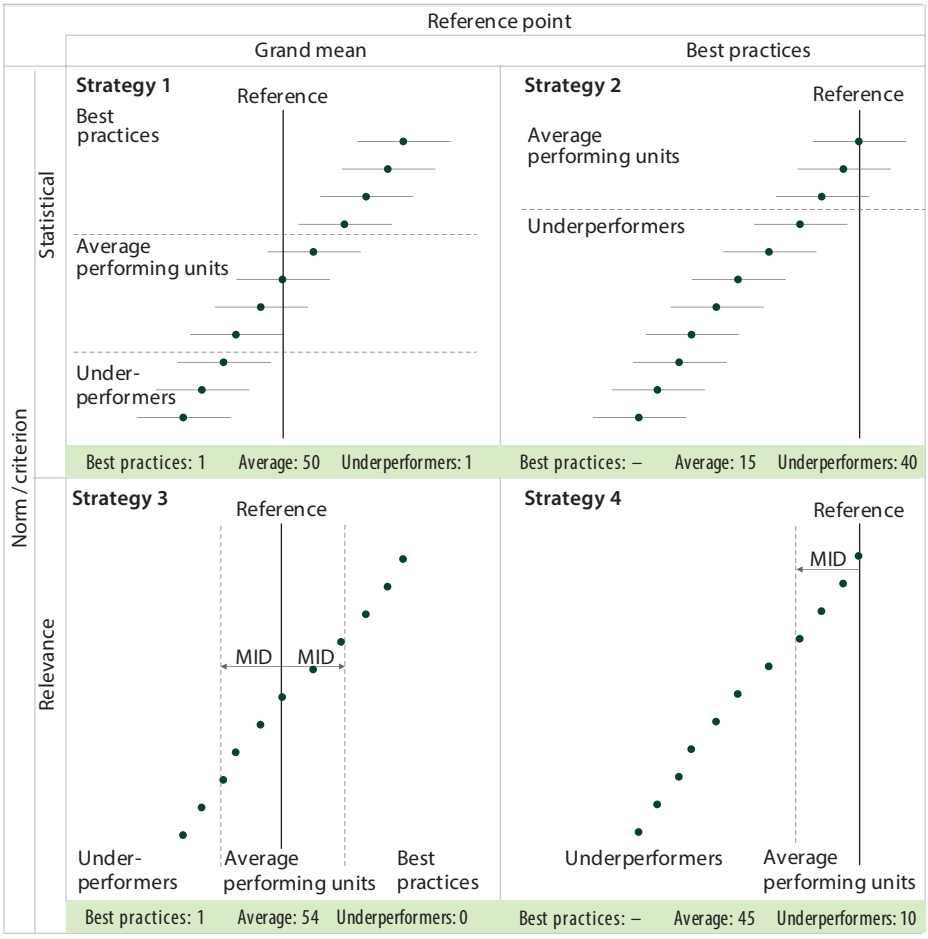
scores require at least 465 filled out ReproQs. As the average number of clients per perinatal unit is about 2000 and the valid response rate is about 30 – 40%, these numbers may be hard to meet on an annual base. Insufficient numbers do not affect the validity of the criterion or reference measure selected, nor the ranking of perinatal units; but it will inflate the units' 95% confidence intervals of the mean experience scores and categorize part of the underperforming units as 'average'.

In search of discrimination with more power, we explored an alternative approach for the criterion to discriminate good from poor performing units. Instead, we investigated the concept of a minimally important (or relevant) difference (MID). A MID in medical research refers to a minimally relevant difference on the patient (client) level[45-48]. Applied to the unit level, a unit is categorized as best- or underperformer when the difference between the unit's mean score and the reference point (grand mean or otherwise) exceeds the MID. A magnitude of 1.0 MID at the unit level means that *all* clients cared for in that unit differ, *on average*, one MID from a reference value, either being much better (best-performing unit) or much worse (underperforming unit). This is an extreme difference. We regard a magnitude of 0.5 MID a more feasible but still a large difference.

The statistical significance criterion and the relevance-based criterion complement each other. The statistical significance criterion focuses on the ability to separate groups with 95% certainty, whereas the MID-based approach focuses on the relevance of the size of the difference. As the relevance approach better matches the desire for client-centered care and avoids problems related to lack of power, we recommend using the MID-based criterion for the identification of underperforming units.

The combination of selected reference point (grand mean, mean of 'best' units) and selected criterion of deviation (95% CI, relevance-based MID) results in four different strategies to identify underperforming units; See Figure 1. In Chapter 5 we identified underperforming units on the basis of strategy #1 (grand mean and 95%CI) and strategy #4 (best 10% of units and MID) using the clients' total experience score during childbirth as outcome measure. With the grand mean as reference, only one unit is underperforming according to the statistical significance approach (strategy #1) and no unit according to the MID-based approach (strategy #3). In contrast, when the mean of the 10 best units is applied as reference, 40 units are identified as underperforming in the significance based approach (strategy #2) and 10 units in the MID-based approach (strategy #4). The different results underline that the selection of the selected reference point and criterion matter for the number of number of identified underperforming units. Based on the considerations discussed above, we slightly prefer strategy #4, with 0.5 MID as criterion.

Figure 1. Four strategies to identify underperforming perinatal units, depending on the selected reference point (grand mean of all units (Strategies #1 and #3, on the left) or mean of 10 best units (Strategies #2 and #4, on the right)) and the selected criterion of deviance (upper limit of 95% CI (Strategies #1 and #2, at the top) or mean score beyond MID (Strategies #3 and #4, at the bottom)).



Identifying underperformance: mean versus negative experience score

Conventional benchmark procedures take the grand mean of the units as outcome measure[14,44]. We explored the discriminative ability of an alternative definition of the outcome measure using the same ReproQ data as in Chapter 5: the negative experience score, i.e. the presence of any experience item (among all items relevant for a sum score) scored below a threshold. Due to the favorable measurement properties of this dichotomous score, we used it as an alternative outcome measure in most of our analyses.

Below we will discuss the definitions and the conceptual differences between the mean and the negative experience scores (section 5.2.1.), followed by a head-to-head comparison (section 5.2.2.).

Mean versus negative experience score: concept

The 'mean domain score' is defined as the unweighted mean score of items for each domain, treating the item response categories numerically. The summary scores are not the mean of all items included in the domains, but the overall mean of the mean domain scores included in that summary measure. For its calculation, each domain has the same weight, even if the domains consist of different numbers of items.

The negative experience score rests on a different principle. A 'negative' experience was defined here as ticking the response 'never' in at least one of the items of a domain (where 'never' indicates a very poor experience), and/or ticking 'sometimes' in a domain that the client identified as being the most important. This creates a personalized negative experience score. After aggregation, the prevalence of negative experiences at the unit level is the percentage of women with at least one negative experience on the domain level or summary level (personal, setting, total score).

The negative experience score and the mean score differ in two respects. Firstly, the negative score reflects the presence of an answer below a certain threshold, regardless of the negative or positive answers to the other items. Typical of the negative score is that positive experiences cannot compensate for negative experiences, and that the expected probability of a negative experience score increases from the domain to the summary level as the number of items increases.

In contrast, the mean score allows the compensation of negative experiences with positive experiences, and the expected score does not change with the number of items or domains included. We believe that quality improvement is most efficient when the focus is on those aspects and units where it is most needed. For this, the percentage of negative experiences appears the obvious outcome measure. Both outcome measures have the disadvantage that they must be decomposed to learn which items or domains are responsible for the suboptimal scores. The decomposition is more laborious for the mean score where positive and negative experiences may compensate. Whatever outcome measure chosen, measures for quality improvement may be different in these two cases.

Secondly, the mean score is an unpersonalized score as it is conventionally unweighted, while the negative score can be easily personalized using individual client's preferences on the importance of the domains. We believe it is justified to restrict the use of 'never' as measure of suboptimality in any domain and restrict 'never' and 'sometimes' as a negative experience to domains that a respondent assigns as important. Consequently, the domains that a client identified as the most important were valued the most.

As personalization of an individual score matches the desire for client-centered care, we prefer the negative experience score to the unweighted mean score. The issue of personalizing and weighting the mean score should be further explored, for example by discrete choice experimentation (DCE). If appropriately designed, DCE like questions could be added, to obtain not only a sophisticated joint preference score on the domains of service quality but also to obtain a composite preference score that joins aspects of outcome and service quality.

Mean versus negative experience score: the two-stage quality cycle

In view of the conceptual differences, one may wonder whether application of the mean and the negative experience score models affect the ranking of units (stage 1 of the two-stage quality cycle), and different abilities to identify areas that need improvement (stage 2).

Table 1 shows the 55 perinatal units from Chapter 5 being classified as best, average and underperforming units, for two outcomes measures (mean score and negative experience score of the total experience score during birth) and two criteria (statistical significance with grand mean as reference, and relevance based criterion with best practices as reference). When the significance-based criterion is used, 4/55 units (7%) are categorized underperformers according to the mean score and 13/55 (24%) according to the negative score. When the MID-based criterion is applied, 10/55 units (18%) are classified as underperforming units according to the mean score, whereas 28/55 (51%) are classified as underperforming units according to the negative score. It appears that the significance-based criterion applied to the mean score (Strategy #1) tends to classify units as being average, whereas the relevance-based MID criterion (Strategy #4) tends to emphasize units being classified as best- or underperformers.

Table 1. Discriminatory power of the mean experience score and negative score for the total experience score during childbirth (n_{pu}=55).

	Overall average	Strategy #1 reference: grand mean | criterion: statistical significance			Strategy #4 reference: best practices | criterion: relevance MID)		
		Best-practices	Average	Under performers	Mean best practices (ΔP90 – P100)	MID	Under performers (ΔP90 1MID)
Mean score	3.73	1	50	4	3.80	0.10	10
Negative score	49%	12	30	13	64%	11%	28

Table 2 shows the agreement of classifying perinatal units in terms of best, average and underperforming between the mean score and negative score according to Strategy #1. Table 3 shows the same comparison for Strategy #4.

Table 2. Classification of perinatal units according to the mean score and the negative score based on Strategy #1 (n=55 perinatal units).

		Negative score			
		Best-practices	Average	Underperformers	Total
Mean score	**Best practices**	0	1	0	*1*
	Average	12	29	9	*50*
	Underperformers	0	0	4	*4*
	Total	**12**	**30**	**13**	**55**

Table 3. Comparison the categorization of units for the mean with the categorization of units for the negative score according to Strategy #4 (n=55 perinatal units).

		Negative score		
		Average	Underperformers	Total
Mean score	**Average**	26	19	**45**
	Underperformers	1	9	**10**
	Total	**27**	**28**	**55**

The classification of perinatal units according to the mean score and the negative experience score corresponds in 60% of units for Strategy #1 and in 64% of units for Strategy #4.This implies that methods choices affect the selection of units that should improve.

The second stage of the two-stage quality cycle is aimed at selecting areas or domains that need improvement. Our results show that the mean score and negative experience score usually identify the same topics to improve (see Chapters 3, 4 and 6). Hence, both outcome models appear equally suitable from this perspective.

Overall, we recommend using the negative experience score rather than the mean score in both stages of the quality cycle for three reasons: 1) the negative experience score is better able to separate units in terms of best, average and underperforming units; 2) the negative experience score enables easy personalization; and 3) the negative score is an easy to interpret signal ('red flag') in quality improvement processes.

Stage 2. Quality improvement

In this section we address three strategies to arrive at recommendations to improve care once the performance ranking of stage 1 has been successfully completed.

Strategy 1 is 'profiling' or comparing the domain and item scores of an underperforming unit with the summary scores of (a group of) best practices that serves as *external* reference. Any outcome measure, criterion and threshold will suffice (see sections 5.1 and 5.2). An example of this strategy was outlined in Chapter 5 Appendix 2. This strategy assumes that the underperforming units adopt te policies and procedures of best practices by exchanging information. The underperforming and best performing units are aware of their ranking. The underperforming units are willing and capable to learn from the best units, and the best units are willing to share the relevant information with the underperforming units. Client interaction does not play a role. The success of this strategy depends on 1) the premise that all units' have a joint responsibility for excellent service delivery, 2) the best performing units know what explains their success, and 3) access to policies to achieve a higher rank. Stage 1 should not be experienced as a competitive analysis and best practices should not be inclined to withhold information, fearing to lose their competitive edge[49-52]. This fear could be real in a highly competitive market.

Strategy 2 assumes that all units, not only the underperforming ones, should improve quality. This could be achieved either on the basis of an *internal* reference or *external* reference. Strategy 2 may be valid even if ranking information form stage 1 is absent. Consulted experts may inform on the causes of suboptimal ReproQ domain and item scores and recommend actions for improvement. Even clients can take up the role of experts. An example of strategy 2 is outlined in Chapter 7. In that study a consensus improvement agenda was set up, but details on the implementation and impact on service delivery are not mentioned.

Strategy 3, not elaborated in this thesis but occasionally used in maternity care[53-56], is to derive recommendations for quality improvement from multidisciplinary meetings or audits, consisting of experts, on the basis of *internal* or *external* references. Strategy 3 rests on the assumption that severe adverse outcomes share modifiable factors rather than 'bad luck', i.e. coincidental occurrence of multiple risks. Strategy 3 assumes that the causes of these adverse outcome from individual cases can be generalized and help to define improvement measures. A ranking of units is not applicable because the focus is on individual cases with adverse outcomes. Clients do not play an active role in this strategy.

Despite their difference in scope and information processes, these three strategies share that health care professionals (and clients where appropriate) should reach consensus on the recommendations. Consensus is the cornerstone of continued involvement and support of professionals or stakeholders for a change[50,52]. Currently Strategy 2 is commonly used for quality improvement in practice, as the conditions for strategy 1 are often not met: the required information is unavailable and/or best practices are unwilling or incapable to share policy information.

Conditions for successful benchmarking

Before successful benchmarking can be completed, the following four conditions have to be met.

First, units should have a culture that actively supports benchmarking and quality improvement, both on the unit level as the sector level. Professionals should feel safe to discuss imperfections, and resources should be made available to implement required quality actions. Essential part of the culture should be that all units and professionals involved acknowledge the principle of joint responsibility for a client/patient[50,52].

In the Netherlands, hospitals with their perinatal units and maternity care organizations already have implemented *internal* quality improvement cycles as part of their HKZ/ISO-certification procedures and as part of their budgetary contracts with health insurance companies. Typically these cycles mimic strategy 2. They are also obliged to participate in the national perinatal audit system (strategy 3), where priorities are set on the adverse cases to be discussed[53-56]. *External, reference guided* quality improvement cycles like strategy 1 are largely absent. As perinatal units are still legally and professionally under construction, strategy 1 could easily lead to withdrawal[49-51].

Secondly, besides a supportive culture, units should support and acclaim the performance indicators, including the outcome measure, scoring model, reference point and criterion, once chosen[51]. Indicator compliance can be difficult to achieve, as most health care professionals prefer their 'own' professional perspective, outcome indicators and norms for deviation to an integrated care perspective. They may feel disappointed with integrated indicators like ReproQ or other comprehensive measures, and may claim that the indicator does not provide relevant information for them. Actually, this reflects a lack of jointly felt responsibility.

The third condition is the presence of a permanent organisational structure above, as well as within, perinatal units. This organisation acts as information node and service center to implement improvements[50,52]. Currently, such a structure is lacking; Dutch maternity care lacks an authoritive independent party trusted by the various professions involved, despite the claim of at least two professional organizations. The current information services are limited and vulnerable for all sorts of biases. Within the hospitals no facilities exist to create the required stage 1 data for internal quality processes. One exception are maternity care organizations who routinely use detailed quality information systems (e.g. Kraamzorgkompas, MATRIQs) both at the national and the organisational level. The collection, appropriate handling and presentation of data and the detection of underperformance is demanding and costly task if every unit takes up this challenge individually. Consequently, when a benchmark is performed, most effort is often put into ensuring the validity of numbers and the production of league tables[50] rather than focussing upon how similar activities are handled by different units with different performance rank[51]. Therefore, the full potential of a benchmark is often not recognized and rarely realized[50].

The final condition to be met is that quality improvement does not rest on opportunistic cherry-picking of high policy impact-low patient relevance aspects of care. For some organizations, improvement measures are only eligible for implementation when a net financial benefit is projected (e.g. simple time saving procedures, instructions, increasing numbers of clients) or immaterial advantage is foreseen (less adverse outcomes, avoiding reputation damage).

Combining the available information on Dutch perinatal care we conclude that at this moment (July 2018), quality improvement in Dutch perinatal care is still in its infancy.

CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

From the previous chapters we derive the following conclusions:

- The WHO Responsiveness model can also be applied to maternity care. All the domains from that model were judged relevant by health care professionals and clients. Additionally, operationalization of the domains into suitable questions resulted in good content and construct validity.
- Since women's experiences with childbirth and postnatal care had a systematic and positive impact on the antenatal experiences measured retrospectively, the measurement of antenatal experiences should be scheduled before birth instead of after birth.
- The criterion chosen to define deviant performance (either relevance-based MID or the significance-based 95% confidence interval) influences the number of units that should improve their care. A homogeneous client population increases the probability of detection of either a best- or underperformer with both criteria.
- While the majority of perinatal units have comparable mean scores, confidence intervals of the unit mean vary considerably. This could point to true variation of patients with comparable performance, or to variation in performance among comparable patients. Additional analysis may suggest which one is true.
- The negative experience score rather than the mean score is superior to separate best-performing, average and underperforming perinatal units. Other method choices can also affect the ranking of perinatal units, although about 2/3s of units are equally grouped regardless of these choices.
- Clients' experiences are predominantly influenced by the care process and the self-reported perceived health outcome and to lesser extent by their socio-demographic background. The relationship between socio-demographic determinants and the mean sore is comparable to negative experience score.
- Although the majority of clients in our datasets showed good outcome and, good experiences, considerable differences existed between subgroups of clients. These outcome differences could qualify as 'health inequalities'. Its potential sources should be studied further. Inequalities related to differences in service delivery and professional factors should be minimized.

- Having extensive information on one uniformly used and validated instrument, including its weaknesses, is to be preferred to the current situation in Dutch maternity care where various instruments are defined 'valid', despite the absence of any psychometrical evidence. Undocumented instruments cannot be used in stage 1 of benchmarking, nor serve as the basis of 'profiling' in stage 2. Their use is limited to strategy 2 quality improvement.

Apart from routine quality assessment of the ReproQ, the following research questions may be addressed in future research:

- Can the ReproQ, now consisting of the questionnaire at large and supportive modules, be abridged using data extracted from other sources? What information can be reliably, completely and legally valid obtained in real time from existing sources?
- Can the mean score be personalized? How would this affect the ranking of units? If the client would be informed on the personalization of her response, would this increase the response rate?
- Is the discriminative power of the negative experience score different from the mean score?
- Is it possible to combine individual clients' experience scores, her health outcome and the fairness of financial distribution into one overall performance score? Would quality improvement profit from inserting context-specific and local specific add-ons to support quality improvement cycles instead of current audit-like procedures?
- Are client experiences (as measured with ReproQ) related to perinatal and health outcomes, such as the Adverse Outcome Index, or maternal outcome measures?

In a recent article it is postulated that there is a lack of evidence on the best way to organize birth care; that care models are being implemented without knowing the effect on perinatal mortality and morbidity and continuity of care; and that implemented care models are often fragmented, small scale or unspecific, and are hardly evaluated[57]. Therefore, it remains unclear which organizational changes are needed to achieve improved quality of birth care[57]. As shown in this thesis, the ReproQ can contribute to this aim.

REFERENCES

1. de Silva A. A framework for measuring responsiveness. *Global Programme on Evidence for Health Policy Discussion Paper 32* 2000.
2. Gostin L. The domains of health responsiveness: a human rights assessment. 2002.
3. Valentine N, de Silva A, Kawabata K, Darby C, Murray C, Evans B. Health system responsiveness: concepts, domains and measurement. In: Murray C, Evans B, eds. Geneva, Switzerland: World Health Organization; 2003:573-596.
4. van der Velden K. *Safe care during pregnancy and birth.* 2009.
5. Bonsel GJ, Birnie E, Denktas S, Poeran J, Steegers EAP. *Trends in the perinatal mortality. Description study pregnancy and birth 2010.* Rotterdam: Erasmus MC; 2010.
6. Ravelli AC, Tromp M, van Huis M, et al. Decreasing perinatal mortality in The Netherlands, 2000-2006: a record linkage study. *Journal of epidemiology and community health.* 2009;63(9):761-765.
7. Murray CJL, Frenk L. Summary measures of population health in the context of the WHO framework for health system performance assessment. In: Murray CJL, Evans DB, eds. *Health Systems Performance Assessment*. Geneva, Swiss: World Health Organisation,; 2002.
8. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483
9. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* Oxford, United Kingdom: National Perintal Epidemiology Unit;2010.
10. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg
11. Wiegers TA, Keirse MJ, Berghs GA, van der Zee J. An approach to measuring quality of midwifery care. *Journal of clinical epidemiology.* 1996;49(3):319-325.
12. PRN Foundation. *Perinatal zorg in Nederland 2013. (Perinatal care in the Netherlands 2013).* Utrecht: PRN Foundation,;2014.
13. CBS. Working population: obtained education according to personal characteristics 2001-2012 [Beroepsbevolking; behaalde onderwijs naar persoonskenmerken 2001-2012]. 2013; http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71822NED&D1=0&D2=l&D3=1-3&D4=0&D5=a&D6=0&D7=l&HDR=T,G6,G2,G5,G3&STB=G1,G4&VW=T%5D/SCP/Emancipatiemonitor.
14. Sixma H, Hendriks M, De Boer D, Delnoij D. Manual CQI development: guidelines and regulations for the development of a CQI measurement [Handboek CQI ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument]. 2008; http://www.nivel.nl/sites/default/files/bestanden/Handboek-CQI-Ontwikkeling.pdf.
15. Valentine NB, Bonsel GJ, Murray CJ. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16(7):1107-1125.
16. Mohangoo AD, Buitendijk SE, Szamotulska K, et al. Gestational age patterns of fetal and neonatal mortality in Europe: results from the Euro-Peristat project. *PLoS One.* 2011;6(11):e24727.
17. Zeitlin J, Mohangoo AD, Delnord M, Cuttini M, Committee E-PS. The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *J Epidemiol Community Health.* 2013;67(12):983-985.
18. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the Dutch obstetric care system. *BMC health services research.* 2014;14:622.

19. Shawe J, Delbaere I, Ekstrand M, et al. Preconception care policy, guidelines, recommendations and services across six European countries: Belgium (Flanders), Denmark, Italy, the Netherlands, Sweden and the United Kingdom. *The European journal of contraception & reproductive health care : the official journal of the European Society of Contraception.* 2015;20(2):77-87.
20. van der Zee B, de Beaufort I, Temel S, de Wert G, Denktas S, Steegers E. Preconception care: an essential preventive strategy to improve children's and women's health. *Journal of public health policy.* 2011;32(3):367-379.
21. van Voorst S, Plasschaert S, de Jong-Potjer L, Steegers E, Denktas S. Current practice of preconception care by primary caregivers in the Netherlands. *The European journal of contraception & reproductive health care : the official journal of the European Society of Contraception.* 2016;21(3):251-258.
22. van Voorst SF, Vos AA, de Jong-Potjer LC, Waelput AJ, Steegers EA, Denktas S. Effectiveness of general preconception care accompanied by a recruitment approach: protocol of a community-based cohort study (the Healthy Pregnancy 4 All study). *BMJ open.* 2015;5(3):e006284.
23. GGD Amsterdam. *Growth Guide 0-4 years [GroeiGids 0-4 jaar].* Amsterdam: GGD Amsterdam,; 2012.
24. Lamkaddem M, van der Straten A, Essink-Bot ML, van Eijsden M, Vrijkotte T. [Ethnic differences in uptake of professional maternity care assistance]. *Nederlands tijdschrift voor geneeskunde.* 2014;158:A7718.
25. Eurostat. Individuals using the internet by place of use. 2015; http://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tin00075.
26. Hitzert M. *The ReproQ4All. Measuring the performance of a client's questionnaire about perinatal care, based on the WHO responsiveness model, in low SES and non-Dutch ethnic groups.* . Rotterdam: Institute of Health Policy and Management, Erasmus University Rotterdam; 2013.
27. Christiaens W, Gouwy A, Bracke P. Does a referral from home to hospital affect satisfaction with childbirth? A cross-national comparison. *BMC health services research.* 2007;7:109.
28. Scheerhagen M, Van Stel HF, Birnie E, Franx A, Bonsel GJ. *Het discriminerend vermogen van de ReproQuestionnaire [the discriminative power of the ReproQuestionnaire].* Rotterdam, The Netherlands: Erasmus Medical Center;2015.
29. Truijens SE, Pommer AM, van Runnard Heimel PJ, Verhoeven CJ, Oei SG, Pop VJ. Development of the Pregnancy and Childbirth Questionnaire (PCQ): evaluating quality of care as perceived by women who recently gave birth. *European journal of obstetrics, gynecology, and reproductive biology.* 2014;174:35-40.
30. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC Pregnancy Childbirth.* 2009;9:18.
31. Feskens RCW, Hox J, Lensvelt-Mulders G, Schmeets H. Collecting Data among Ethnic Minorities in an International Perspective. *Field Methods.* 2006;18:284–304.
32. Hox J, de Leeuw E. Non-response in written surveys, interviews by phone, and face-to-face interviews; a comparing meta-analyses. [Nonrespons bij schriftelijk vragenlijsten, telefonische interviews en face-to-face interviews; een vergelijkende meta-analyse]. *Kwantitatieve Methoden.* 1993;43:59-78.
33. Hox J, de Leeuw E. Non-response in surveys: an overview. [Nonrespons in surveys: Een overzicht]. *Kwantitatieve Methoden.* 1998;19(57):31–53.
34. Statistics Netherlands. Labor fource: obtained education according to personal characteristics 2001-2012. *Statline* 2013; http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71822NED&D1=0&D2=l&D3=1-3&D4=0&D5=a&D6=0&D7=l&HDR=T,G6,G2,G5,G3&STB=G1,G4&VW=T%5D/SCP/Emancipatiemonitor.
35. Bobak M, Pikhart H, Hertzman C, Rose R, Marmot M. Socioeconomic factors, perceived control and self-reported health in Russia. A cross-sectional survey. *Soc Sci Med.* 1998;47(2):269-279.
36. Agyemang C, Denktas S, Bruijnzeels M, Foets M. Validity of the single-item question on self-rated health status in first generation Turkish and Moroccans versus native Dutch in the Netherlands. *Public Health.* 2006;120(6):543-550.

37. Fransen M, Harris VC, Essink-Bot ML. Low health literacy in ethnic minority patients: understandable language is the beginning of good healthcare. *Nederlands Tijdschrijft voor Geneeskunde.* 2013;157(14): A5581.
38. PRN Foundation. Netherlands Perinatal Registry. *2004-2013* 2013.
39. Bos N, Sturms LM, Stellato RK, Schrijvers AJ, van Stel HF. The Consumer Quality Index in an accident and emergency department: internal consistency, validity and discriminative capacity. *Health Expect.* 2015;18(5):1426-1438.
40. de Boer D, Delnoij D, Rademakers J. The discriminative power of patient experience surveys. *BMC health services research.* 2011;11:332.
41. Krol MW, De Boer D, Sixma H, Van Der Hoek L, Rademakers JJ, Delnoij DM. Patient experiences of inpatient hospital care: a department matter and a hospital matter. *Int J Qual Health Care.* 2015;27(1):17-25.
42. Streiner DL, Norman GR, Cairney J. *Health measurement scales.* Oxford: University Press; 2014.
43. Stubbe JH, Brouwer W, Delnoij DM. Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire. *BMC Ophthalmol.* 2007;7:14.
44. Twisk JWR. *Applied Multilevel Analysis.* Cambridge University Press; 2014.
45. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2005;25(3):250-261.
46. Engel L, Beaton DE, Touma Z. Minimal Clinically Important Difference: A Review of Outcome Measure Score Interpretation. *Rheumatic diseases clinics of North America.* 2018;44(2):177-188.
47. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2):171-184.
48. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2017;20(4):644-650.
49. Collins-Fulea C, Mohr JJ, Tillett J. Improving midwifery practice: the American College of Nurse-Midwives' benchmarking project. *J Midwifery Womens Health.* 2005;50(6):461-471.
50. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
51. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
52. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
53. Alderliesten ME, Stronks K, Bonsel GJ, et al. Design and evaluation of a regional perinatal audit. *European journal of obstetrics, gynecology, and reproductive biology.* 2008;137(2):141-145.
54. Eskes M, Waelput AJ, Erwich JJ, et al. Term perinatal mortality audit in the Netherlands 2010-2012: a population-based cohort study. *BMJ open.* 2014;4(10):e005652.
55. Kurinczuk JJ, Draper ES, Field DJ, et al. Experiences with maternal and perinatal death reviews in the UK--the MBRRACE-UK programme. *BJOG : an international journal of obstetrics and gynaecology.* 2014;121 Suppl 4:41-46.
56. Mancey-Jones M, Brugha RF. Using perinatal audit to promote change: a review. *Health policy and planning.* 1997;12(3):183-192.
57. Goodarzi B, van der Post J, Schellevis F, de Jonge A. Implementing of integral maternity care is premature [Invoering integrale geboortezorg voorbarig]. *Medisch Contact.* 2018;73(24):18-20.

10

Summary / samenvatting

SUMMARY

In 2000 the World Health Organization (WHO) presented a comprehensive model to compare different health systems, monitor its performance, and evaluate system changes[1]. According to this model, a health care system's performance is good if the average levels of both medical outcome and responsiveness (defined as aspects related to the way individuals are treated and the environment in which they are treated)[2] are high, when there are no inequalities in health status and responsiveness, and minor variability across individuals in fairness in financial contribution[1].

This WHO model seems also appropriate to evaluate the system changes in the Dutch maternity care. The performance of the Dutch maternity care is currently unknown. Although the system's clinical outcomes can be evaluated in terms of indicators of perinatal morbidity and perinatal mortality –which we know are suboptimal[3-9], the system's responsiveness and financial fairness have not been evaluated before. This thesis is especially dedicated to assessing the responsiveness of the Dutch maternity care.

As an independent indicator of performance, clients' experiences are often measured as part of a two-stage quality cycle, a so-called benchmark[10,11]. In the first stage, the best and worst practices are identified. In the second stage, underperformers are invited to improve their results, by quality improvement efforts based on an internal interpretation of the performance results[12-15].

To structurally evaluate the clients' experiences within and between units, several questionnaires already exist, e.g. NHS and CQ instruments[16-19]. However, existing questionnaires are unable to deal with the current transition in Dutch maternity care to integrated care. Therefore, we developed and extensively tested a patient reported measure addressing the client's experience with perinatal care conform the WHO responsiveness model: the ReproQuestionnaire (ReproQ).

The ReproQ consists of two complementory versions: an antenatal version, which covers the experiences during early and late pregnancy, and a postnatal version, which covers the experiences during childbirth and postnatal care.

The conceptual basis of the ReproQ was the WHO responsiveness model[2,20,21]. This model consists of four domains on the interactions of the client with the health professional (Dignity, Autonomy, Confidentiality, and Communication), and four domains on the client orientation of the organizational setting (Prompt attention, Access to family and community support, Quality of Basic amenities, and Choice and continuity of care)[2,20,21].

Additional sections of the ReproQ address the client's socio-demographic characteristics, details about the care process during pregnancy and childbirth, a global experience rating, and maternal and baby's health outcomes in non-medical terms as perceived by the mother. We added also a relevance question on which two out of eight domains were most important to the client.

To summarize women's experiences we used the eight individual domains, summary scores of the four personal domains and the four setting domains, and a total score across all domains. Most analyses were performed using three scoring models each with its own merits: mean score, above/below the median, and having a negative experience.

In **chapters 2 and 3** the validity and reliability of the ReproQ were examined. The content validity of the null version of the ReproQ was sufficient: both clients and maternity care professionals judged all responsiveness domains relevant, and the main missing topic (the expertise of the health professional) does not fit the responsiveness concept. After first adaptation, a web-based survey was conducted. The performance of the digital self-report ReproQ appeared favorable for routine measurement. Respondents highly appreciated the system's responsiveness (median of the total score was 3.70 (range 1 – 4)). The assumed domain structure was supported by the exploratory factor analysis, but that analysis suggested several small adaptations (**chapter 2**). Moreover, the ReproQ provides reliable data to be used in a benchmark: the reliability for the negative and mean score was both 'good' (**chapter 3**).

To be suitable for a two-stage quality cycle, the ReproQ should be able to identify differences both between units and between subgroups of clients. These differences can be identified in terms of statistically significant differences and in terms of a 'clinically' relevant difference. To identify relevant differences we estimated the Minimally Important Difference (MID), which was 9% for having a negative experience and 0.10 for the mean experience score during birth. Considering the skewedness of the clients' experiences, these estimated MIDs seem rather large (**chapter 3**).

Pre-stated group comparisons confirmed the expected significant difference following a good versus adverse birth outcome. Fully integrated organizations performed slightly better than less integrated organizations (**chapter 2**). Application of the MIDs revealed relevant differences in women's experience with regard to professional continuity, setting continuity and having travel time (**chapter 3**).

Chapters 4 and 5 explicitly focus on the discriminative power of the ReproQ. **Chapter 4** discusses which determinants should be considered for case mix correction when a benchmark is performed. For a fair ranking of the perinatal units, case mix correction should be applied when a determinant influences the experiences score and is distributed unequally across health care organizations, but is beyond the influence of and usually unrelated to the organization. Determinants applicable for case mix correction are the client's socio-demographic characteristics and the client's reported outcome of mother and child. **Chapter 5** indicates that the ReproQ can distinguish care providers that perform above or below average with both the statistical and relevance approaches. However, the clients' experiences are only to a limited extent related to the perinatal unit in which they received care – the within-unit variability was much larger than the between-unit variability. The D-study to assess sample size for achieving excellent reliability (G-coefficient of 0.8)

indicated that for the total score each perinatal unit needs to include 272 (antenatal), 432 (delivery), and 258 (postnatal period) valid responses. Considering the second stage of the quality improvement cycle, our findings imply that a low experience score can be assigned to variation in specific subgroups of clients, i.e.: women with low antenatal experiences, women who did not experience professional and setting continuity, and women who did not have expectations towards the place of birth (**chapter 4**). Care for these subgroups of clients require improvement.

Chapters 6, 7 and 8 focus on the applicability of the ReproQ in a quality improvement cycle and its implementation. **Chapter 6** explores the validity of retrospective measurement of the antenatal experiences, and consequently when the antenatal ReproQ needs to be administered. Our results show that the association of the antenatal experiences measured before and after birth was moderate. This supports our approach to measure the antenatal experiences during pregnancy instead of after birth.

Chapters 7 and 8 are the reports on two different applications of ReproQ and its outcomes (in the second phase of the quality cycle) for maternity care improvement. **Chapter 7** describes how the results of ReproQ can be used as basis for quality improvement. Clients and health care professionals made recommendations to improve care, prioritized these recommendations and eventually consented on an improvement agenda. **Chapter 8** studies the use of ReproQ as evaluation instrument for health care interventions (here: the implementation of birth centers) and its role in quality improvement. These results show that the domains Autonomy and Prompt attention should be improved for women who plan to give birth in a birth center as well as for the women who were referred to the hospital during childbirth. These studies confirm the ReproQ's suitability for the second stage of a quality improvement cycle.

In **chapter 9**, the main findings are discussed and conclusions of this thesis are drawn and summarised. These studies show that the ReproQ is suitable as instrument in a two-stage quality cycle. This chapter also focuses on several considerations that affect the benchmark, e.g. determinants for which case mix correction is required, scoring model (negative or mean score), the norm (best-practices or overall average of all units), and the criterion (statistically or relevance).

REFERENCES

1. Murray CJL, Frenk L. Summary measures of population health in the context of the WHO framework for health system performance assessment. In: Murray CJL, Evans DB, eds. *Health Systems Performance Assessment*. Geneva, Swiss: World Health Organisation,; 2002.
2. Valentine NB, De Silva A, Kawabata K, Darby C, Murray CJ, Evans BE. Health system responsiveness: concepts, domains and measurement. In: Murray CJ, Evans BE, eds. *Health systems performance assessment*. Geneva: World Health Organization; 2003:573-596.
3. Bonsel GJ, Birnie E, Denktas S, Poeran J, Steegers EAP. *Trends in the perinatal mortality. Description study pregnancy and birth 2010.* Rotterdam: Erasmus MC; 2010.
4. de Graaf JP, Ravelli AC, de Haan MA, Steegers EA, Bonsel GJ. Living in deprived urban districts increases perinatal health inequalities. *J Matern Fetal Neonatal Med.* 2013;26(5):473-481.
5. de Graaf JP, Schutte JM, Poeran JJ, van Roosmalen J, Bonsel GJ, Steegers EA. Regional differences in Dutch maternal mortality. *Bjog.* 2012;119(5):582-588.
6. Mohangoo AD, Buitendijk SE, Szamotulska K, et al. Gestational age patterns of fetal and neonatal mortality in Europe: results from the Euro-Peristat project. *PLoS One.* 2011;6(11):e24727.
7. van der Velden K. *Safe care during pregnancy and birth.* 2009.
8. Zeitlin J, Mohangoo AD, Delnord M, Cuttini M, Committee E-PS. The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *J Epidemiol Community Health.* 2013;67(12):983-985.
9. Zeitlin J, Szamotulska K, Drewniak N, et al. Preterm birth time trends in Europe: a study of 19 countries. *Bjog.* 2013;120(11):1356-1365.
10. Haugum M, Danielsen K, Iversen HH, Bjertnaes O. The use of data from national and other large-scale user experience surveys in local quality work: a systematic review. *Int J Qual Health Care.* 2014;26(6):592-605.
11. Weinick RM, Quigley DD, Mayer LA, Sellers CD. Use of CAHPS patient experience surveys to assess the impact of health care innovations. *Joint Commission journal on quality and patient safety.* 2014;40(9):418-427.
12. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
13. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
14. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
15. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
16. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483
17. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* Oxford, United Kingdom: National Perintal Epidemiology Unit;2010.
18. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg
19. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC Pregnancy Childbirth.* 2009;9:18.
20. de Silva A. A framework for measuring responsiveness. *Global Programme on Evidence for Health Policy Discussion Paper 32* 2000.
21. Gostin L. The domains of health responsiveness: a human rights assessment. 2002.

SAMENVATTING[1]

In 2000 presenteerde de Wereldgezondheidsorganisatie (WHO) een model om verschillende gezondheidszorgsystemen te vergelijken op hun performance, deze te monitoren en veranderingen te evalueren[1]. Dit model bestaat uit drie onafhankelijke indicatoren: medische uitkomsten, responsiviteit en eerlijke verdeling van de zorgkosten. Centraal in dit proefschrift staat de responsiviteit: de manier waarop cliënten worden behandeld en de omgeving waarin dat gebeurt als zij zorg ontvangen[2]. Volgens het WHO model is de performance van een zorgsysteem goed als de medische uitkomsten en responsiviteit gemiddeld goed zijn, wanneer er geen ongelijkheden zijn in gezondheidstoestand en responsiviteit, en wanneer alle cliënten toegang hebben tot de noodzakelijke zorg en hierin niet beperkt worden door de financiële bijdrage[1].

Dit WHO model lijkt ook geschikt om de veranderingen in de Nederlandse geboortezorg te evalueren, waaronder de invoering van verloskundige samenwerkingsverbanden (VSVs). Op dit moment is de performance van de geboortezorg onbekend. Over de medische uitkomsten van de Nederlandse geboortezorg (ziekten en sterfte rondom de bevalling) is al het een en ander bekend[3-9]. De responsiviteit en financiële gelijkheid van de geboortezorg zijn nog niet eerder geëvalueerd. Dit proefschrift richt zich op de responsiviteit van de Nederlandse geboortezorg, welke gemeten wordt aan de hand van ervaringen van de cliënt met de zorg.

Als een onafhankelijke indicator van performance worden de cliëntervaringen vaak gemeten als onderdeel van een kwaliteitscyclus of benchmark die bestaat uit twee fasen[10,11]. In de eerste fase worden de beste en slechtst presenterende organisaties geïdentificeerd. In de tweede fase, onderzoeken VSVs de oorzaak van slechte scores, en gaan ze aan de slag met het verbetermaatregelen[12-15].

Om de cliëntervaringen tussen en binnen VSVs structureel te evalueren zijn er verschillende vragenlijsten beschikbaar, bijvoorbeeld de vragenlijst van de NHS en enkele CQ-indexen[16-19]. De bestaande vragenlijsten zijn echter niet geschikt voor het evalueren van de geïntegreerde zorg, waarin de verschillende ketenpartners nauw samenwerken. Daarom hebben wij een cliëntervaringsvragenlijst ontwikkeld en uitgebreid getest, die wel geschikt is om de geïntegreerde zorg te evalueren: de ReproQuestionnaire (ReproQ).

De ReproQ bestaat uit twee versies: een antenatale versie, die de ervaringen tijdens de vroege en late zwangerschap beslaat, en een postnatale versie die de ervaringen tijdens de bevalling en de kraamzorg evalueert.

Als basis voor de ReproQ is het responsiviteitsmodel van de WHO gekozen[2,20,21]. Dit model bestaat uit vier domeinen die gaan over de interactie tussen de cliënt en de zorgprofessional (Respect, Autonomie, Privacy en Communicatie), en vier domeinen die gaan over de mate waarin de organisatie cliënt-vriendelijk is (Tijd tot geboden hulp, Sociale ondersteuning,

[1] De Nederlandse samenvatting is geen letterlijke vertaling van de Engelse samenvatting.

Faciliteiten en Keuze in en continuïteit van de zorg)[2,20,21]. De kern van de vragenlijst bestaat uit vragen die gaan over bovenstaande domeinen. Aanvullende modules van de ReproQ gaan over de sociaal-demografische karakteristieken van de cliënten, details in het zorgproces van de cliënt tijdens de zwangerschap en geboorte, een globale ervaringsscore, en de gezondheidsuitkomsten van moeder en kind zoals die door de cliënt worden ervaren. Tot slot is een wegingsvraag toegevoegd, waarin cliënten aangeven welke twee van de bovenstaande acht domeinen ze het meest belangrijk vinden.

De cliëntervaringen worden samengevat in acht domeinscores, een samenvattende score van de vier professionele domeinen, een samenvattende score van de vier setting domeinen, en een totaal score van alle acht domeinen. De meeste analyses zijn uitgevoerd voor drie scoremodellen: de gemiddelde score, boven of onder de mediaan (middelste waarde), en het hebben van een negatieve ervaring.

In **hoofdstukken 2 en 3** is de validiteit en betrouwbaarheid van de ReproQ onderzocht. De inhoudsvaliditeit van de concept versie van de ReproQ was voldoende: zowel cliënt als professionals uit de geboortezorg vonden alle domeinen relevant. Het belangrijkste onderwerp dat werd gemist was de expertise van de professional. De expertise van de professional past echter niet past binnen het concept responsiviteit . Immers, hoe kundig een professional is, is niet goed te beoordelen door de cliënt die deze expertise niet heeft. Dit onderwerp is daarom niet toegevoegd. Na de eerste aanpassingen is er een proefonderzoek uitgevoerd. Tijdens dit onderzoek vulden vrouwen de ReproQ zelfstandig in. Uit dit onderzoek bleek dat de ReproQ geschikt is voor routinematige metingen. Daarnaast bleek dat de cliënten over het algemeen goede ervaringen hadden met de geboortezorg: de mediaan van de totaal score was 3,70 op een schaal van 1-4. Onze indeling van de vragen over de domeinen werd ondersteund door de exploratieve factor analyse. Desondanks hebben we enkele kleine aanpassingen gedaan naar aanleiding van de factoranalyse (**hoofdstuk 2**). Uit aanvullend onderzoek bleek daarnaast dat de ReproQ voorziet in betrouwbare data voor de benchmark: de betrouwbaarheid van de gemiddelde en negatieve score was voor beide 'goed' te noemen (**hoofdstuk 3**).

Voor een succesvolle benchmark moet de ReproQ verschillen tussen VSVs kunnen identificeren, en onderscheid kunnen maken tussen cliëntgroepen. Dit onderscheid kan worden gemaakt op basis van 'statistische significantie' en op basis van een 'medisch' relevant verschil. Om relevante verschillen te identificeren is de Minimally Important Difference (MID), het zogenaamde minimale relevante verschil, berekend. Als een verschil tussen twee groepen cliënten of VSVs groter is dan de MID doet een VSV er verstandig aan de zorg voor deze groep te verbeteren. De MID was 9% voor het hebben van een negatieve ervaring en 0,10 voor de gemiddelde score tijdens de bevalling. Voor VSVs kan het lastig zijn om hun score met 9% of 0,10 te verbeteren, omdat het merendeel van de vrouwen heeft goede ervaringen en dit de ruimte voor verbetering beperkt (**hoofdstuk 3**).

Als voorbereiding op fase 2 van de benchmark hebben we de ervaringsscores van verschillende cliëntgroepen vergeleken. Dat bevestigde het verwachte significant verschil tussen vrouwen met goede bevallingsuitkomsten en vrouwen met slechte bevallingsuitkomsten. Daarnaast verschilden de scores significant van cliënten die zorg kregen in volledig geïntegreerde en minder geïntegreerde VSVs (**hoofdstuk 2**). Toepassing van de MID toonde eveneens relevante verschillen in ervaringsscores met betrekking tot professionele continuïteit, setting continuïteit en het hebben van reistijd tijdens de bevalling (**hoofdstuk 3**).

Hoofdstukken 4 en 5 gaan over het discriminerend vermogen van de ReproQ. Voor een eerlijke rangordering van VSVs in fase 1 van de benchmark moet casemix correctie worden overwogen. Hierbij worden de scores van de VSVs gecorrigeerd voor de verschillen door samenstelling van de cliënten van VSVs en niet door het VSV zelf. Een kenmerk van de cliënt komt voor casemix correctie in aanmerking als het kenmerk de ervaringen van de cliënt beïnvloedt, en de verdeling tussen VSVs verschilt, maar VSVs het cliëntenkenmerk niet kunnen beïnvloeden. Op basis van onze resultaten komen de volgende kenmerken in aanmerking voor casemix correctie: de sociaal-demografische kenmerken van de cliënt, en de gezondheidsuitkomsten van moeder en kind zoals ervaren door de cliënt (**hoofdstuk 4)**. **Hoofdstuk 5** toont aan dat de ReproQ onderscheid kan maken tussen VSVs, zowel op basis van statistische significantie als op basis van een relevant verschil. Echter, de ervaringen van de cliënten bleken maar in beperkte mate gerelateerd aan het VSV waarin ze zorg kregen. Voor een betrouwbare schatting van de responsiviteit van een VSV zijn tenminste de antwoorden van 272 (antenataal), 432 (bevalling), en 258 (kraamzorg) cliënten nodig. Voor fase 2 van de benchmark zijn een aantal cliëntgroepen geïdentificeerd waarvoor de zorg verbeterd moet worden, gezien hun lage ervaringsscores. Dit betrof vrouwen met een lage antenatale ervaring, vrouwen die geen professionele en setting continuïteit ervoeren en vrouwen die geen verwachting hadden over waar ze zouden bevallen (**hoofdstuk 4**).

Hoofdstukken 6, 7, en 8 gaan over de toepassing van de ReproQ in kwaliteitscycli, dus fase 2, en de implementatie van de ReproQ in de praktijk. In **Hoofdstuk 6** is onderzocht of de ervaringen tijdens de zwangerschap ook na de bevalling kunnen worden gemeten. Het blijkt echter dat de samenhang tussen de antenatale ervaringen die tijdens en na de zwangerschap worden gemeten maar beperkt met elkaar samenhangen en bovendien systematisch van elkaar verschillen. Dit ondersteunt onze aanpak om de ervaringen tijdens de zwangerschap te meten tijdens de zwangerschap in plaats van na de bevalling. **Hoofdstukken 7 en 8** gaan over twee verschillende toepassingen van de ReproQ en zijn uitkomsten in fase 2 van de kwaliteitscyclus. **Hoofdstuk 7** beschrijft hoe de resultaten van de ReproQ kunnen worden gebruikt voor kwaliteitsverbetering. Cliënten en zorgprofessionals deden aanbevelingen om de zorg te verbeteren, deze te prioriteren en werden het vervolgens eens over de verbeteragenda. **Hoofdstuk 8** gaat in op het gebruik van de ReproQ als evaluatie instrument voor zorginterventies, hier de implementatie van geboortecentra.

Het blijkt dat de Autonomie en Tijd tot geboden hulp verbetering behoeven voor vrouwen die de bevalling in een geboortecentrum gepland hebben, en voor vrouwen die tijdens de bevalling zijn verwezen naar het ziekenhuis. Beide studies bevestigen dat de ReproQ' ook geschikt is voor toepassingen in fase 2 van een benchmark.

In **hoofdstuk 9** worden de belangrijkste bevindingen en conclusies uit dit proefschrift samengevat en bediscussieerd. Geconcludeerd wordt dat de ReproQ geschikt is als instrument voor een benchmark. Dit hoofdstuk gaat ook in op de implementatie van de ReproQ dataverzameling en verschillende modellen om fase 1 en 2 van de benchmark vorm te geven, bijvoorbeeld ten aanzien van hethet gebruikte scoremodel (gemiddelde of het hebben van een negatieve ervaring), de norm waarmee vergeleken wordt (best-practices of het totale gemiddelde van alle VSVs), en het criterium op basis waarvan VSVs worden gecategoriseerd (statistisch of relevantie). Hoofdstuk 9 sluit af met aanbevelingen voor vervolgonderzoek.

REFERENTIES

1. Murray CJL, Frenk L. Summary measures of population health in the context of the WHO framework for health system performance assessment. In: Murray CJL, Evans DB, eds. *Health Systems Performance Assessment*. Geneva, Swiss: World Health Organisation,; 2002.
2. Valentine NB, De Silva A, Kawabata K, Darby C, Murray CJ, Evans BE. Health system responsiveness: concepts, domains and measurement. In: Murray CJ, Evans BE, eds. *Health systems performance assessment*. Geneva: World Health Organization; 2003:573-596.
3. Bonsel GJ, Birnie E, Denktas S, Poeran J, Steegers EAP. *Trends in the perinatal mortality. Description study pregnancy and birth 2010.* Rotterdam: Erasmus MC; 2010.
4. de Graaf JP, Ravelli AC, de Haan MA, Steegers EA, Bonsel GJ. Living in deprived urban districts increases perinatal health inequalities. *J Matern Fetal Neonatal Med.* 2013;26(5):473-481.
5. de Graaf JP, Schutte JM, Poeran JJ, van Roosmalen J, Bonsel GJ, Steegers EA. Regional differences in Dutch maternal mortality. *Bjog.* 2012;119(5):582-588.
6. Mohangoo AD, Buitendijk SE, Szamotulska K, et al. Gestational age patterns of fetal and neonatal mortality in Europe: results from the Euro-Peristat project. *PLoS One.* 2011;6(11):e24727.
7. van der Velden K. *Safe care during pregnancy and birth.* 2009.
8. Zeitlin J, Mohangoo AD, Delnord M, Cuttini M, Committee E-PS. The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *J Epidemiol Community Health.* 2013;67(12):983-985.
9. Zeitlin J, Szamotulska K, Drewniak N, et al. Preterm birth time trends in Europe: a study of 19 countries. *Bjog.* 2013;120(11):1356-1365.
10. Haugum M, Danielsen K, Iversen HH, Bjertnaes O. The use of data from national and other large-scale user experience surveys in local quality work: a systematic review. *Int J Qual Health Care.* 2014;26(6):592-605.
11. Weinick RM, Quigley DD, Mayer LA, Sellers CD. Use of CAHPS patient experience surveys to assess the impact of health care innovations. *Joint Commission journal on quality and patient safety.* 2014;40(9):418-427.
12. Ellis J. All inclusive benchmarking. *J Nurs Manag.* 2006;14(5):377-383.
13. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. *Healthcare Policy.* 2012;7(4):e101-e119.
14. Kay JFL. Health care benchmarking. *Hong Kong Medical Diary.* 2007;12(2):22-27.
15. Department of Health. *Essence of Care 2010.* Crown: The Stationery Office;2010.
16. Hay H. A report on the development of the questionnaire for the 2010 maternity survey. 2010; http://www.nhssurveys.org/surveys/483
17. Redshaw M, Heikkila K. *Delivered with care: a national survey of women's experience of maternity care.* Oxford, United Kingdom: National Perintal Epidemiology Unit;2010.
18. van Wagtendonk I, Hoek vd, Wiegers T. Development of a consumer quality index of post- natal care. 2010; http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/cqi-vragenlijsten - CQIKraamzorg
19. Wiegers TA. The quality of maternity care services as experienced by women in the Netherlands. *BMC Pregnancy Childbirth.* 2009;9:18.
20. de Silva A. A framework for measuring responsiveness. *Global Programme on Evidence for Health Policy Discussion Paper 32* 2000.
21. Gostin L. The domains of health responsiveness: a human rights assessment. 2002.

11

Appendix A. Postnatal ReproQ as Used in the Discriminative Study

Appendix B. Postnatal Experience Items as Currently Used

APPENDIX A. Postnatal ReproQ as Used in the Discriminative Study

Beste mevrouw {achternaam},

Uw zorgverleners willen graag weten wat uw ervaringen zijn met de zorg die u kreeg. Daarom ontvangt u deze vragenlijst. De vragen gaan over uw ervaringen met de zorg tijdens de bevalling en kraamperiode. We zouden het erg fijn vinden als u de vragenlijst invult. Door het invullen van de vragenlijst helpt u met het verbeteren van de geboortezorg!

Het invullen van de vragenlijst

Via deze link kunt u de vragenlijst openen. {link}

Het invullen van de vragenlijst duurt ongeveer 15 minuten.

Wilt u bij het invullen van de vragenlijst alleen denken aan de zorg die u kreeg tijdens de laatste keer dat u bent bevallen.

Uw antwoorden

Het invullen van de vragenlijst is anoniem. Dat betekend dat niemand weet welke antwoorden u heeft gegeven. Ook worden uw antwoorden niet doorgegeven aan anderen, ook niet aan uw zorgverleners. Uw e-mailadres gebruiken we hierna alleen nog om u een herinneringsmail te sturen. Daarna gooien we het weg; we bewaren het dus niet.

Vragen

Heeft u nog vragen? Stuur dan een e-mail naar ReproQ@umcutrecht.nl of bel naar 010 - 70 38 712.

Alvast heel erg bedankt voor het invullen!

Met vriendelijke groet,
Marisja Scheerhagen
(onderzoeker in opleiding)

Uw zwangerschap

1. Bij wie had u de eerste verloskundige controle?
(de intake / het intakegesprek)
- ○ verloskundige [→ ga naar vraag 2]
- ○ gynaecoloog [→ ga naar vraag 3]
- ○ huisarts [→ ga naar vraag 4]
- ○ andere zorgverlener [→ ga naar vraag 5]

2. Bij wie waren de controles tot aan de bevalling?
- ○ vanaf het begin bij de verloskundige [→ ga naar vraag 7]
- ○ tijdelijk naar de gynaecoloog maar daarna weer bij de verloskundige [→ ga naar vraag 5]
- ○ naar de gynaecoloog verwezen en daar gebleven [→ ga naar vraag 5]
- ○ afwisselend bij de verloskundige en de gynaecoloog [→ ga naar vraag 5]

3. Bij wie waren de controles tot aan de bevalling?
- ○ vanaf het begin bij de gynaecoloog [→ ga naar vraag 7]
- ○ tijdelijk naar de verloskundige maar daarna weer bij de gynaecoloog [→ ga naar vraag 5]
- ○ naar de verloskundige verwezen en daar gebleven [→ ga naar vraag 5]
- ○ afwisselend bij de gynaecoloog en de verloskundige [→ ga naar vraag 5]

4. Bij wie waren de controles tot aan de bevalling?
- ○ vanaf het begin bij de huisarts [→ ga naar vraag 7]
- ○ naar de verloskundige en daar gebleven [→ ga naar vraag 5]
- ○ tijdelijk naar de gynaecoloog maar daarna weer bij de huisarts [→ ga naar vraag 5]
- ○ naar de gynaecoloog verwezen en daar gebleven [→ ga naar vraag 5]

5. Bij wie had u de meeste controles in de eerste helft van de zwangerschap?
- ○ verloskundige
- ○ gynaecoloog
- ○ huisarts
- ○ verloskundige en gynaecoloog ongeveer gelijk
- ○ huisarts en gynaecoloog ongeveer gelijk

6. Bij wie had u de meeste controles in de tweede helft van de zwangerschap?

- ○ verloskundige
- ○ gynaecoloog
- ○ huisarts
- ○ verloskundige en gynaecoloog ongeveer gelijk
- ○ huisarts en gynaecoloog ongeveer gelijk

Uw bevalling en het kraambed

7. Waar verwachtte u te bevallen ongeveer één maand voor de bevalling?

- ○ thuis
- ○ geboortecentrum (of vergelijkbare plaats,zoals kraamkliniek, kraamsuite, geboorte-hotel of bevalcentrum)
- ○ ziekenhuis
- ○ ik wist het toen nog niet

8. Bij wie verwachtte u te bevallen ongeveer één maand voor de bevalling?

- ○ verloskundige uit de praktijk
- ○ verloskundige uit het ziekenhuis
- ○ gynaecoloog
- ○ huisarts

9. Waar was u toen de bevalling begon?

- ○ thuis
- ○ geboortecentrum (of vergelijkbare plaats)
- ○ ziekenhuis

10. Wie begeleidde u toen uw bevalling begon?

- ○ verloskundige uit de praktijk
- ○ verloskundige uit het ziekenhuis
- ○ gynaecoloog
- ○ huisarts

11. Hoe is de bevalling begonnen?

- ○ vanzelf (spontaan)
- ○ ingeleid of opgewekt, bijvoorbeeld met infuus of gel
- ○ meteen met keizersnee

12. Hoe laat begon uw bevalling?

- ○ tussen 8.00 ’s ochtends en 17.00 ‘s middags
- ○ tussen 17.00 ‘s middags en 8.00 ’s ochtends

13. Op welke dag begon uw bevalling?

- ○ een gewone doordeweekse dag (ma-di-woe-do-vrij)
- ○ zaterdag
- ○ zondag of een officiële feestdag

14. Moest u vlak voor of tijdens de bevalling met spoed naar het ziekenhuis?

- ○ nee
- ○ ja

15. Waar is uw baby uiteindelijk geboren?

- ○ thuis
- ○ geboortecentrum (of vergelijkbare plaats)
- ○ ziekenhuis

16. Wie begeleidde u toen de baby uiteindelijk werd geboren?

- ○ verloskundige uit de praktijk
- ○ verloskundige uit het ziekenhuis
- ○ gynaecoloog
- ○ huisarts

17. Hoe is uw baby geboren?

- ○ op natuurlijke wijze (zonder ingreep)
- ○ met een knip
- ○ met een tang
- ○ met een vacuümpomp of zuignap
- ○ met een geplande keizersnee
- ○ met een spoed keizersnee

18. Heeft u tijdens de bevalling iets gehad tegen de pijn?

- ○ nee
- ○ ja

19. Hoeveel dagen kraamzorg heeft u gehad?

(Als u op een plek geen enkele kraamzorg hebt gehad, moet u het getal 0 invullen.)

(Als u kraamzorg had, tel dan de dag waarop u bevallen bent, als eerste dag.)

aantal dagen thuis _____dagen

aantal dagen in het geboortecentrum/geboortehotel _____dagen

aantal dagen in het ziekenhuis _____dagen

Uw gezondheid en die van uw baby

20. Hoe gezond was uw baby kort na de bevalling, volgens u of uw partner?

(Bij kleine of grote problemen gaat het om uw ervaring of gevoel)

- ○ gezond, geen problemen
- ○ kleine problemen
- ○ grote problemen
- ○ problemen, maar onduidelijk hoe erg het was
- ○ overleden

21. Moest uw baby de eerste 24 uur naar het ziekenhuis, of in het ziekenhuis blijven?

- ○ nee [→ ga naar vraag 24]
- ○ ja [→ ga naar vraag 22]

22. Waarom moest uw baby toen naar het ziekenhuis, of in het ziekenhuis blijven?

- ○ niet vanwege de gezondheid van mijn baby
- ○ ter observatie van mijn baby
- ○ omdat mijn baby ziek was

23. Waar lag uw baby toen?

- ○ op de kraamafdeling bij mij – de moeder
- ○ op de couveuse afdeling
- ○ op de speciale intensive care of high care afdeling voor pasgeborenen (NICU)
- ○ overgeplaatst naar een ander ziekenhuis dan waar ik lag

24. Nu gaat het over u. Hoe gezond was u na de bevalling?

(Bij kleine of grote problemen gaat het om uw ervaring of gevoel)

- ○ gezond, geen problemen
- ○ kleine problemen
- ○ grote problemen
- ○ problemen, maar onduidelijk hoe erg het was

25. Moest u in de eerste 24 uur na de bevalling naar het ziekenhuis, of in het ziekenhuis blijven?

○ nee [→ ga naar vraag 27]
○ ja [→ ga naar vraag 26]

26. Waarom moest u na de bevalling naar het ziekenhuis toe, of in het ziekenhuis blijven?

○ vanwege nazorg (1 dag)
○ vanwege mijn eigen gezondheid (meer dan 1 dag)
○ omdat mijn baby was opgenomen (meer dan 1 dag)
○ vanwege mijzelf en mijn baby (meer dan 1 dag)

We vragen hierna steeds uw mening over 2 situaties: tijdens de bevalling, en in de dagen na de bevalling. Bij elke vraag moeten dus **2** kruisjes worden gezet.

Als we iets vragen over uw zorgverleners dan bedoelen we de zorgverlener die uw ervaring het meest heeft bepaald.

Omgang tussen u en uw zorgverleners

27. Hielden uw zorgverleners rekening met uw privacy?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

28. Behandelden uw zorgverleners u met respect?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

29. Kreeg u persoonlijke aandacht van uw zorgverleners?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

30. Behandelden uw zorgverleners u vriendelijk?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

31. Hielden uw zorgverleners rekening met uw wensen en gebruiken rondom zwanger zijn en bevallen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

32. Had u het gevoel dat u alles kon vertellen aan uw zorgverleners?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Zelf kunnen beslissen

33. Besliste u mee over uw behandeling als dat kon?

(Het gaat hier niet om noodsituaties)

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

34. Kon u een voorgestelde behandeling weigeren?

(Het gaat hier niet om noodsituaties)

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

35. Besliste u mee over de pijnbehandeling tijdens de bevalling?

- ○ ja, ik besliste helemaal zelf
- ○ ja, ik besliste gedeeltelijk mee
- ○ nee, maar ik wilde wel meebeslissen
- ○ nee, maar ik wilde ook niet meebeslissen
- ○ niet van toepassing, bijvoorbeeld door een keizersnee
- ○ niet van toepassing, pijnbehandeling is niet van tevoren besproken

36. Hoeveel invloed had u op uw geboorteplan?

(In een geboorteplan beschrijft u uw wensen rondom de bevalling, bijvoorbeeld over de plaats van bevallen, de manier van bevallen, en over de pijnbehandeling en kraamzorg)

- ○ ik heb het helemaal zelf bepaald
- ○ ik had veel invloed
- ○ ik had weinig invloed
- ○ ik had geen invloed, zonder dat hier een medische reden voor was
- ○ ik had geen invloed, vanwege een medische reden
- ○ er is niet over gesproken / het is niet aan de orde geweest

Privacy

37. Bespraken uw zorgverleners uw medische situatie met uw familie, alleen als u dat goedvond?

tijdens de bevalling	altijd	meestal	soms	nooit	weet ik niet
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	weet ik niet

38. Kon u belangrijke zaken met uw zorgverleners bespreken, zonder dat anderen dat hoorden?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

39. Gingen uw zorgverleners zorgvuldig om met uw medische gegevens en uw dossier?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Praten met uw zorgverleners

40. Beantwoordden uw zorgverleners uw vragen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

41. Gaven uw zorgverleners dezelfde adviezen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

42. Begreep u de uitleg van uw zorgverleners?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

43. Vertelden uw zorgverleners steeds aan u wat er ging gebeuren?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Tijd totdat u hulp kreeg

44. Werd u snel geholpen als het dringend was?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

45. Werd u snel geholpen als het niet dringend was?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

46. Hadden uw zorgverleners tijd voor u als u daar om vroeg?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

47. Was de kraamzorg op tijd aanwezig?

tijdens de bevalling	altijd	meestal	soms	nooit	geen kraamzorg
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	geen kraamzorg

48. Was de plaats waar u zorg kreeg bereikbaar?

(Denk bij plaats bijvoorbeeld aan het ziekenhuis of geboortecentrum)

(Denk bij bereikbaarheid aan het openbaar vervoer, parkeren en wegwijzers)

tijdens de bevalling	altijd	meestal	soms	nooit	ik was thuis
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	ik was thuis

49. Waren de zorgverleners bereikbaar, bijvoorbeeld via de telefoon, of bel bij het bed?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Gezin en huishouden

50. Werd uw partner of familie bij de zorg betrokken?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

51. Hielden uw zorgverleners rekening met uw gezin en huishouden?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

52. Voelde u zich gesteund door uw partner, familie of anderen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Voorzieningen

De volgende vragen gaan over de voorzieningen tijdens de bevalling en het kraambed.

53. Waren de ruimtes comfortabel?

de ruimte waarin u bent bevallen	altijd	meestal	soms	nooit
de ruimte waarin u kraamzorg kreeg	altijd	meestal	soms	nooit

54. Waren de ruimtes schoon?

de ruimte waarin u bent bevallen	altijd	meestal	soms	nooit
de ruimte waarin u kraamzorg kreeg	altijd	meestal	soms	nooit

55. Waren de ruimtes toegankelijk?

(Het gaat hier om de ruimtes waar u verbleef, zoals de kamer, de douche, het toilet, maar ook om voldoende ruimte rond het bed)

de ruimte waarin u bent bevallen	altijd	meestal	soms	nooit
de ruimte waarin u kraamzorg kreeg	altijd	meestal	soms	nooit

Wisselen van zorgverlener

56. Was bij wisseling van zorgverlener uw nieuwe zorgverlener goed geïnformeerd over uw situatie?

tijdens de bevalling	altijd	meestal	soms	nooit	altijd dezelfde zorgverlener
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	altijd dezelfde zorgverlener

57. Was bij verwijzing naar het ziekenhuis uw nieuwe zorgverlener goed geïnformeerd over uw situatie?

tijdens de bevalling	altijd	meestal	soms	nooit	niet verwezen
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	niet verwezen

58. Kreeg u de soort zorgverlener die u wilde?

(Denk aan: huisarts, verloskundige, ziekenhuis verloskundige, gynaecoloog, kraamverzorgster, kinderarts)

tijdens de bevalling	altijd	meestal	soms	nee, er was geen keuze
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nee, er was geen keuze

59. Was het u steeds duidelijk wie de leiding had over de zorg die u kreeg?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Overige vragen

60. Kende u voor de bevalling de zorgverlener die de leiding had over uw bevalling?

- ○ ja, ik heb haar/hem ontmoet en gesproken
- ○ ja, ik kende haar/hem alleen van naam
- ○ nee, ik kende haar/hem niet
- ○ nee, ik wist niet wie de leiding had

61. Kon u uw zorg krijgen op de plaats die nodig was?

- ○ ja
- ○ nee, er was geen plaats (vol)
- ○ nee, andere reden

62. Hoe zou u over het algemeen uw gezondheid noemen?

slecht	matig	goed	zeer goed	uitstekend

63. We gaan er vanuit dat iedereen zijn best doet om u goede zorg te geven. Soms lukt dat niet. Niet alles is even erg. Hieronder staan 8 situaties die kunnen voorkomen. Lees eerst alle 8 mogelijkheden door.

Welke 2 situaties vindt u het ergst? *(geef beide aan met 2 kruisjes.)*

	onvriendelijk en onbeleefd behandeld
	niet betrokken bij keuze behandeling
	slordige omgang met mijn persoonlijke gegevens
	weinig uitleg gekregen over ziekte en behandeling
	lang moeten wachten op hulp
	weinig aandacht voor mijn gezin en huishouden
	behandeld in vieze en oncomfortabele ruimtes
	weinig keuze om voor een eigen zorgverlener te kiezen

64. Hieronder staan opnieuw dezelfde 8 situaties die kunnen voorkomen.

Welke 2 situaties vindt u het MINST erg? *(geef beide aan met 2 kruisjes)*

	onvriendelijk en onbeleefd behandeld
	niet betrokken bij keuze behandeling
	slordige omgang met mijn persoonlijke gegevens
	weinig uitleg gekregen over ziekte en behandeling
	lang moeten wachten op hulp
	weinig aandacht voor mijn gezin en huishouden
	behandeld in vieze en oncomfortabele ruimtes
	weinig keuze om voor een eigen zorgverlener te kiezen

65. Hierna komt de laatste vraag over uw ervaringen tijdens de bevalling en kraambed. Alles bij elkaar, hoe zou u uw ervaringen beschrijven?

Zet een kruisje of streepje op of tussen beide uiteinden op de plaats die voor uw gevoel het beste past.

	1	2	3	4	5	6	7	8	9	10	
ik had een erg slechte ervaring											ik had een erg goede ervaring

Eerdere ervaringen

De volgende vragen gaan over uw eerdere ervaring met zwangerschap en bevallen.

66. Bent u al eens eerder bevallen?

- ○ nee [→ ga naar vraag 68]
- ○ ja, ik ben 1 keer eerder bevallen [→ ga naar vraag 67]
- ○ ja, ik ben 2 keer eerder bevallen [→ ga naar vraag 67]
- ○ ja, ik ben 3 keer eerder of meer bevallen [→ ga naar vraag 67]

67. Bent u eerder bevallen met een keizersnee?

- ○ nee
- ○ ja

68. Heeft u ooit een miskraam of buitenbaarmoederlijke zwangerschap gehad?

- ○ nee
- ○ ja

69. Heeft u ooit een abortus gehad?

- ○ nee
- ○ ja

70. Heeft u ooit een baby gekregen die rond de geboorte is overleden?

- ○ nee
- ○ ja, al voor de geboorte
- ○ ja, tijdens de geboorte
- ○ ja, in de eerste week na de bevalling

71. Was deze zwangerschap een bewuste keuze?

- ○ bewuste keuze
- ○ geen bewuste keuze, maar wel gewenst
- ○ geen bewuste keuze, eigenlijk ongewenst

Algemene gegevens

Tot slot willen we u vragen om nog wat algemene gegevens over u zelf in te vullen. Deze informatie kan in de toekomst helpen bij het leveren van verloskundige "zorg op maat".

72. Wat is uw geboortejaar? __ __ __ __

73. Wat is uw postcode? __ __ __ __ __ __

74. Wat is uw hoogst afgeronde opleiding?

- ○ (nog) geen [ga naar vraag 76]
- ○ lagere school / basisschool [ga naar vraag 76]
- ○ lager beroepsonderwijs (LWOO, VMBO K, VMBO B, LTS enz.) [→ ga naar vraag 76]
- ○ MAVO, VMBO T, 3 jaar HAVO [ga naar vraag 76]
- ○ HAVO [→ ga naar vraag 76]
- ○ VWO, atheneum, gymnasium [ga naar vraag 76]
- ○ middelbaar beroepsonderwijs (MBO) [→ ga naar vraag 76]
- ○ hoger beroepsonderwijs (HBO, HTS, HBO-V, enz.) [→ ga naar vraag 76]
- ○ universiteit [→ ga naar vraag 76]
- ○ opleiding in het buitenland [→ ga naar vraag 75]

75. Hoeveel jaar heeft u onderwijs gehad? _____ **jaar**

76. Wat is uw relatie met de vader van de baby?

- ○ gehuwd of samenwonend
- ○ niet samenwonend, wel relatie
- ○ geen

77. Wij weten dat de behoefte aan zorg rond de zwangerschap verschilt afhankelijk van waar iemand vandaan komt. Daarom willen we graag weten bij welke bevolkingsgroep u en de vader van de baby horen?

	Uzelf	De vader van de baby
Nederlands	0	0
Antilliaans / Arubaans	0	0
Kaapverdiaans	0	0
Turks/Koerdisch	0	0
Surinaams-Creools	0	0
Surinaams-Hindoestaans	0	0
Surinaams-anders	0	0
Marokkaans-Berbers	0	0
Marokkaans-Arabisch	0	0
Indonesisch	0	0
Oost-Europees	0	0
Aziatisch	0	0
Anders	0	0

78. Bij welke zorginstellingen heeft u zorg gehad?

○ verloskundige praktijk: _____

○ ziekenhuis: _____

○ kraamzorg: _____

79. Wilt u nog iets kwijt over uw ervaringen met de geboortezorg?

Hartelijk dank voor het invullen van de vragenlijst!

APPENDIX B. Postnatale Experience Items as Currently Used

Omgang tussen u en uw zorgverleners

1. Hielden uw zorgverleners rekening met uw privacy?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

2. Behandelden uw zorgverleners u met respect?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

3. Kreeg u persoonlijke aandacht van uw zorgverleners?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

4. Behandelden uw zorgverleners u vriendelijk?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

5. Hielden uw zorgverleners rekening met uw wensen en gebruiken rondom zwanger zijn en bevallen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

6. Had u het gevoel dat u alles kon vertellen aan uw zorgverleners?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

7. Besliste u mee over uw behandeling als dat kon?

(Het gaat hier niet om noodsituaties)

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Zelf kunnen beslissen

8. Kon u een voorgestelde behandeling weigeren?

(Het gaat hier niet om noodsituaties)

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

9. Besliste u mee over de pijnbehandeling tijdens de bevalling?

- ○ ja, ik besliste helemaal zelf
- ○ ja, ik besliste gedeeltelijk mee
- ○ nee, maar ik wilde wel meebeslissen
- ○ nee, maar ik wilde ook niet meebeslissen
- ○ niet van toepassing, bijvoorbeeld door een keizersnee
- ○ niet van toepassing, pijnbehandeling is niet van tevoren besproken

10. Hoeveel invloed had u op uw geboorteplan?

(In een geboorteplan beschrijft u uw wensen rondom de bevalling, bijvoorbeeld over de plaats van bevallen, de manier van bevallen, en over de pijnbehandeling en kraamzorg)

- ○ ik heb het helemaal zelf bepaald
- ○ ik had veel invloed
- ○ ik had weinig invloed
- ○ ik had geen invloed, zonder dat hier een medische reden voor was
- ○ ik had geen invloed, vanwege een medische reden
- ○ er is niet over gesproken

11. In welke mate werd u betrokken bij het aantal uren kraamzorg dat u kreeg?

- ○ ik heb het helemaal zelf bepaald
- ○ ik had veel invloed
- ○ ik had weinig invloed
- ○ ik had geen invloed, zonder dat hier een medische reden voor was
- ○ ik had geen invloed, vanwege een medische reden
- ○ er is niet over gesproken

12. Hoeveel invloed had u op de keuze voor borstvoeding, flesvoeding of een combinatie?

- ○ ik heb het helemaal zelf bepaald
- ○ ik had veel invloed
- ○ ik had weinig invloed
- ○ ik had geen invloed, zonder dat hier een medische reden voor was
- ○ ik had geen invloed, vanwege een medische reden

Privacy

13. Bespraken uw zorgverleners uw bevalling met anderen, bijv. met uw familie of vrienden?

- ○ niet besproken met familie of vrienden
- ○ weet niet of dit is besproken
- ○ wel besproken, maar vooraf niet met mij overlegd
- ○ wel besproken, nadat het met mij overlegd was

14. Bespraken uw zorgverleners uw kraamperiode met anderen, bijv. met uw familie of vrienden?

- ○ niet besproken met familie of vrienden
- ○ weet niet of dit is besproken
- ○ wel besproken, maar vooraf niet met mij overlegd

15. Kon u belangrijke zaken met uw zorgverleners bespreken, zonder dat anderen dat hoorden?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

16. Gingen uw zorgverleners zorgvuldig om met uw medische gegevens en uw dossier?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

17. Beantwoordden uw zorgverleners uw vragen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Praten met uw zorgverleners

18. Gaven uw zorgverleners dezelfde adviezen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

19. Begreep u de uitleg van uw zorgverleners?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

20. Vertelden uw zorgverleners steeds aan u wat er ging gebeuren?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Tijd totdat u hulp kreeg

21. Werd u snel geholpen als het dringend was?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

22. Werd u snel geholpen als het niet dringend was?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

23. Hadden uw zorgverleners tijd voor u als u daar om vroeg?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

24. Was de kraamzorg op tijd en volgens afspraak aanwezig?

tijdens de bevalling	direct	na korte tijd	na lange tijd	nee	geen kraamzorg afgesproken
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	geen kraamzorg afgesproken

25. Was de plaats waar u zorg kreeg bereikbaar?

(Denk bij plaats bijvoorbeeld aan het ziekenhuis of geboortecentrum)

(Denk bij bereikbaarheid aan het openbaar vervoer, parkeren en wegwijzers)

tijdens de bevalling	altijd	meestal	soms	nooit	ik was thuis
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	ik was thuis

26. Waren de zorgverleners bereikbaar, bijvoorbeeld via de telefoon, of bel bij het bed?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

27. Werd uw partner of familie bij de zorg betrokken?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

Gezin en huishouden

28. Hielden uw zorgverleners rekening met uw situatie, zoals uw gezin, werk en dagelijkse bezigheden?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

29. Voelde u zich gesteund door uw partner, familie of anderen?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

30. Waren de ruimtes comfortabel?

de ruimte waarin u bent bevallen	altijd	meestal	soms	nooit
de ruimte waarin u kraamzorg kreeg	altijd	meestal	soms	nooit

Voorzieningen

De volgende vragen gaan over de voorzieningen tijdens de bevalling en het kraambed.

31. Waren de ruimtes schoon?

de ruimte waarin u bent bevallen	altijd	meestal	soms	nooit
de ruimte waarin u kraamzorg kreeg	altijd	meestal	soms	nooit

32. Waren de ruimtes toegankelijk?

(Het gaat hier om de ruimtes waar u verbleef, zoals de kamer, de douche, het toilet, maar ook om voldoende ruimte rond het bed)

de ruimte waarin u bent bevallen	altijd	meestal	soms	nooit
de ruimte waarin u kraamzorg kreeg	altijd	meestal	soms	nooit

Wisselen van zorgverlener

33. Was bij wisseling van zorgverlener, uw nieuwe zorgverlener goed geïnformeerd over uw situatie?

tijdens de bevalling	altijd	meestal	soms	nooit	altijd dezelfde zorgverlener
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	altijd dezelfde zorgverlener

34. Was bij verwijzing naar het ziekenhuis, uw nieuwe zorgverlener goed geïnformeerd over uw situatie?

tijdens de bevalling	altijd	meestal	soms	nooit	niet verwezen
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit	niet verwezen

35. Kreeg u de soort zorgverlener die u wilde?

(Denk aan: huisarts, verloskundige, ziekenhuis verloskundige, gynaecoloog, kraamverzorgende, kinderarts)

tijdens de bevalling	altijd	meestal	soms	nee, er was geen keuze
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nee, er was geen keuze

36. Was het u steeds duidelijk wie de leiding had over de zorg die u kreeg?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

37. Werd uw gezondheid en die van uw baby in de gaten gehouden?

tijdens de bevalling	altijd	meestal	soms	nooit
in de dagen na de bevalling (kraambed)	altijd	meestal	soms	nooit

38. Was er tijdens uw bevalling een zorgverlener aanwezig?

tijdens de bevalling	altijd	meestal	soms	nooit

12

Authors and affiliations

PhD Portfolio

Word of thanks (dankwoord)

About the author

Authors and Affiliations

M.E. van den Akker-van Marle
Department of Medical Decision Making, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands

E. Birnie
Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands,
University Medical Center Utrecht, Department of Obstetrics and Gynecology, Academic Collaborative Maternity Care, Utrecht, The Netherlands.

G.J. Bonsel
Erasmus Medical Centre, Department of Obstetrics and Gynaecology, Division of Obstetrics and Prenatal Medicine, Rotterdam, The Netherlands
University Medical Center Utrecht, Department of Obstetrics and Gynecology, Academic Collaborative Maternity Care, Utrecht, The Netherlands.

I.C. Boesveld
Jan van Es Institute, Netherlands Expert Centre Integrated Primary Care, Wisselweg 33, 1314 CB Almere, The Netherlands

P. van Dommelen
Department of Life Style, TNO, PO Box 2215, 2316 ZL Leiden, The Netherlands

N.T.L. van Duijnhoven
Radboud University Medical Center, Department of Obstetrics and Gynaecology, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands

M.J. Faber
Radboud University Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare), P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands

J.P. de Graaf JP
Erasmus Medical Centre, Department of Obstetrics and Gynaecology, Division of Obstetrics and Prenatal Medicine, Rotterdam, The Netherlands

C.J.M. Groenen
Radboud University Medical Center, Department of Obstetrics and Gynaecology, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands

A. Franx
University Medical Center Utrecht, Division Women and Baby, Department of Obstetrics, Utrecht, The Netherlands

M.A. Hermus
Department of Child Health, TNO, PO Box 2215, 2316 ZL Leiden, The Netherlands;
Department of Obstetrics, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands;
Midwifery Practice Trivia, Werkmansbeemd 2, 4907 EW Oosterhout, The Netherlands.

M.F. Hitzert
Erasmus Medical Centre, Department of Obstetrics and Gynaecology, Division of Obstetrics and Prenatal Medicine, Rotterdam, The Netherlands

J.A.M. Kremer
Radboud University Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare), P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands
K.M van der Pal-de Bruin
Department of Child Health, TNO, PO Box 2215, 2316 ZL Leiden, The Netherlands

H.F. van Stel
University Medical Center Utrecht, Julius Center for Health Sciences and Primary Care, Department of Healthcare Innovation and Evaluation, Utrecht, The Netherlands

D.J.C. Tholhuijsen
Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

F.P.H.A. Vandenbussche
Radboud University Medical Center, Department of Obstetrics and Gynaecology, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands

T.A. Wiegers
NIVEL (Netherlands Institute for Health Services Research), PO Box 1568, 3500 BN Utrecht, The Netherlands

PhD Portfolio

Summary of PhD training and teaching

Name PhD Candicate	Marisja Scheerhagen
Erasmus MC department	Obstetrics & Gynecology
Research school	Netherlands Institute for Health Science (NIHES)
PhD period:	January 2012 – June 2019
Promotors	Prof. dr. G.J. Bonsel
	Prof. dr. A. Franx
Copromotors	Dr. E. Birnie
	Dr. H.F. van Stel

PhD training

General and specific courses

Course	*Year*	*ECTS*
External auditing	2018	1.0
Lead auditor ISO-9001	2017	0.5
Auditing quality management systems	2017	0.5
Writing policy statements and advice reports	2015	1.0
Multilevel analyze	2014	1.0
Health Economics	2014	1.5
Research Integrity	2014	0.3
Biomedical English Writing and Communication	2014	3.0
Systematic Literature Retrieval	2014	1.0
Introduction in Endnote	2014	0.1
BROK ('Basiscursus Regelgeving Klinisch Onderzoek')	2013	1.0
Introduction in Syntax in SPSS	2012	0.1

Conferences	*Year*	*ECTS*
EuroQol Academy, Noordwijk, The Netherlands: **oral presentation** *EQ-5D5L in pregnant women and women who recently gave birth*	2017	3.0
International conference of the European Public Health Association (EUPHA), Glasgow, Schotland: **oral presentation** *Do client experiences with maternity care differ according to ethnicity?*	2014	2.0
Symposium 'Movements in maternity care' (in Dutch: 'Beweging in geboortezorg'), Utrecht, The Netherlands	2014	0.3

Symposium 'Pregnancy and Birth' of ZonMW (in Dutch: 'Bijeenkomst Zwangerschap en Geboorte'), **oral presentation** *ReproQuestionnaire – a national field study*	2013	1.0
National Congress – Organization of maternity care (in Dutch: 'Nationaal Congres Organisatie van Geboortezorg'), Utrecht, The Netherlands	2013	0.3
Symposium 'Movements in maternity care' (in Dutch: 'Beweging in geboortezorg'), Utrecht, The Netherlands	2012	0.3
Symposium of the Dutch Society of Quality of Life Research (ISOQOL-NL), Utrecht, The Netherlands: **poster presentation** *Client experiences in maternity care: development of the ReproQuestionnaire*	2012	1.0
Conference of the 'National network research Youth & Health' (in Dutch: 'Landelijk network onderzoek jeugd&gezondheid'), Soesterberg, The Netherlands	2012	1.0
Symposium 'Agrualy better' (in Dutch: 'Aantoonbaar beter'), Utrecht, The Netherlands	2012	0.3

Seminars, workshops and research meetings	***Year***	***ECTS***
BO college concerning quality policy and legislation, Utrecht, The Netherlands	2017 – 2018	5.0
Symposia of the Academic Collaborative Maternity Care with four oral presentations concerning measuring and implementing quality indicators, Utrecht, The Netherlands	2015 – 2018	5.0
Symposium of the knowledge center for maternity care, Nieuwegein, The Netherlands: **oral presentation** *On to a new National Indication Protocol*	2017	1.0
Attending weekly research meeting of the Department of Obstetrics and Gynecology with two **oral presentations**, Rotterdam, The Netherlands	2013 – 2015	5.0
Regional Obstetric Collaboration meetings ('VSV') all over The Nederlands with several **oral presentations**	2013 – 2014	5.0
Minisymposium of the Regional Consortium Southwest Netherlands, Rotterdam, The Netherlands: **oral presentation** *The ReproQuestionaire and the Regional Consortium*	2013	0.5
Minisymposium of the Regional Consortium Northwest Netherlands, Amsterdam, The Netherlands: **oral presentation** *The ReproQuestionaire and the Network maternity care Northwest Netherlands*	2013	0.5
Minisymposium of the Regional Consortium Middle Netherlands, Utrecht, The Netherlands: **oral presentation** *The ReproQuestionaire*	2013	0.5

Visit from the management of Kraamvogel B.V., Amstelveen, The Netherlands: **oral presentation** *ReproQuestionnaire – CQ-index Maternity care*	2013	0.5
Julius Seminar, Utrecht, The Netherlands: **oral presentation** *Client experiences in maternity care: development of the ReproQuestionnaire*	2012	1.0
Visit from the management of Kraamvogel B.V., Amstelveen, The Netherlands: **oral presentation** *ReproQuestionnaire - CQ-index Maternity care*	2013	0.5
Julius Seminar, Utrecht, The Netherlands: **oral presentation** *Client experiences in maternity care: development of the ReproQuestionnaire*	2012	1.0

Teaching

Lecturing	***Year***	***ECTS***
Lectures on 'Evidence based medicine' for midwifery students, Midwifery Academy Rotterdam, The Netherlands	2014	1.0
Lectures on 'Quality of care' for midwifery students, Midwifery Academy Rotterdam, The Netherlands	2014	0.5
Lectures on 'Quality of care' for Physician assisants, University Rotterdam, The Netherlands	2014	0.5

Supervising	***Year***	***ECTS***
Anne-Loes Jonkers – *Measuring patient experiences combined with background factors in perinatal care using the eight domain WHO Responsiveness model.* Master thesis HCM. EUR: IBMG.	2015	2.0
Jacqueline Coenraads-van der Horst – *Valuations and trade-offs of responsiveness in perinatal care using the ReproQ questionnaire*. Master thesis HEPL. EUR: IBMG.	2015	0.5
Martine van Leeuwen – *Vrouwen die zwanger of bevallen zijn, en de domeinen van zorg die zij belangrijk vinden. De invloed van kenmerken van vrouwen en hun ervaringen met geboortezorg.* Bachelor scriptie. EUR: IBMG.	2015	0.5
Dominique Tholhuijsen – *De betrouwbaarheid van de postnatale versie van de ReproQ vragenlijst.* Bachelor scriptie. EUR: IBMG.	2015	1.0
Esma Tuerkeli – *Empowerment als indicator voor de kwaliteit van kraamzorg: Een onderzoek naar de test-hertest betrouwbaarheid van 'Maternity Empowerment Questionnaire' (MEQ).* Bachelor scriptie. EUR: IBMG.	2015	1.0

Bas Brummelhuis – *Komen de ervaringen met zorg tijdens de zwangerschap gemeten tijdens de zwangerschap overeen met het meten na de bevalling? Bachelor scriptie.* EUR: IBMG.	2015	1.0
Lotte Huitema – *The responsiveness of the Dutch perinatal health care system.* Master theses Health Science. VU University.	2014	2.0

Dankwoord

Prof. dr. G.J. Bonsel, beste Gouke, jij was de kapitein op ons schip: je zette de koers uit en bepaalde point north. Het gevolg was een goede gedegen vragenlijst die geschikt was om de geboortezorg te verbeteren. In de afgelopen jaren deed je me af en toe denken aan dr. House: soms onnavolgbaar maar ook een absoluut genie. Ik bewonder je eindeloze stroom aan ideeën, je passie voor onderzoek en je gedrevenheid om die kennis over te brengen. Door je grote kennis ben ik enorm gegroeid met als kers op de taart dit proefschrift *en* een landelijke vragenlijst waarbij de wetenschap het won van de politiek.

Prof. dr. A. Franx, beste Arie, onze vuurtorenwachter. Zolang alles goed ging was je op afstand, maar je hield ons altijd in het zicht en zodra we een ondiepte naderden schoot je meteen te hulp.

Henk, in scheepsmetaforen was jij duidelijk de machinist. Je gaf me mijn eerste kans en maakte daarmee het team compleet. Ik kon bij jou altijd terecht voor vragen over methodes, technieken en analyses en je was de motor achter de multilevel analyse. Helaas hebben we het niet samen mogen afmaken.

Erwin, jij was mijn reddingsbrigade. Je gooide me de life-line toe op het moment dat ik het niet meer zag zitten en gaf me daarna de structuur, rust en het vertrouwen om mijn proefschrift af te maken. Zonder jou geen proefschrift!

Leuke collega's van de Westzeedijk, voor het eerst vind ik het jammer om ergens weg te gaan. Jullie waren goud waard.

Alle cliënten en deelnemende VSVs wil ik bedanken voor hun inzet en betrokkenheid bij het onderzoek. Ik heb genoten van de grote variatie in organisaties, maar vooral van jullie enorme motivatie om de zorg te verbeteren. Op naar een betere geboortezorg, door en voor jullie!

Vrienden, jullie waren mijn wereld buiten de wetenschap. Altijd paraat voor een paar uur zonder statische analyses en wetenschappelijke artikelen, maar gewoon een lekker glas wijn bij de vuurkorf, een potje poolen of abseilen van de Euromast. Om daarna met hernieuwde energie weer verder te gaan met mijn proefschrift.

Pap en mam, dank jullie wel voor jullie oneindige stroom aan liefde, steun en support in dit uitdagende project. Jullie hebben elk hoogtepunt met me gevierd en stonden 100% achter me als uitdagingen obstakels werden.

Nico, lief, jij was *en* bent mijn veilige haven. Hoewel dit mijn project was heb ik het nooit alleen hoeven doen. Dank voor je onvoorwaardelijk steun en altijd aanwezige geloof in mij! Zonder jou was dit proefschrift er niet gekomen. Sophie, met jou komst mocht ik de geboortezorg op een nieuwe manier ervaren. Ik wist niet dat het zo mooi kon zijn.

About the Author

Marisja Scheerhagen shows with her thesis on "Performance of maternity care from the client's perspective" that she thrives on difficult subjects, and that defines her.

Her organizational skill-set and her emphatic way of approaching people let her succeed in implementing the ReproQ in more than half of the perinatal units in the Netherlands. The ReproQ is a questionnaire that measures the quality of the care delivered during pregnancy, birth and postnatal care and is based on the WHO responsiveness model.

Prior to her PhD research she studied Prevention and Public health at the Free University of Amsterdam. She combined her scientific interest and communication skills in a successful internship: Evaluating how children and their parents appreciated the entertainment and educational aspects of a comic book on tooth brushing. Today she works as a policy officer in a maternity care organization, which has quality management and quality improvement as its focus.

Marisja is a professional who is not afraid to set goals that are hard to accomplish, and to make the effort to get the job done. This was also manifest in her scouting days where she organized several regional and international events.

She chooses her own path, when she chose a school against the advice of her teacher, when she joined scouting as the first girl in her village.

Let's illustrate her perseverance with a story dating from 2001 when she was 12 years old. Marisja stood at the bottom of a challenging mountaintop in the Dolomites. Several climbers advised an easier route and thought that she would accomplish halfway at most, but she chose her tough route and succeeded. She made her own plan to the top, path or no path. Marisja, dear daughter, congratulations with the result of your hard work. We are very proud of you!

Martin & Elly Scheerhagen