Implementing personalized obstetric care

Implementing personalized obstetric care

The research described in this thesis was funded by the Netherlands Organization for Health Research and Development (ZonMw, research programme Pregnancy and Childbirth, project no. 209020007 & 505200098150).

The studies presented in this thesis were conducted within the Care and Public Health Research Institute (CAPHRI) at the Department of Epidemiology, Maastricht University, the Netherlands, and the Department of Obstetrics and Gynaecology, Maastricht University Medical Centre+, the Netherlands.

ISBN: 978-94-6375-693-8
Design cover and chapter pages: © evelienjagtman.com
Layout: Pim van Montfort
Portrait photo: © JULE Fotografie
Printed by: Ridderprint BV, Alblasserdam|www.ridderprint.nl

Implementing personalized obstetric care

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op woensdag 19 februari 2020 om 16.00 uur

door

Pim van Montfort

Promotores

Prof. dr. L.J.M. Smits
Prof. dr. M.E.A. Spaanderman

Copromotor

Dr. H.C.J. Scheepers

Beoordelingscommissie

Prof. dr. I.C.W. Arts (voorzitter)
Prof. dr. R.F.P.M. Kruitwagen
Dr. R. Hermens (RadboudUMC Nijmegen)
Dr. M.J. Nieuwenhuijze (Academie Verloskunde Maastricht)
Prof dr. M.A. Joore

Contents

Chapter 1

General introduction

Preface

World's shortest story is a powerful one; 'For sale: baby shoes, never worn.'
Unfortunately, a story that could be told too often in The Netherlands according to the first Europeristat reports of 2008 and 2013 [1,2]. These reports, comparing all European countries with data originating from 2004 and 2010 respectively, showed that the Dutch perinatal mortality rate was above average. A surprise, considering that the unique Dutch obstetric care model long served as an example of well-organized maternity care [3].
In response Europeristat's reports, the Dutch Health ministry organized a steering committee that published recommendations to improve the obstetric healthcare system. These recommendations set a base for the Pregnancy and Childbirth research program organized by ZonMw, a Dutch governmental organization aimed at innovation and healthcare research [4]. Two of the main pillars of this program were improving the risk selection of pregnant women and integrating obstetric care. This resulted into the start of two projects in Limburg: 1) Installation of the Limburg Obstetric Consortium (LOC), intended to jointly reorganize obstetric healthcare in the region and establish an infrastructure for scientific research, and 2) The Expect Study, aimed at improving risk selection during early pregnancy.

Risk selection and prevention of adverse outcomes

In obstetric healthcare, risk selection is the process of quantifying and judging a woman's risk of an adverse pregnancy outcome. The methods used to identify women at increased risk of adverse outcomes varies greatly among countries. In the Netherlands, autonomous midwives (primary care) or gynecologists (secondary care) monitor pregnant women [3]. The obstetric indication list (Verloskundige IndicatieLijst, VIL) is used to check whether there is a predefined risk factor present (e.g. chronic hypertension, diabetes mellitus), or a complication arises (e.g. pregnancy induced hypertension, gestational diabetes mellitus) that warrants transfer from primary to secondary care [5]. Although this list is a national guideline used to judge pregnant women's risk, it is not an individual risk assessment tool, nor does it describe the contents of primary or secondary healthcare.
The majority of perinatal deaths in the Netherlands are related to either asphyxia (Apgar score <7 after 5 minutes), preterm birth (PTB), small-for-gestational-age infancy (SGA), or congenital anomalies [6]. Hypertensive disorders in pregnancy, such as pre-eclampsia (PE), are strongly associated with SGA and PTB [7]. On the other hand, gestational diabetes mellitus (GDM) increases the risk of large-for-gestational-age infants (LGA) [8], which in turn is associated with birth injuries and asphyxia [9]. As a result, PE, GDM, PTB, SGA, and LGA are all related to perinatal mortality. Therefore, preventing these adverse outcomes would eventually lead to a reduction of perinatal mortality.
Identification of women at increased risk for these adverse events may improve outcomes due to increased awareness of both pregnant women as healthcare professionals regarding the occurrence of these events. However, risk selection is even more useful if appropriate and effective interventions exist. A number of interventions may prevent or reduce the risk of adverse pregnancy outcomes, some examples are: low-dose aspirin treatment in case of PE [10-12], adequate management of GDM [13,14], and progesterone administration in women at risk of spontaneous PTB [15]. However, most of these interventions are not suitable for all pregnant women, due to either possible adverse effects, patient burden, or costs. Algorithms by which it would be possible to predict adverse outcomes such as PE accurately

during early pregnancy, would give healthcare professionals the opportunity to apply these preventive measures based on women's individual risk profile.
Often, such algorithms, or prediction models, are logistic regression models. In case of prediction, the coefficients of the model parameters are used to estimate the absolute probability of a certain outcome instead of just describing the correlation between the parameters and the outcome [16]. Consequently, such models take the weighted risk of multiple factors into account simultaneously and allow for a more fine-tuned estimation of the weight of multiple risk factors and possible inter-relations [17]. Therefore, these models may me be more accurate in identifying women at increased risk then guidelines that recommend to merely check whether one of the listed risk factors is present in a woman (e.g. BMI >35, age >40, history of PE) [18].

Development of prediction models for clinical practice

Scientific research aimed at the use of a prediction model in clinical practice can be divided in three to four categories [19,20]. Each category resembles a crucial step in order to achieve the ultimate goal of widespread adoption of the prediction model in clinical practice.
Model development is the first step. Preferably, candidate predictors are selected with the aid of existing literature and an expert opinion panel. Using an observational study design, ideally a prospective cohort, the initial model can be trained by using the selected candidate predictors to predict the outcome [21]. During model development, several variables are eventually selected from the candidate predictors to create a final model [22]. Predictor selection can be a difficult process, with several pitfalls that may affect the reliability of the final model. There are several methods to select the predictors, but there is no consensus yet regarding the best strategy to achieve a final model [16].
Often, results indicating the predictive performance of a model are overestimated when retrieved from the development dataset [20]. For this reason, a prediction model always needs to be validated after development. During validation the model's reliability is tested. There are roughly two kinds of validation: internal and external validation. Internal validation is the validation of the model within the observational study used to develop the model, procedures such as bootstrapping can be applied to correct the initial model with an shrinkage factor [20].
For external validation, the model is applied to a new dataset that has not been used for its development. This dataset represents another cohort which differs in either time, geographical location, or the participants are selected differently [20]. Since most models have a tendency to show too optimistic results even after the internal validation, external validation is strongly recommended before applying the model in clinical practice [23]. If necessary, the results of the external validation process can be used to update the model to improve its accuracy [24].
When a prediction model successfully passes the external validation, the model accurately predicts the outcome in the external validation dataset, the next step is analyzing the potential impact of the model. In other words, the potential usefulness of adopting the model in clinical practice should be studied. Depending on the specific setting and goal of the model an impact analysis is performed with respect to clinical outcomes, healthcare costs, patient satisfaction, or allocations of healthcare resources [19]. When these three phases are successfully completed and the prediction model appears to be clinically beneficial (the model has the potential to improve current clinical practice) the final step is widespread

implementation of the model.
Impact analysis and implementation of the model are, however, not per se mutually exclusive processes. It may be impossible to address several aspects of the impact analysis without implementing the model at a certain level due to a lack of specific data, for example patient satisfaction related with the use of the model.
Implementing a prediction model in clinical practice, thus changing the current clinical practice, can be a complex process. A multitude of barriers and incentives are often at play that may either hinder or facilitate the implementation process. Grol and Wensing describe a 10-step model to induce change of professional behavior [25]. The evidence regarding the most effective strategies to produce behavioral change, however, remains inconclusive and vary greatly depending on the setting and target groups [25]. Still, the chance of successful implementation increases by using a tailored strategy that identifies and addresses potential barriers during the entire process [26,27].
Impact and implementation studies are an essential step in translating predictive research to clinical practice. First, such studies may facilitate the implementation itself, by providing an easy accessible format of the prediction model. Second, they may improve our insight regarding the effects in daily practice. These effects may differ substantially from the results expected from study results, since usage of the prediction tool as well as adherence rates of both healthcare professionals as patients contribute to the observed effect in daily practice [28]. Impact and implementation studies will improve our understanding of how a prediction model is used, whether recommendations correlated to the risks are applied, and whether the effects suggested from earlier studies is achieved [29].

The Expect Study and the Limburg Obstetric Consortium

The Limburg Obstetric Consortium (LOC) consists of five regions representing the Southeastern part of the Netherlands. Every region consists of a hospital providing secondary obstetric care (gynecologists and clinical midwives) and a corresponding group of independent midwives providing primary obstetric care. The LOC committee consists of two to four representatives per region (midwives and gynecologists), representatives of maternity care, representatives of Maastricht University, and a manager. With the aid of numerous surveys consulting all obstetric healthcare professionals of Limburg they reorganized the obstetric healthcare of the province. The main goal was to achieve a uniform set of recommendations that form the base of risk-based care pathways. These care pathways would standardize the obstetric healthcare of the region and would enable a system of integrated client-centered care.

Validation Study

The Expect Study was designed to improve risk selection during early pregnancy and to provide a starting point for personalized obstetric healthcare. Prediction models may be useful tools to achieve an individual assessment of important risks upon adverse pregnancy outcomes. Several models trying to predict the risks of PE, GDM, PTB, SGA, and LGA during early pregnancy have been published. Unfortunately, most models were not externally validated and consequently were not yet ready for usage in clinical practice [30].
The first part of the Expect Study, Expect Study I, aimed to evaluate the validity of published prediction models. The Expect Study specifically focused on models that are applicable during the first trimester and solely relied on non-invasive predictors: predictors that are

collected routinely in Dutch obstetric health care, or are easily to obtain in an outpatient midwifery setting.
For the validation study, 2,614 women were enrolled in a multicenter prospective cohort study from 2013 to 2015 throughout Limburg [31]. The results of Expect Study I indicated that implementing prediction models predicting PE, or GDM may be clinical beneficial and have the potential to improve obstetric care. The non-invasive models predicting fetal growth (SGA, and LGA) were unable to predict these outcomes accurately enough in order to improve current obstetric healthcare. Moreover, the definitions of LGA and SGA also include constitutionally larger or smaller infants. Clinical relevant fetal growth deviations, on the other hand, are often related to underlying disorders such as gestational diabetes and hypertensive disorders. Models predicting the underlying disorders may therefore be more specific [8]. The results regarding the external validation of models predicting spontaneous preterm birth are covered in chapter two of this thesis.

Risk-based care pathways

During the recruitment period of Expect Study I, the LOC developed healthcare pathways that are tailored to women's individual risk profiles. This resulted in pathways consisting of basic antenatal care for all women and additional recommendations for women at risk for pregnancy related complications. For example, women with an increased PE-risk or GDM-risk are recommended to consider a low-dose aspirin prophylaxis or an oral glucose tolerance test, respectively. A detailed description of the specific content of the healthcare pathways is provided in chapter 4 of this thesis.
Members of the LOC agreed to use the best performing prediction models externally validated in Expect Study I to assess women's risk of PE and GDM. Furthermore, consensus was reached regarding suitable cut-off values as risk-threshold. In case women's risk exceeds the selected threshold, it is advised to discuss additional recommendations using a shared decisional approach.

Implementation and impact study

Despite the increasing amount of published prediction models and external validation studies, outside the realm of research, such models have rarely been implemented in daily obstetric practice [32]. The second part of the Expect Study, Expect Study II, was aimed at analyzing the impact of the risk-based care paths assigned to women by the aid of the validated prediction models. To be able to perform an impact analysis and evaluate the effect of risk-based care, Expect Study II also played an important role in facilitating the implementation of the prediction models.
An online prediction tool, the Expect Calculator, embedding externally validated prediction models and LOC's risk-based healthcare pathways, was developed and made available to all healthcare professionals of the region. To facilitate the shared decisional approach regarding the additional recommendations for women with an increased risk, the results of the risk assessment were visualized at a linear scale and provided with corresponding patient brochures.
To evaluate the impact of the prediction tool we used a before-after study design. During Expect Study II, a second prospective multicenter cohort was recruited. Besides a smaller population size and recruitment being facilitated by the prediction tool, Expect cohort I and

II share the same recruitment regimen. Consequently, Expect cohort I represents the former care-as-usual approach and Expect cohort II represents the risk-based care approach in this before-after analysis.

Aims and outline of this thesis

This thesis consists of two parts: the first part describes the preparations that have been performed to facilitate the impact study and its impact analysis, the second part describes the results of the implementation and impact study. The main purposes of the studies in the first part were to analyze the previous care-as-usual approach.
The second part of the thesis addresses several aspects of the implementation process and focusses on the impact of risk-based care. These studies provide insight to what extent the risk-based care approach was implemented and whether discussed interventions were applied in case of an increased risk. Moreover, the impact of risk-based care upon perinatal health is analyzed and a cost-benefit analysis is performed to evaluate the economic impact of risk-based care compared to former care-as-usual.

Part I – Framework of conditions for implementing personalized obstetric care

Chapter two describes the external validation of published models predicting spontaneous preterm birth. It evaluates the clinical potential of these models and whether implementation of these models may be clinical beneficial. Furthermore, strategies and methods that may improve these models are suggested for future research.
Chapter three analyzes women´s appreciation of the obstetric healthcare services during the care-as-usual period (Expect Study I). This chapter specifically focusses on determinants that may cause women to be less satisfied regarding the obstetric healthcare system, in order to increase our understanding how obstetric healthcare could be improved from a client's perspective.
Chapter four describes the protocol of the impact study and how the impact analysis will be performed. Additionally, the specific content of risk-based care is discussed.

Part II – Implementation and impact of personalized obstetric care

The process of selecting cut-of values that indicate which women have an increased risk of PE, is described in chapter five. Furthermore, healthcare professional's adherence to the recommendation to discuss low-dose-aspirin usage with women with an increased PE-risk is analyzed as well in this chapter.
The usage of low-dose-aspirin by pregnant women with an increased PE-risk is analyzed in chapter six, along with potential reasons for non-use.
Chapter seven focusses at the recommendation of an oral glucose tolerance test (OGTT), indicated for women with an increased GDM-risk. Furthermore, this chapter gives insight regarding the burden of the OGTT as experienced by women and we discuss the pro- and cons of universal versus selective GDM screening
The economic impact of risk-based care is discussed in chapter eight. The cost-effectiveness of risk-based care compared to former obstetric care-as-usual is analyzed as well as its impact on perinatal health.
The final chapter, chapter nine, provides a general discussion of the main findings in this dissertation. Along with the results, limitations as well as implications and recommendations for future research and clinical practice will be discussed.

References

1. Euro-Peristat Project with SCPE and EUROCAT, European Perinatal Health Report. The Health and care of pregnant women and babies in Europe in 2010. 2013: p. 113.
2. Euro-Peristat project with SCPE EUROCAT EURONEOSTAT. European Perinatal Health Report 2004. 2008.
3. De Vries, R., Nieuwenhuijze, M., et al., What does it take to have a strong and independent profession of midwifery? Lessons from the Netherlands. Midwifery, 2013. 29(10): p. 1122-1128.
4. Stuurgroep Zwangerschap en Geboorte, Een goed begin; veilige zorg rond zwangerschap en geboorte. Advies Stuurgroep Zwangerschap en Geboorte. Utrecht Stuurgroep Zwangerschap en Geboorte, 2009.
5. Commissie Verloskunde van het College voor zorgverzekeringen, verloskundig vademecum - Verloskundige indicatielijst (VIL), 2003.
6. Bonsel, G., Birnie, E., et al., Lijnen in de perinatale sterfte, Signalementstudie Zwangerschap en Geboorte 2010. 2010, Rotterdam: Erasmus MC.
7. Koullali, B., Oudijk, M.A., et al., Risk assessment and management to prevent preterm birth. Semin Fetal Neonatal Med, 2016. 21(2): p. 80-8.
8. Meertens, L., Smits, L., et al., External validation and clinical usefulness of first-trimester prediction models for small- and large-for-gestational-age infants: a prospective cohort study. Bjog, 2019. 126(4): p. 472-484.
9. Henriksen, T., The macrosomic fetus: a challenge in current obstetrics. Acta Obstet Gynecol Scand, 2008. 87(2): p. 134-45.
10. Meertens, L.J.E., Scheepers, H.C.J., et al., Should women be advised to use calcium supplements during pregnancy? A decision analysis. Matern Child Nutr, 2017.
11. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. American journal of obstetrics and gynecology, 2017. 216(2): p. 110-120. e6.
12. Rolnik, D.L., Wright, D., et al., Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med, 2017.
13. HAPO Study Cooperative Research Group, Hyperglycemia and adverse pregnancy outcomes. N Engl j Med, 2008. 2008(358): p. 1991-2002.
14. Poolsup, N., Suksomboon, N., et al., Effect of treatment of gestational diabetes mellitus: a systematic review and meta-analysis. PLoS One, 2014. 9(3): p. e92485.
15. Dodd, J.M., Jones, L., et al., Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. Cochrane Database Syst Rev, 2013(7): p. Cd004947.
16. Moons, K.G., Kengne, A.P., et al., Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. Heart, 2012. 98(9): p. 683-690.
17. Steyerberg, E.W., Vickers, A.J., et al., Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology, 2010. 21(1): p. 128-38.
18. Wright, D., Syngelaki, A., et al., Competing risks model in screening for preeclampsia by maternal characteristics and medical history. American journal of obstetrics and gynecology, 2015. 213(1): p. 62. e1-62. e10.
19. Adams, S.T. and Leveson, S.H., Clinical prediction rules. Bmj, 2012. 344: p. d8312.
20. Moons, K.G.M., Kengne, A.P., et al., Risk prediction models: II. External validation, model updating, and impact assessment. Heart, 2012. 98(9): p. 691-698.
21. Royston, P., Moons, K.G.M., et al., Prognosis and prognostic research: Developing a prognostic model. BMJ, 2009. 338: p. b604.
22. Moons, K.G.M., Royston, P., et al., Prognosis and prognostic research: what, why, and how? BMJ, 2009. 338: p. b375.

23. Altman, D.G., Vergouwe, Y., et al., Prognosis and prognostic research: validating a prognostic model. BMJ, 2009. 338: p. b605.
24. Toll, D.B., Janssen, K.J.M., et al., Validation, updating and impact of clinical prediction rules: A review. Journal of Clinical Epidemiology, 2008. 61(11): p. 1085-1094.
25. Grol, R. and Wensing, M., What drives change? Barriers to and incentives for achieving evidence-based practice. Medical Journal of Australia, 2004. 180(S6): p. S57-S60.
26. Baker, R., Camosso-Stefinovic, J., et al., Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. Cochrane Database of Systematic Reviews, 2010(3).
27. Brown, B., Cheraghi-Sohi, S., et al., Understanding clinical prediction models as 'innovations': a mixed methods study in UK family practice. BMC Med Inform Decis Mak, 2016. 16: p. 106.
28. Kappen, T.H., van Klei, W.A., et al., Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagnostic and Prognostic Research, 2018. 2(1): p. 11.
29. Wallace, E., Uijen, M.J., et al., Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. BMJ Open, 2016. 6(3): p. e009957.
30. Steyerberg, E.W. and Vergouwe, Y., Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J, 2014. 35(29): p. 1925-31.
31. Meertens, L.J.E., Scheepers, H.C., et al., External Validation Study of First Trimester Obstetric Prediction Models (Expect Study I): Research Protocol and Population Characteristics. JMIR Res Protoc, 2017. 6(10): p. e203.
32. Kleinrouweler, C.E., Cheong-See, F.M., et al., Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol, 2016. 214(1): p. 79-90.e36.

Part I

Framework of conditions for implementing personalized obstetric care

Chapter 2

Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation

Linda J.E. Meertens*, Pim van Montfort*, Hubertina C.J. Scheepers, Sander M.J. van Kuijk, Robert Aardenburg, Ivo M.A. van Dooren, Josje Langenveld, Iris M. Zwaan, Marc E.A. Spaanderman, and Luc J.M. Smits
* Contributed equally

Acta Obstetricia et Gynecologica Scandinavica, 2018; 97(8):907-20

Abstract

Introduction

Prediction models may contribute to personalized risk-based management of women at high risk of spontaneous preterm delivery. Although prediction models are published frequently, often with promising results, external validation generally is lacking. We performed a systematic review of prediction models for the risk of spontaneous preterm birth based on routine clinical parameters. Additionally, we externally validated and evaluated the clinical potential of the models.

Methods

Prediction models based on routinely collected maternal parameters obtainable during first 16 weeks of gestation were eligible for selection. Risk of bias was assessed according to the CHARMS guideline. We validated the selected models in a Dutch multicentre prospective cohort study comprising 2,614 unselected pregnant women. Information on predictors was obtained by a web-based questionnaire. Predictive performance of the models was quantified by the area under the receiver operating characteristic curve and calibration plots for the outcomes spontaneous preterm birth <37 weeks and <34 weeks of gestation. Clinical value was evaluated by means of decision curve analysis and calculating classification accuracy for different risk thresholds.

Results

Four studies describing five prediction models fulfilled the eligibility criteria. Risk of bias assessment revealed a moderate to high risk of bias in three studies. The AUROC of the models ranged from 0.54 to 0.67 and 0.56 to 0.70 for the outcomes spontaneous preterm birth <37 weeks and <34 weeks of gestation, respectively. A subanalysis showed that the models discriminated poorly (AUROC 0.51 to 0.56) for nulliparous women. Although we recalibrated the models, two models retained evidence of overfitting. The decision curve analysis showed low clinical benefit for the best performing models.

Discussion

This review revealed several reporting and methodological shortcomings of published prediction models for spontaneous preterm birth. Our external validation study indicated that none of the models had the ability to adequately predict spontaneous preterm birth in our population. Further improvement of prediction models, using recent knowledge about both model development and potential risk factors, is necessary in order to provide an added value in personalized risk assessment of spontaneous preterm birth.

Introduction

Preterm birth (PTB), usually defined as birth before 37 weeks of gestation, occurs in 5-10% of singleton pregnancies in Europe [1]. The majority of preterm deliveries, approximately 70%, start spontaneously (sPTB) [2]. As both perinatal mortality and morbidity are inversely related to gestational age, health benefits may be achieved by increased monitoring and preventive interventions resulting in a prolongation of pregnancy [3,4].

Progesterone treatment has been reported to reduce the risk of sPTB before 34 weeks of gestation in women at high risk [5,6]. Cervical cerclage or application of a pessary may also protect against sPTB [7-9]. Evidence whether which of the three interventions is most effective is limited [7-9].

Women with a history of sPTB, cervical surgery or a mid-pregnancy short cervix are considered to be at high risk [10]. Without routine cervical length screening, the majority of nulliparous women are regarded as low risk and thus do not receive any preventive treatment. However, universal cervical length screening in women without a history of sPTB results in relatively high numbers needed to screen (1147 in low-risk nulliparous women) [11,12]. Universal cervical length screening is not performed in Dutch obstetric care. Besides a history of sPTB, other risk factors have been associated with PTB, including socioeconomic status, psychological characteristics, family history, height, weight and smoking [13]. Early risk assessment may be useful in order to identify women at risk who may benefit from effective follow-up management strategies.

In the past, several risk assessment tools for sPTB based on a list of single risk factors were developed showing low accuracy rates [14]. In the last decade, a number of promising prediction models based on multivariable regression analysis for the risk of sPTB have been published [15]. Prediction models may be more accurate in identifying women at high risk as regression allows for a more fine-tuned estimation of the weight of multiple risk factors and possible inter-relations [16]. A review of all existing models assessing their methodological quality is lacking. Moreover, most models have not been externally validated, an essential step before implementation in clinical practice [17]. In this article, we performed a systematic review of all existing models predicting sPTB based on routine clinical parameters obtained in first 16 weeks of pregnancy. We externally validated and compared the selected models in a Dutch multicenter prospective cohort of pregnant women.

Methods

Search strategy

This systematic review is reported in accordance with the recently published guidelines for systematic reviews and meta-analyses of prediction model performance [18]. We systematically searched PubMed and EMBASE up to June 26, 2017. Keywords for prediction studies were combined with synonyms for the outcome sPTB appearing in the title, abstract, or MeSH terms. Reference lists of included studies and related articles (i.e. reviews) were manually checked to identify additional eligible articles. The detailed search strategy is provided in Supplementary File S2.1.

Selection criteria

We aimed to identify all published prediction models for the risk of sPTB that are applicable in the first 16 weeks of pregnancy and are based on non-invasive predictors (Supplementary Table S2.1). Studies were eligible if they met the following criteria: (1) the article presented a newly developed prediction model, or a validation or update of a previously developed model in pregnant women, (2) the outcome of the model was the risk of sPTB, (3) the model contained more than one predictor, (4) predictors were available in Dutch obstetric practice (maternal characteristics, anthropometric measures, or blood pressure measurements), 5) predictor values were obtainable during first 16 weeks of pregnancy, and (6) these predictor values were based on regression coefficients. Authors of the original articles were contacted if the model algorithm or definitions of predictors were not available. Studies were excluded in a language other than English, German, French, or Dutch, or if it was a non-original study (for example review). Two researchers (LM, PvM) screened the retrieved titles and abstracts and assessed the eligibility of the full-text papers independently. Discrepancies were resolved by discussion. A third reviewer (LS) was available in case no consensus was reached.

Data extraction and critical appraisal

The risk of bias of the included studies was assessed using the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [19]. The following data were extracted for each included study: source of data, participants, outcome(s) to be predicted, candidate predictors, sample size, handling of missing data, model development, model performance, model evaluation, model presentation, and model interpretation. The risk of bias was critically assessed for eight risk domains: source of data, participant selection, predictor assessment, outcome assessment, sample size, attrition, analysis, and presentation of the model. Risk of bias was rated as low if bias was unlikely, moderate if there were no fatal shortcomings and high if essential errors were made. Previously published risk of bias criteria were used and slightly adapted [20]. Data extraction and critical appraisal was performed independently by two reviewers (LM, PvM). Discrepancies were resolved by discussion and a third reviewer (LS) was available in case of no consensus.

Validation cohort

The included prediction models were externally validated in the Expect Study I [21]. The main purpose of the Expect Study I was to validate published prediction models for several obstetric complications in an independent population. A multicentre prospective cohort study was performed in 36 midwifery practices (primary care) and six hospitals (secondary and tertiary care) in the south-eastern part of the Netherlands between July 1, 2013 and January 1, 2015. Follow-up took place until December 31, 2015. All pregnant women up to 16 weeks of gestation and aged 18 years or older were eligible. Eligible pregnant women were asked to complete two web-based questionnaires (a paper version was available upon request), one before 16 weeks of gestation and one six weeks after the estimated due date. The online questionnaires were accessible via the study website using a unique login code provided with the study information. Automatic reminders were sent in case of incompleteness or nonresponse. Medical records and discharge letters were requested

from caregivers. Pregnancies ending in a miscarriage or termination before 24 weeks of gestation, and women lost-to-follow-up, were excluded. For this study, we also excluded multiple pregnancies and cases of iatrogenic preterm onset of parturition.
The Medical Ethical Committee of the Maastricht University Medical Centre evaluated the study protocol and declared that no ethical approval was necessary (MEC 13-4-053). All participating women gave informed consent through the Internet. The study was registered at The Netherlands Trial Registry on 21 August 2013 (NTR4143, www.trialregister.nl).

Predictor and outcome assessment

Predictors in the included prediction models were assessed by the pregnancy questionnaire completed before 16 weeks of gestation. We used the same definitions as defined in the original articles (Supplementary Table S2.2).
The primary outcome sPTB was defined as a delivery before 37 weeks of gestation with spontaneous onset of parturition (primary contractions or preterm premature rupture of membranes). Secondly, we defined early sPTB as a spontaneously delivery before 34 weeks of gestation. The outcome was obtained from a combination of the medical record and postpartum questionnaire. Cause of labour onset (i.e. spontaneous or not) was available in both data sources. Duration of pregnancy was also available in both data sources and was moreover calculated based on estimated due date and date of birth. Discrepancies between the two variables and data sources were checked. In the absence of the postpartum questionnaire (n=421 sPTB <37 weeks and n=424 sPTB <34 weeks), the medical record was used as reference standard and vice versa (n=16 for both sPTB <37 weeks and sPTB <34 weeks).

Data analysis

A sample size of 2500 women was expected to provide a minimum of 100 cases and 100 non-cases, assuming a 4.5% incidence rate of spontaneous preterm birth <37 weeks of gestation [22].
We imputed missing data for predictors using stochastic regression imputation with predictive mean matching as the imputation model [23]. Characteristics of the validation cohort were described as an absolute value with percentage for categorical variables and as mean ± standard deviation (SD) for continuous variables. We evaluated the relatedness of development samples and validation cohort by comparing the distribution of population characteristics.
The original formulas were used to calculate individual predicted probabilities for each model (Supplementary Table S2.3). We assessed the predictive performance of each model by means of discrimination and calibration for the outcomes sPTB <37 and <34 weeks of gestation, as described in the framework reported by Steyerberg et al. [16]. Discrimination indicates the ability of the model to distinguish between women who will have a sPTB and those who will not. For each model, we computed the area under the receiver operating characteristic curve (AUROC) with 95%-confidence interval (CI). A subgroup analysis was performed among nulliparous women as a history of sPTB is a strong risk factor for recurrent sPTB. Calibration refers to the agreement between the actual outcome and predicted probabilities by the model. We constructed calibration plots in which women were divided into 10 groups with similar predicted risks, and calculated calibration-in-

the-large and the slope. Calibration-in-the-large (intercept), which compares the mean predicted probabilities with mean observed risk, indicates the extent to which predictions are systematically too low or too high. The slope refers to the average strength of predictor effects. Perfect predictions have an intercept of zero and a slope of one [17]. The prediction models were recalibrated by adjusting the intercept and slope using the linear predictor as the only covariate. Discriminative performance (AUROC) of the models is not affected as this recalibration method does not change the ranking of the predicted probabilities [24]. A discriminative performance below 0.70 is generally considered moderate [16].

Lastly, we performed decision curve analysis to evaluate the potential clinical utility of the models. Decision curve analysis assesses the net benefit (proportion of true positives and false positives) of the prediction models over a range of risk thresholds compared with considering all and no women to be at high risk for sPTB [25]. Sensitivity, specificity, and positive and negative predictive values at certain risk thresholds were calculated for the model with the highest overall net benefit.

Statistical analyses were performed with R version 3.4.1, packages rms, pROC, and DecisionCurve.

Results

General characteristics of the studies

The search identified 2018 unique articles. After title and abstract screening, full t ext assessment was performed for 47 articles. Four articles fulfilled the eligibility criteria[26-29]. Reference cross-checking provided no additional articles. An overview of the systematic study selection is shown in Supplementary Figure S2.1.

The four included studies were all development studies describing five models predicting the risk for sPTB based on maternal characteristics. The studies were conducted in four different countries and published between 2011 and 2014. Two studies used a prospective cohort design and the other two were based on registry data. The number of predictors in the published prediction models varied between 2 and 16. Common predictors were body mass index (BMI), smoking, and previous preterm delivery. The prevalence of sPTB, defined as sPTB <34 weeks of gestation by two studies and <37 weeks of gestation by the other two studies, ranged from 0.9% to 1.1% for sPTB <34 weeks of gestation and from 3.7% to 5.7% for sPTB <37 weeks of gestation. Discriminative performance (AUROC) varied from 0.62 to 0.70. Only one study performed internal validation by bootstrapping and the study of Sananes et al. performed an external validation of which the results were not reported. The key characteristics of the included studies are shown in Table 2.1.

Risk of bias

A summary of potential bias per domain is shown in Figure 2.1. Two studies used registry data for model development, which may be less effective for research purposes due to the likelihood of missing data on promising predictors. Moreover, the outcome was extracted at the same time as the predictors which may lead to bias. Nevertheless, sPTB is an objective outcome so assessment may be less biased. The domain participants was rated as liable to a moderate to high risk of bias due to selective reporting of patient characteristics. Para-Cordero et al. used criteria which are not available at the intended moment of prediction. Besides, women may be treated for spontaneous onset of PTB. Only Alleman et al. explicitly

Table 2.1 Characteristics included prediction models for spontaneous preterm birth

Study, Author (year)	Study design	Population	Time of assessment	No. cases/ total (%)	Definition sPTB	Predictors	Prediction model
Parra-Cordero et al. (2014)	Prospective cohort (n=3480)	Singleton pregnancies Exclusion: iatrogenic delivery <34 weeks of gestation, early-onset pre-eclampsia, early-onset SGA, spontaneous miscarriage, intrauterine fetal death, fetal abnormalities, placental abruption, cerclage, and history of cervical surgery	11^{+0}-13^{+6} weeks of gestation.	31/3310 (0.9)	sPTB <34 weeks of gestation	Prior preterm delivery, smoking	Odds ratios reported
Sananes et al. (2013)	Registry Data 2000-2011 (n=33,761)	Singleton pregnancies Exclusion: fetal deaths, medical terminations, iatrogenic delivery <37 weeks of gestation, and delivery <24 weeks of gestation	<14 weeks of gestation	NR/17,341 (NR)	sPTB <37 weeks of gestation	Age, BMI, prior late miscarriage, prior preterm delivery, prior term delivery, smoking	Odds ratios reported. Full algorithm received by email
Alleman et al. (2013)	Registry Data 2009-2010 (n=12,057)	Singleton pregnancies Exclusion: congenital anomaly, birth weight >3 standard deviations from mean, serious infection, cerclage, tocolysis, and delivery <20 weeks of gestation	First trimester (precise period NR)	153/2699 (5.7)	sPTB <37 weeks of gestation	BMI, diabetes mellitus, education, prior preterm delivery, prior live birth	Algorithm and odds ratios not reported. Full algorithm received by email
Beta et al. (2011)	Prospective Cohort 2006-2009 (n=36,743)	Singleton pregnancies Exclusion: major fetal abnormalities, termination, miscarriage or fetal death before 24 weeks of gestation, and iatrogenic delivery <34 weeks of gestation	11^{+0} to 13^{+6} weeks of gestation	353/33,370 (1.0)	sPTB <34 weeks of gestation	Age, ethnicity, height, method of conception, nulliparous fetal loss, nulliparous late miscarriage, prior preterm birth, prior iatrogenic preterm delivery, prior term delivery, smoking Model 2 with obstetric history subdivided according to number of previous preterm deliveries	Odds ratios reported

BMI, body mass index; NR, not reported; SGA, small-for-gestational-age; sPTB, spontaneous preterm birth. Sananes et al. externally validated their model, but results of this validation are not reported. Alleman et al. performed an internal validation step by 1000-fold bootstrapping, but did not report the results

reports exclusion of women undergoing cerclage or tocolysis from their study population. Parra-Cordero et al. merely excluded women with a history of cerclage. Sample size was scored at moderate risk for the model of Parra-Cordero et al. because the overall number of cases was low (n=31) which probably led to the inclusion of only two predictors. The domains attrition and analysis had the highest risk of bias for all included models. All studies either had incomplete data (loss-to-follow-up or missing predictor values), or did not report any information about missing data (Parra-Cordero et al.). The other three studies were scored as moderate risk because they had a substantial amount of missing data and performed a complete case analysis. Methods of analysis were not reported in enough detail by Parra-Cordero et al.. All studies selected predictors based on statistical significance and only one study performed shrinkage of the regression coefficients. For the models of two studies, only odds ratios were available. As the intercept was unavailable, no initial calibration plots could be drawn. Alleman et al. reported their final model including serum markers. The algorithm consisting only maternal characteristics was provided after contacting the authors. Overall, the study of Beta et al. showed the lowest risk of bias. A detailed description of the data extraction and risk of bias assessment according to the CHARMS checklist is provided in Supplementary Table S2.4 and S2.5.

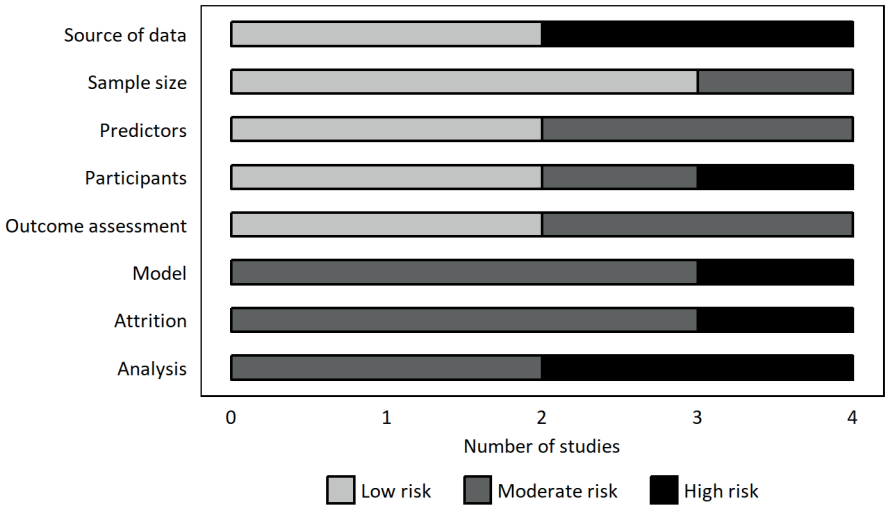


Figure 2.1 Risk of bias assessment of the four included studies according to CHARMS checklist [19]

Validation cohort

The validation cohort consisted of 2,540 women of which 118 (4.6%) had a sPTB <37 weeks of gestation (Figure 2.2). Patient characteristics are shown in Table 2.2. There were ≤1.2% missing values per predictor and the cohort was generally similar after imputation of incomplete predictor variables. Supplementary Table S2.6 provides an overview of complete cases and the imputed validation cohort. The study population for the outcome sPTB <34 weeks of gestation comprised 2,576 women, since fewer women were excluded because of an iatrogenic preterm onset of labour, of which 34 women (1.3%) delivered spontaneously before 34 weeks of gestation.

The distribution of predictors and predictor effects in the original cohorts and our validation

cohort are available in Supplementary Table S2.7. In contrast to the original cohorts, women in our validation cohort were nearly all of Caucasian origin. Almost all population characteristics of Sananes et al. differed considerably compared with the validation cohort. Women in the cohort of Alleman et al. had a higher BMI and higher prevalence of pre-existing diabetes mellitus. The populations of Parra-Cordero et al. and Beta et al. were more comparable, but Parra-Cordero et al. had a higher prevalence of smoking during pregnancy and women in the cohort of Beta et al. were shorter and had a higher prevalence of previous fetal loss. The prevalence of sPTB <37 weeks of gestation was higher in Alleman et al. (5.7%) and lower in the overall population of Sananes et al. (3.7%) compared with the validation cohort (4.6%). The outcome sPTB <34 weeks of gestation was comparable with our prevalence.

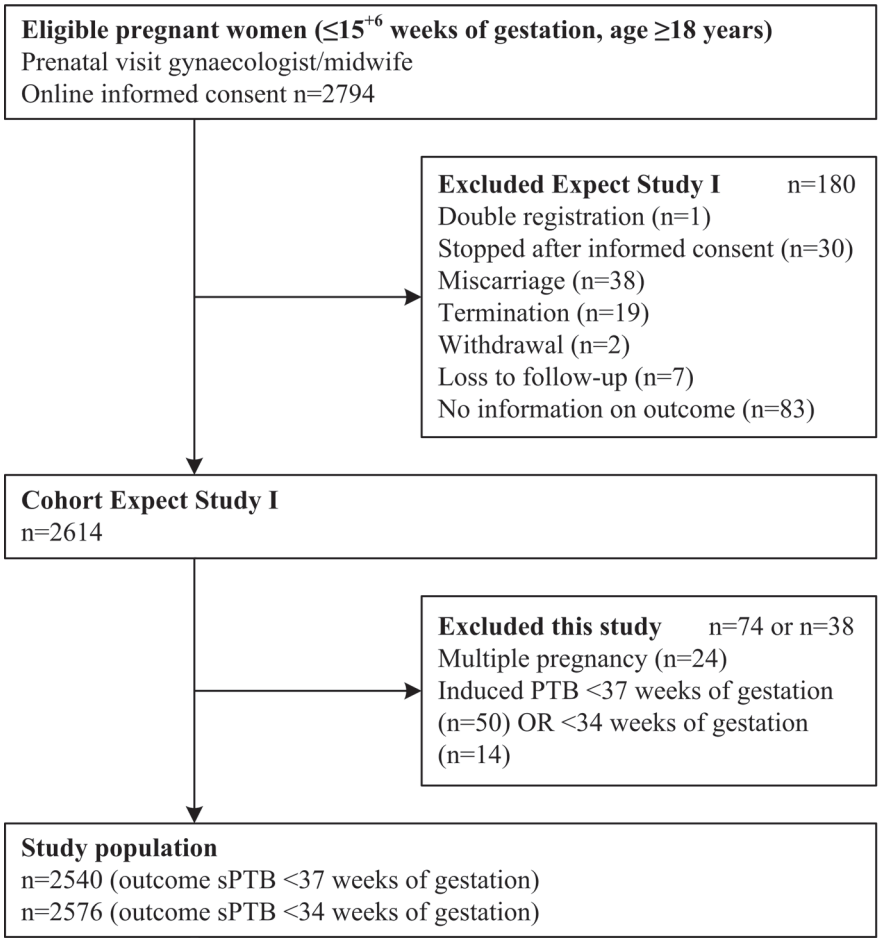


Figure 2.2 Flowchart validation cohort spontaneous preterm birth (sPTB)

Table 2.2 Baseline characteristics of the validation cohort (Expect Study I)

Characteristics	Missing values, n (%)	Observed validation cohort (Expect Study I)[a]		
		Overall (n=2540)	sPTB <37 weeks (n=118)	No sPTB ≥37 weeks (n=2422)
Age, years	0 (0.0)	30.2 (3.9)	30.1 (3.8)	30.2 (3.9)
Ethnicity	0 (0.0)			
Caucasian		2462 (96.9)	115 (97.5)	2347 (96.9)
Afro-Caribbean		3 (0.1)	1 (0.8)	2 (0.1)
South Asian		4 (0.2)	0 (0.0)	4 (0.2)
East Asian		4 (0.2)	1 (0.8)	3 (0.1)
Other Asian		11 (0.4)	1 (0.8)	10 (0.4)
Hispanic		11 (0.4)	0 (0.0)	11 (0.5)
Mixed		45 (1.8)	0 (0.0)	45 (1.9)
Tertiary level of education	3 (0.1)	1380 (54.3)	69 (58.5)	1311 (54.1)
Height, cm	3 (0.1)	168.8 (6.4)	167.3 (6.6)	168.9 (6.4)
Weight, kg	5 (0.2)	68.9 (13.0)	65.6 (11.5)	69.0 (13.0)
Body mass index, kg/m^2	5 (0.2)	24.1 (4.3)	23.4 (3.8)	24.2 (4.3)
Smoking during pregnancy	1 (0.0)	149 (5.9)	8 (6.8)	141 (5.8)
Diabetes mellitus	0 (0.0)	10 (0.4)	1 (0.8)	9 (0.4)
Type 1		8 (0.3)	1 (0.8)	7 (0.3)
Type 2		1 (0.0)	0 (0.0)	1 (0.0)
Other		1 (0.0)	0 (0.0)	1 (0.0)
History of chronic hypertension	0 (0.0)	24 (0.9)	0 (0.0)	24 (1.0)
Parity	0 (0.0)			
Nulliparous		1284 (50.6)	77 (65.3)	1207 (49.8)
Primiparous		1003 (39.5)	35 (29.7)	968 (40.0)
Multiparous		253 (9.9)	6 (5.0)	247 (10.2)
Conception	0 (0.0)			
Spontaneous		2375 (93.5)	114 (96.6)	2261 (93.4)
Ovulation induction		88 (3.5)	3 (2.5)	85 (3.5)
IVF/ICSI		77 (3.0)	1 (0.8)	76 (3.1)
History of fetal loss <16 weeks of gestation	0 (0.0)	702 (27.6)	24 (20.3)	678 (28.0)
History of recurrent miscarriages (≥3)	0 (0.0)	49 (1.9)	1 (0.8)	48 (2.0)
Vaginal bleeding (≥2 days)	0 (0.0)	277 (10.9)	27 (20.3)	250 (10.3)
History of sPTB	30 (1.2)	76 (3.0)	16 (13.6)	60 (2.5)
16-23 weeks of gestation		4 (0.2)	1 (0.8)	3 (0.1)
24-27 weeks of gestation		7 (0.3)	1 (0.8)	6 (0.2)
28-30 weeks of gestation		2 (0.1)	2 (1.7)	0 (0.0)
31-33 weeks of gestation		13 (0.5)	3 (2.5)	10 (0.4)
34-36 weeks of gestation		52 (2.0)	9 (7.6)	43 (1.8)
History of iatrogenic preterm delivery ≥24 weeks of gestation	29 (1.1)	44 (1.7)	0 (0.0)	44 (1.8)
History of term delivery	29 (1.1)	1130 (44.5)	29 (24.6)	1101 (45.5)
History of live birth	18 (0.7)	1221 (48.1)	40 (33.9)	1181 (48.8)

[a]Original data (not imputed) presented as mean (SD) or absolute number (%) ICSI, intracytoplasmic sperm injection; IVF, in vitro fertilisation; sPTB, spontaneous preterm birth

Performance of the models

The discriminative performance of the included models is shown in Table 2.3. For the primary outcome sPTB <37 weeks of gestation, the AUROC ranged from 0.54 to 0.67. The AUROC of the model of Alleman et al. decreased considerably from 0.70 to 0.57 (95% CI 0.52-0.62). The model of Sananes et al. had a slightly higher discrimination compared with the original cohort. All models performed better for the outcome sPTB <34 weeks of gestation. Model 2 of Beta et al. yielded the highest discriminative performance (AUROC 0.70, 95% CI 0.61-0.78). Wide confidence intervals were observed due to the low number of cases for sPTB <34 weeks of gestation. The subgroup analysis among nulliparous women showed a drastic decrease towards almost no discriminative performance for all models. The ROCs in the overall cohort are presented in Supplementary Figure S2.2.

Calibration plots of the two models that provided a complete algorithm are provided in Figure 2.3. The model of Alleman et al. underestimated the risk of sPTB and was overfitted (slope <1). Besides the difference in baseline risk, Sananes et al. was fitted well to our population (slope = 1). Recalibration showed closer fitting to the ideal calibration line (Supplementary Figure S2.3). The models of Alleman et al. and Beta et al. retained some overfitting.

The decision curve analysis of the two best performing models is presented in Figure 2.4. The models had a positive net benefit compared with classifying all or no women as high-risk over a small range of probability thresholds (2.5-10%). However, net benefit remained low throughout this range. This low clinical usefulness is also shown in Table 2.4. Choosing a high sensitivity leads to a large proportion of women that will be indicated unnecessarily as having a high risk of sPTB <37 weeks of gestation. Conversely, a higher specificity leads to a minimal amount of true positives. The model performed especially insufficient among nulliparous women. The moderate performance is predominantly determined by a history of sPTB or term delivery.

Table 2.3 Discrimination of selected prediction models for spontaneous preterm birth

Study, first author (year)	AUROC (95% CI) Original publication	AUROC (95% CI) Validation cohort sPTB <37 weeks (n=2540)	AUROC (95% CI) Validation cohort sPTB <34 weeks (n=2576)	AUROC (95% CI) Validation cohort, nulliparous sPTB <37 weeks (n=1284)	AUROC (95% CI) Validation cohort, nulliparous sPTB <34 weeks (n=1305)
Parra-Cordero et al. (2014)	NR	0.54 (0.50,0.57)	0.56 (0.49,0.63)	0.52 (0.50,0.54)	0.51 (0.46,0.55)
Sananes et al. (2013)	0.618 (0.595,0.641)	0.64 (0.60,0.68)	0.68 (0.59,0.76)	0.53 (0.48,0.57)	0.53 (0.43,0.63)
Alleman et al. (2013)	0.703 (NR)	0.57 (0.52,0.62)	0.61 (0.51,0.71)	0.55 (0.49,0.60)	0.51 (0.39,0.63)
Beta et al. (2011)	Model 1: 0.668 (0.639,0.698) Model 2: NR	0.65 (0.60,0.70) 0.67 (0.62,0.72)	0.68 (0.59,0.77) 0.70 (0.61,0.78)	0.51 (0.45,0.57) 0.54 (0.48,0.60)	0.52 (0.39,0.65) 0.56 (0.44,0.68)

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; NR, not reported; sPTB, spontaneous preterm birth

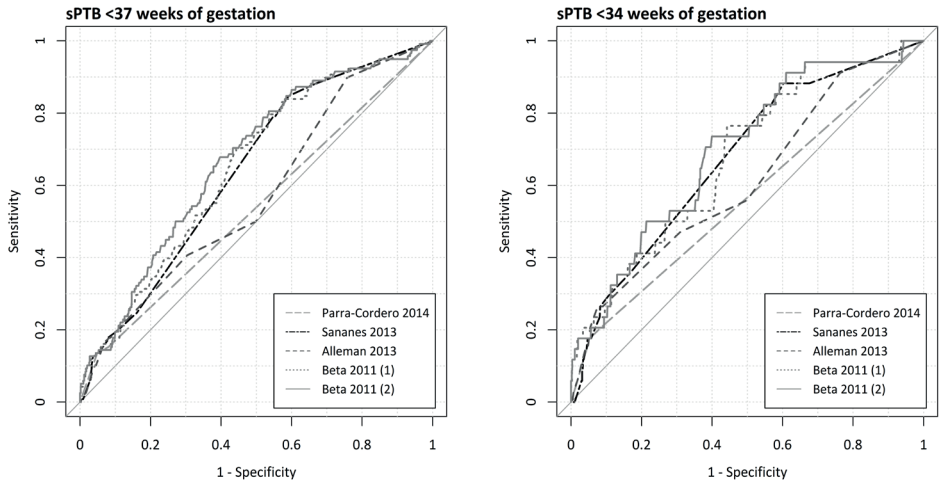


Figure 2.3 ROC curves of externally validated first trimester prediction models for spontaneous preterm birth (sPTB) <37 weeks and <34 weeks of gestation

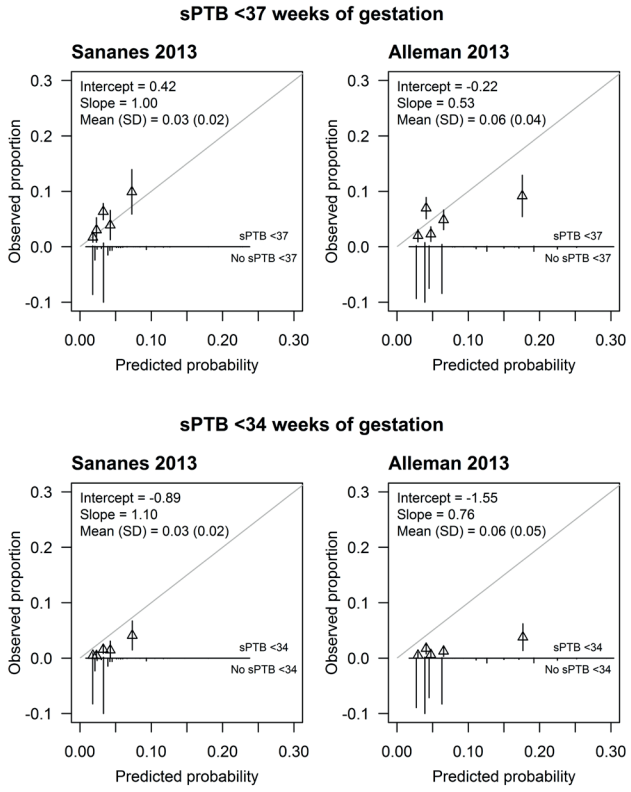


Figure 2.4 Calibration plots of externally validated first trimester prediction models for spontaneous preterm birth (sPTB) <37 weeks and <34 weeks of gestation. The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines)

Table 2.4 Sensitivities, specificities and predictive values at different risk thresholds for recalibrated model 2 of Beta et al., outcome sPTB <37 weeks of gestation

Risk threshold[a], %	High risk, % (n/n)			Sensitivity, % (n/n)			Specificity, % (n/n)			PPV, % (n/n)			NPV, % (n/n)		
	All	Nulli-parous	Multi-parous	All	Nulli-parous	Multi-parous	All	Nulli-parous	Multi-parous	All	Nullipa-rous	Multi-parous	All	Nulli-parous	Multi-parous
2	98.3 (2496/2540)	100 (1284/1284)	96.5 (1212/1256)	99.2 (117/118)	100 (77/77)	97.6 (40/41)	1.8 (43/2422)	0 (0/1207)	3.5 (43/1215)	4.7 (117/2496)	6.0 (77/1284)	3.3 (40/1212)	97.7 (43/44)	100 (0/0)	97.7 (43/44)
3	70.8 (1799/2540)	98.5 (1265/1284)	42.5 (534/1256)	89.8 (106/118)	100 (77/77)	70.7 (29/41)	30.1 (729/2422)	1.6 (19/1207)	58.4 (710/1215)	5.9 (106/1799)	6.1 (77/1265)	5.4 (29/534)	98.4 (729/741)	100 (19/19)	98.3 (710/722)
4	51.7 (1312/2540)	83.3 (1069/1284)	19.3 (243/1256)	76.3 (90/118)	89.6 (69/77)	51.2 (21/41)	49.5 (1200/2422)	17.1 (207/1207)	81.7 (993/1215)	6.9 (90/1312)	6.5 (69/1069)	8.6 (21/243)	97.7 (1200/1228)	96.3 (207/215)	98.0 (993/1013)
5	28.1 (715/2540)	41.4 (532/1284)	14.6 (183/1256)	50.0 (59/118)	49.4 (38/77)	51.2 (21/41)	72.9 (1766/2422)	59.1 (713/1207)	86.7 (1053/1215)	8.3 (59/715)	7.1 (38/532)	11.5 (21/183)	96.8 (1766/1825)	94.8 (713/752)	98.1 (1053/1073)
6	15.2 (385/2540)	20.3 (261/1284)	9.9 (124/1256)	26.3 (31/118)	16.9 (13/77)	43.9 (18/41)	85.3 (2068/2422)	79.5 (959/1207)	91.3 (1109/1215)	8.1 (31/385)	5.0 (13/261)	14.5 (18/124)	96.0 (2068/2155)	93.7 (959/1023)	98.0 (1109/1132)
7	10.7 (271/2540)	13.9 (179/1284)	7.3 (92/1256)	19.5 (23/118)	9.1 (7/77)	39.0 (16/41)	89.8 (2174/2422)	85.7 (1035/1207)	93.7 (1139/1215)	8.5 (23/271)	3.9 (7/179)	17.4 (16/92)	95.8 (2174/2269)	93.7 (1035/1105)	97.9 (1139/1164)

[a]Predicted risk at or above this level was considered as high risk. NPV, negative predictive value; PPV, positive predictive value

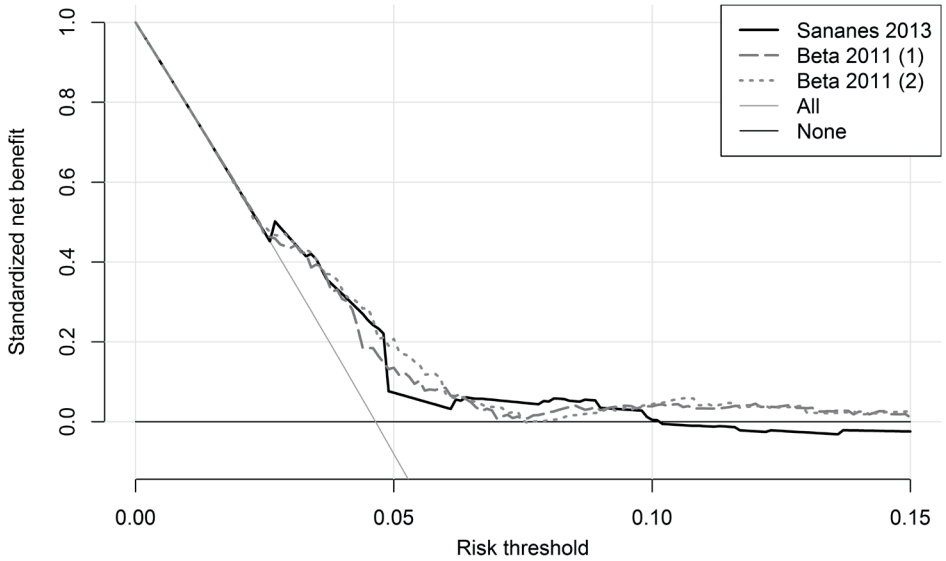


Figure 2.5 Decision curve analysis of three best performing models for the risk of spontaneous preterm birth <37 weeks of gestation. Decision curve analysis assesses the net benefit (vertical axis; proportion of true positives and false positives) of the prediction models over a range of risk thresholds compared to considering all (solid grey line) and no women (horizontal solid black line) to be at high risk for sPTB

Discussion

Main findings

In this systematic review we provided an overview of the currently available prediction models of sPTB based on routine clinical parameters. We identified four articles describing five models fulfilling the eligibility criteria. Assessment of methodological quality revealed several shortcomings in reporting of models. Furthermore, there is a moderate to high risk of bias in the development of the models according to the CHARMS criteria. External validation resulted in a decreased discriminative ability for all models. Model 2 of Beta et al. had the highest AUROC (sPTB <37 weeks: 0.67, and sPTB <34 weeks: 0.70) after validation. This model was based on age, ethnicity, height, method of conception, nulliparous fetal loss, nulliparous late miscarriage, prior PTB (subcategories), prior iatrogenic PTB, prior term delivery, and smoking. The model of Sananes et al. showed the best calibration (slope of one) for sPTB <37 weeks of gestation.

Interpretation

Our systematic review identified a moderate reporting quality of most studies according to the CHARMS criteria. Reporting shortcomings were also noted in a general systematic review about obstetric prediction models [15]. The recently published transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement, may lead to improvements in the reporting quality of future studies [30]. Risk of bias assessment revealed a moderate to high risk of bias in three out of four studies. The main sources of bias were in the domains of analysis, attrition and modeling. All studies selected predictors

on the basis of statistical significance, which leads to a model that fits the data too closely [24,31]. Next, continuous variables were often dichotomized, in example age and BMI in two of our selected models, leading to loss of information [32]. Moreover, only one study, Beta et al., applied the regression shrinkage technique and only Alleman et al. performed an internal validation by bootstrapping. The methodological limitations mentioned could have been one of the reasons why the reported model performance was not achieved in our validation cohort.

Only Sananes et al. mentioned that they validated their model in another population, but the results were not reported. To our knowledge, no other independent external validation study of prediction models for sPTB exists. External validation is recommended to assess the generalizability to other 'related' populations [24]. Our comprehensive independent validation study indicated that all models overestimated performance measures. This illustrates the need for external validation of models before clinical implementation.

Nevertheless, performance measures do not indicate whether a model is clinically useful. Assessment of the clinical utility of the best discriminating model showed a very high false positive rate at acceptable sensitivity rates. These cut-off points result in a major proportion of nulliparous women being unnecessarily considered to be at high risk. Furthermore, for multiparous women the most important predictors are derived from a previous sPTB. In summary, we think that the clinical utility of currently available models is low.

Implications

This systematic review demonstrates shortcomings in the quality and performance of existing non-invasive prediction models for sPTB. Improvement of non-invasive models is necessary. The currently available prediction models mainly rely on previous PTB as predicting variable. However, models mainly relying upon a prior event as the discriminative factor do not add much clinical value since caregivers are already aware that these women are at high risk. Obstetric care would benefit from valid prediction of sPTB in nulliparous women [11].

Future research should focus on the variety of published association studies when selecting candidate predictors. Another important well-known risk factor is cervical surgery [10,33]. However, only a minority of women will be identified as high risk by adding this predictor [11]. Other routine clinical parameters that may also contribute to the prediction of sPTB in nulliparous women are: socio-economic status, psychological characteristics, family history, medical history, and smoking status [10]. Predictive performance of a model might improve by taking into account biomarkers or ultrasound imaging (i.e. cervical length). A few models based on cervical length measurements and biomarkers such as pregnancy-associated plasma protein A (PAPP-A) or alpha-fetoprotein (AFP) have been published [29,34,35]. The reported discriminative performance of these models was only slightly better than the performance of models using maternal characteristics alone. We focused in this review on routine clinical parameters, as these 'specialized' tests are not always routinely performed or readily available in general care, and may generate substantial additional costs [36]. Lastly, different modeling methods can be employed as well. In this review, all selected studies used a multiple logistic regression model. Other methods that can be used are machine learning methods using health records, such as tree-based algorithms or neural networks [37,38]. However, despite all efforts, sPTB may remain a tough outcome to predict due to its heterogeneous and often unknown causes [2].

Nevertheless, a future model with a moderate performance may still be useful. The tradeoff

between the benefit of identifying women at high risk and the false positive rate is important. Using cervical length screening in all women results in the need to screen relatively high numbers of women [11]. A non-invasive model combined with a high sensitivity cut-off point will be able to identify women at very low risk of sPTB who could be excluded from cervical length screening, resulting in the need to screen a smaller number of women. Furthermore, such an approach creates the opportunity to identify women at high risk whom may benefit from preventive interventions such as progesterone treatment [3-5].

Strengths and limitations

To our knowledge, this is the first systematic review of studies reporting non-invasive prediction models for the risk of sPTB. We had to exclude several published models as three models contained predictors which are not available in the first 16 weeks of pregnancy, in example fetal gender, since this is crucial for early prediction of sPTB. Moreover, three other models did not provide the algorithm, which is essential for independent external validation.

A strength of our study is that we validated all included prediction models in a large independent multicentre prospective cohort of unselected pregnant women. The data were very complete with a maximum of only 1.2% of missing values. However, although our cohort contained a sufficient number of cases for sPTB <37 weeks of gestation, there were only 34 cases for the secondary outcome sPTB <34 weeks of gestation. An inadequate sample size decreases the precision of external validation measures [22,39].

Our cohort might suffer from treatment bias to a small extent since we did not exclude women who had received treatment such as a cerclage or tocolysis. This may have resulted into the prevention of sPTB and thus an underestimation of model discrimination and calibration [40]. One of the selected studies, Alleman et al., explicitly reported exclusion of women undergoing cerclage or tocolysis from their study population [27]. Parra-Cordero et al. only excluded women with a history of cerclage [28].

Conclusion

This review revealed several reporting and methodological shortcomings of published prediction models for sPTB. Our external validation indicated that none of the models had the ability to adequately predict sPTB in our population. Obstetric care would benefit most from models predicting sPTB accurately among nulliparous women since most of these women are indicated as low risk in current practice.

Supplementary Files

Supplementary File S2.1	Search strategy
Supplementary Table S2.1	Framework of systematic research aim according to the CHARMS checklist [19]
Supplementary Figure S2.1	Flowchart study selection
Supplementary Table S2.2	Definition and assessment predictors included prediction models for spontaneous preterm birth
Supplementary Table S2.3	Model algorithms for prediction of spontaneous preterm birth.
Supplementary Table S2.4	Data extraction of included studies according to the CHARMS checklist [19]
Supplementary Table S2.5	Risk of bias assessment according to the CHARMS checklist [19] and a study of Smit et al. (2015) [20]
Supplementary Table S2.6	Characteristics of pregnancies in the observed and imputed validation cohort
Supplementary Table S2.7	Baseline characteristics original cohorts and validation cohort
Supplementary Figure S2.2	Calibration plots of recalibrated first trimester prediction models for spontaneous preterm birth (sPTB) <37 weeks of gestation. The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF, correction factor
Supplementary Figure S2.3	Calibration plots of recalibrated first trimester prediction models for spontaneous preterm birth (sPTB) <34 weeks of gestation. The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF, correction factor

References

1. Euro-Peristat Project with SCPE and EUROCAT, European Perinatal Health Report. The Health and care of pregnant women and babies in Europe in 2010. 2013: p. 113.
2. Goldenberg, R.L., Culhane, J.F., et al., Epidemiology and causes of preterm birth. Lancet, 2008. 371(9606): p. 75-84.
3. Saigal, S. and Doyle, L.W., An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet, 2008. 371(9608): p. 261-9.
4. Iams, J.D., Romero, R., et al., Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. Lancet, 2008. 371(9607): p. 164-75.
5. Dodd, J.M., Jones, L., et al., Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. The Cochrane Library, 2013.
6. Romero, R., Conde-Agudelo, A., et al., Vaginal progesterone for preventing preterm birth and adverse perinatal outcomes in singleton gestations with a short cervix: a meta-analysis of individual patient data. Am J Obstet Gynecol, 2018. 218(2): p. 161-180.
7. Gilner, J. and Biggio, J., Management of Short Cervix during Pregnancy: A Review. Am J Perinatol, 2016. 33(3): p. 245-52.
8. Alfirevic, Z., Stampalija, T., et al., Cervical stitch (cerclage) for preventing preterm birth in singleton pregnancy. Cochrane Database Syst Rev, 2017. 6: p. Cd008991.
9. Zheng, L., Dong, J., et al., Cervical pessaries for the prevention of preterm birth: a systematic review and meta-analysis. J Matern Fetal Neonatal Med, 2017: p. 1-10.
10. Koullali, B., Oudijk, M.A., et al., Risk assessment and management to prevent preterm birth. Semin Fetal Neonatal Med, 2016. 21(2): p. 80-8.
11. Ven, J., Os, M.A., et al., The capacity of mid-pregnancy cervical length to predict preterm birth in low-risk women: a national cohort study. Acta obstetricia et gynecologica Scandinavica, 2015. 94(11): p. 1223-1234.
12. Esplin, M.S., Elovitz, M.A., et al., Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. Jama, 2017. 317(10): p. 1047-1056.
13. Goffinet, F., Primary predictors of preterm labour. BJOG: An International Journal of Obstetrics & Gynaecology, 2005. 112(s1): p. 38-47.
14. Honest, H., Bachmann, L.M., et al., The accuracy of risk scores in predicting preterm birth--a systematic review. J Obstet Gynaecol, 2004. 24(4): p. 343-59.
15. Kleinrouweler, C.E., Cheong-See, F.M., et al., Prognostic models in obstetrics: available, but far from applicable. American journal of obstetrics and gynecology, 2016. 214(1): p. 79-90. e36.
16. Steyerberg, E.W., Vickers, A.J., et al., Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology, 2010. 21(1): p. 128-38.
17. Steyerberg, E.W. and Vergouwe, Y., Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J, 2014. 35(29): p. 1925-31.
18. Debray, T.P., Damen, J.A., et al., A guide to systematic review and meta-analysis of prediction model performance. BMJ, 2017. 356: p. i6460.
19. Moons, K.G., de Groot, J.A., et al., Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med, 2014. 11(10): p. e1001744.
20. Smit, H.A., Pinart, M., et al., Childhood asthma prediction models: a systematic review. Lancet Respir Med, 2015. 3(12): p. 973-84.
21. Meertens, L.J.E., Scheepers, H.C., et al., External Validation Study of First Trimester Obstetric Prediction Models (Expect Study I): Research Protocol and Population Characteristics. JMIR Res Protoc, 2017. 6(10): p. e203.
22. Vergouwe, Y., Steyerberg, E.W., et al., Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol., 2005. 58(5): p. 475-83.

23. Van Buuren, S., Flexible imputation of missing data. 2012: CRC press.
24. Steyerberg, E., Clinical prediction models: a practical approach to development, validation, and updating. 2008: Springer Science & Business Media.
25. Vickers, A.J. and Elkin, E.B., Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making, 2006. 26(6): p. 565-74.
26. Sananes, N., Meyer, N., et al., Prediction of spontaneous preterm delivery in the first trimester of pregnancy. European Journal of Obstetrics & Gynecology and Reproductive Biology, 2013. 171(1): p. 18-22.
27. Alleman, B.W., Smith, A.R., et al., A proposed method to predict preterm birth using clinical data, standard maternal serum screening, and cholesterol. American journal of obstetrics and gynecology, 2013. 208(6): p. 472. e1-472. e11.
28. Parra-Cordero, M., Sepúlveda-Martínez, A., et al., Is there a role for cervical assessment and uterine artery Doppler in the first trimester of pregnancy as a screening test for spontaneous preterm delivery? Ultrasound in Obstetrics & Gynecology, 2014. 43(3): p. 291-296.
29. Beta, J., Akolekar, R., et al., Prediction of spontaneous preterm delivery from maternal factors, obstetric history and placental perfusion and function at 11–13 weeks. Prenatal diagnosis, 2011. 31(1): p. 75-83.
30. Collins, G.S., Reitsma, J.B., et al., Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC medicine, 2015. 13(1): p. 1.
31. Harrell, F., Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2015: Springer.
32. Royston, P., Altman, D.G., et al., Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in medicine, 2006. 25(1): p. 127-141.
33. Castanon, A., Landy, R., et al., Risk of preterm delivery with increasing depth of excision for cervical intraepithelial neoplasia in England: nested case-control study. Bmj, 2014. 349: p. g6223.
34. van Ravenswaaij, R., Tesselaar-van der Goot, M., et al., First-trimester serum PAPP-A and fbeta-hCG concentrations and other maternal characteristics to establish logistic regression-based predictive rules for adverse pregnancy outcome. Prenat Diagn, 2011. 31(1): p. 50-7.
35. Poon, L.C., Nekrasova, E., et al., First-trimester maternal serum matrix metalloproteinase-9 (MMP-9) and adverse pregnancy outcome. Prenat Diagn, 2009. 29(6): p. 553-9.
36. Steyerberg, E.W., Moons, K.G., et al., Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med, 2013. 10(2): p. e1001381.
37. Goldstein, B.A., Navar, A.M., et al., Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J, 2016.
38. Peissig, P.L., Santos Costa, V., et al., Relational machine learning for electronic health record-driven phenotyping. J Biomed Inform, 2014. 52: p. 260-70.
39. Collins, G.S., Ogundimu, E.O., et al., Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Statistics in medicine, 2016. 35(2): p. 214-226.
40. Pajouheshnia, R., Peelen, L.M., et al., Accounting for treatment use when validating a prognostic model: a simulation study. BMC Medical Research Methodology, 2017. 17(1): p. 103.

Supplementary File 2.1. Search strategy

PubMed

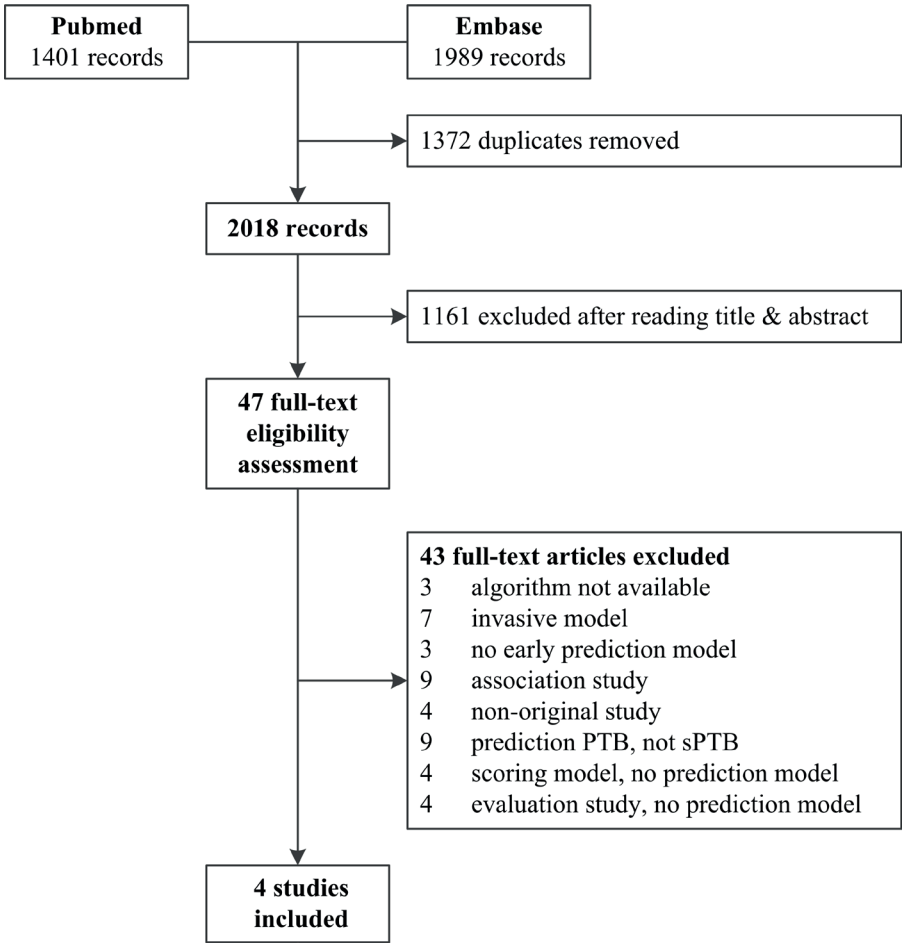
("predictive model"[tiab] OR "predictive models"[tiab] OR prediction[tiab] OR "risk calculator"[tiab] OR "risk calculators"[tiab] OR "risk model"[tiab] OR "risk models"[tiab] OR "risk score"[tiab] OR algorithm*[tiab] OR "risk assessment"[tiab] OR nomogram[tiab] OR "prognostic model"[tiab] OR "prognostic models"[tiab] OR "scoring system"[tiab] OR "scoring systems"[tiab] OR "screening model"[tiab] OR "screening models"[tiab] OR "decision rule"[tiab] OR "decision rules"[tiab]) AND ("preterm labour"[tiab] OR "premature labour"[tiab] OR "premature labor"[tiab] OR "premature delivery"[tiab] OR "premature deliveries"[tiab] OR "premature parturition"[tiab] OR "premature birth"[tiab] OR "preterm labor"[tiab] OR "preterm birth"[tiab] OR "preterm delivery"[tiab] OR "preterm deliveries"[tiab] OR "preterm parturition"[tiab] OR "Premature Birth"[Mesh])

Embase

1. predictive model.ab,ti.
2. predictive models.ab,ti.
3. prediction.ab,ti.
4. risk calculator.ab,ti.
5. risk calculators.ab,ti.
6. risk model.ab,ti.
7. risk models.ab,ti.
8. risk score.ab,ti.
9. algorithm.ab,ti.
10. risk assessment.ab,ti.
11. nomogram.ab,ti.
12. prognostic model.ab,ti.
13. prognostic models.ab,ti.
14. scoring system.ab,ti.
15. scoring systems.ab,ti.
16. screening model.ab,ti.
17. screening models.ab,ti.
18. decision rule.ab,ti.
19. decision rules.ab,ti.
20. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19
21. preterm labour.ab,ti.
22. preterm labor.ab,ti.
23. premature labour.ab,ti.
24. premature labor.ab,ti.
25. premature delivery.ab,ti.
26. premature deliveries.ab,ti.
27. premature parturition.ab,ti.
28. premature birth.ab,ti.
29. preterm birth.ab,ti.
30. preterm delivery.ab,ti.
31. preterm deliveries.ab,ti.
32. preterm parturition.ab,ti.
33. 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34. 20 and 33
35. remove duplicates from 34

Supplementary Table S2.1 Framework of systematic research aim according to the CHARMS checklist [19]

Item	Systematic review aim
Type of prediction model	Prognostic prediction model
Intended scope of review	Reviewing prediction models that may help identifying women who are at high risk for spontaneous preterm birth to aid decision-making regarding preventive interventions or closer monitoring
Type of prediction modelling studies	Model development studies and model validation studies
Target population	Overall pregnant population
Outcome to be predicted	Probability of spontaneous preterm birth
Time span of prediction	First trimester prediction for probability of the outcome in current pregnancy
Intended moment of using the model	First trimester of pregnancy



Supplementary Figure S2.1 Flowchart study selection

Supplementary Table S2.2 (continued) Definition and assessment predictors included prediction models for spontaneous preterm birth

Predictor	Definition (D)/measurement (M) original studies	Definition/measurement validation cohort (Expect Study I)
Ethnicity	Beta 2011 D: Ethnic origin divided into Caucasian, Afro-Caribbean, Indian or Pakistani or Bangladeshi (South Asian), Chinese or Japanese (East Asian) and mixed. M: Self-reported questionnaire 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, ethnicity was divided into ten subgroups: Dutch, Turkish/Kurdish, Moroccan (Moroccan, Algerian, North African), African (African, Surinamese/Antillean of Negroid origin), Hindustani (Hindustani, Pakistani, Indian, Surinamese / Antillean of Hindu origin), Middle East (Iran, Iraqi, Afghan), Asian (Chinese, Japanese, Indonesian, Albanian, Vietnamese), Other Western (European, North American, Australian), Other Non-Western (South and Central American), and mixed. Ethnicity was recoded to Caucasian, Asian, Afro-Caribbean, Hispanic, and mixed (combination of other categories). Subdivision of Asian ethnicity was based on country of birth biological parents. Beta 2011: we added women with an Asian ethnicity other than South Asian or East Asian to the category mixed. Hispanics were categorized as Caucasians.
Height, cm	Beta 2011 D: Continuous in centimetres. M: Height measured at routine assessment at 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, height in centimetres.
History of fetal loss <16 weeks of gestation	Beta 2011 D: Previous miscarriage or termination before 16 weeks. M: Self-reported questionnaire 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, previous pregnancies (miscarriages and terminations <16 weeks of gestation). Obstetric records were checked for discrepancies.
History of iatrogenic preterm birth	Beta 2011 D: Parous iatrogenic preterm delivery before 37 weeks. M: Self-reported questionnaire 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, previous pregnancies (gestational age at delivery and spontaneous onset labour). Obstetric records were checked for additional information about onset of labour. We defined history of iatrogenic preterm delivery as a prior iatrogenic preterm birth ≥24 weeks of gestation.

Supplementary Table S2.2 (continued) Definition and assessment predictors included prediction models for spontaneous preterm birth

Predictor	Definition (D)/measurement (M) original studies	Definition/measurement validation cohort (Expect Study I)
History of live birth	Alleman 2013 D: Previous live birth. M: Neonatal birth certificates.	Self-reported pregnancy questionnaire 1, previous pregnancies (live birth).
History of preterm birth	Parra-Cordero 2014 D: Prior preterm delivery <37 weeks of gestation. M: Interview before the ultrasound scan at 11^{+0}-13^{+6} weeks of gestation. Sananes 2013 D: Previous preterm deliveries categorized as 24-27 weeks, 28-33 weeks, and 34-36 weeks of gestation. M: Electronic medical records. Alleman 2013 D: Previous preterm birth. M: Neonatal birth certificates.	Self-reported pregnancy questionnaire 1, previous pregnancies (gestational age at delivery) and checked for discrepancies by obstetric record. We defined preterm birth as a delivery <37 weeks of gestation. Categorical variables generated according to definition original prediction model.
History of spontaneous preterm birth	Sananes 2013 D: History of miscarriage between 16 and 24 weeks of gestation. M: Electronic medical records. Beta 2011 D: Previous spontaneous deliveries ≥24 weeks of gestation, subdivided into: 24-27^{+6} weeks, 28-30^{+6} weeks, 31-33^{+6} weeks, and 34-36^{+6} weeks of gestation. In a second model, the categories were subdivided according to the number or previous preterm deliveries: one or at least two spontaneous deliveries between 16-30^{+6} weeks of gestation with and without additional deliveries between 31-36^{+6} weeks or ≥37 weeks of gestation, and spontaneous delivery between 31-36^{+6} weeks of gestation with and without additional deliveries ≥37 weeks of gestation. M: Self-reported questionnaire 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, previous pregnancies (gestational age at delivery and spontaneous onset labour). Obstetric records were checked for additional information about onset of labour. Categorical variables generated according to definition original prediction model.
History of term delivery	Sananes 2013 D: Term delivery ≥37 weeks of gestation. M: Electronic medical records. Beta 2011 D: Deliveries at or after 37 weeks. M: Self-reported questionnaire 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, previous pregnancies (gestational age at delivery). We defined term delivery as a delivery ≥37 weeks of gestation.

Supplementary Table S2.2 (continued) Definition and assessment predictors included prediction models for spontaneous preterm birth

Predictor	Definition (D)/measurement (M) original studies	Definition/measurement validation cohort (Expect Study I)
Smoking	Parra-Cordero 2014 D: Smoking during pregnancy. M: Interview before the ultrasound scan at 11^{+0}-13^{+6} weeks of gestation. Sananes 2013 D: Smoking status during pregnancy. M: Electronic medical records. Beta 2011 D: Cigarette smoker. M: Self-reported questionnaire 11^{+0}-13^{+6} weeks of gestation.	Self-reported pregnancy questionnaire 1, cigarette smoking status (non-smoker, stopped during pregnancy, current smokers) and number of cigarettes a day. Smoking status was recoded to definition original prediction model. We defined cigarette smoking as current smoker at completion pregnancy questionnaire 1.

Supplementary Table S2.3 Model algorithms for prediction of spontaneous preterm birth

Original study	The probability of spontaneous preterm birth was calculated as $e^{lp}/(1+ e^{lp})$, where:
Parra-Cordero 2014	Lp = α + 1.163 (if nulliparous and smoking) + 1.526 (if parous with previous preterm delivery).
Sananes 2013	Lp = -3.3772 + 0,2490 (if age ≤22 or ≥35) + 0,3290 (if BMI ≤19 kg/m^2) + 0,2880 (if smoking) + 0,7722 (if prior late miscarriage 16-23 weeks of gestation) + 1,6249 (if prior preterm delivery 24-27 weeks of gestation) + 0,6622 (if prior preterm delivery 28-33 weeks of gestation) + 1,1326 (if prior preterm delivery 34-36 weeks of gestation) - 0,62 (if prior term delivery ≥37 weeks of gestation).
Alleman 2013	Lp = -2.6603 – 0.4949 (if maternal education postsecondary degree) + 1.0524 (if diabetes mellitus) + 1.5801 (if prior preterm delivery) – 0.3396 (if prior live birth) + 1.4385 (if BMI <18.5 kg/m^2) + 0.7352 (if BMI >40 kg/m^2).
Beta 2011	Lp = α + 0.025 (age, years) – 0.019 (height, cm) + 0.589 (if Afro-Caribbean) + 0.554 (if South Asian) + 0.168 (if East Asian) – 0.4 (if Mixed) + 0.567 (if smoker) + 0.535 (if assisted conception) + 0.239 (if nulliparous, fetal loss <16 weeks of gestation) + 1.976 (if nulliparous, miscarriage at 16-23 weeks of gestation) + 1.734 (if parous, preterm delivery 24-27 weeks of gestation) + 1.503 (if parous, preterm delivery 28-30 weeks of gestation) + 1.142 (if parous, preterm delivery 31-33 weeks of gestation) + 0.907 (if parous, preterm delivery 34-36 weeks of gestation) – 0.414 (if parous, term delivery >37 weeks of gestation) + 0.309 (if parous, iatrogenic preterm delivery). Lp = α + 0.027 (age, years) – 0.019 (height, cm) + 0.568 (if Afro-Caribbean) + 0.554 (if South Asian) + 0.149 (if East Asian) – 0.387 (if Mixed) + 0.595 (if smoker) + 0.538 (if assisted conception) + 1.766 (if delivery at 16-30 weeks of gestation, one event) + 2.93 (if delivery at 16-30 weeks of gestation, two events) + 1.992 (if delivery at 16-30 weeks of gestation, one event plus 31-36 weeks of gestation) + 0.437 (if delivery at 16-30 weeks of gestation, one event plus ≥37 weeks of gestation) + 2.277 (if delivery at 16-30 weeks of gestation, two events plus ≥37 weeks of gestation) + 0.846 (if delivery 31-36 weeks of gestation) + 0.627 (if delivery 31-36 weeks of gestation plus ≥37 weeks) – 0.54 (if delivery ≥37 weeks of gestation).

BMI, body mass index; Lp, linear predictor

Supplementary Table S2.4 Data extraction of included studies according to the CHARMS checklist[19]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Source of data	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	Prospective cohort	Registry data	Registry data	Prospective cohort
Participants	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria) Participant description Details of treatments received, if relevant Study dates	3310 participants recruited between 11^{+0}-13^{+6} GA in University of Chile Hospital. Study dates NR. Exclusion criteria: multiple pregnancy, iatrogenic delivery <34 GA, early-onset pre-eclampsia (delivery <34 GA, blood pressure >140/90 mmHg, and proteinuria >300 mg/24h), early-onset SGA (birth weight <10th percentile and delivery before 34 GA), spontaneous miscarriage, intrauterine fetal death, fetal congenital malformations, chromosomal abnormalities, placental abruption, and patients with a history of cervical surgery or cerclage. Patients may be treated for spontaneous onset of preterm delivery, not specifically reported. Baseline characteristics reported.	Antenatal consultation before 14 weeks and delivery after 24 weeks in single medical center France between 1 January 2000 and 30 November 2011. Exclusion criteria: multiple pregnancy, pregnancies with imprecise dates, fetal deaths, medical terminations, and induced preterm births before 37 GA. Patients may be treated for spontaneous onset of preterm delivery, not specifically reported. Baseline characteristics reported.	All Iowa women that underwent routine prenatal testing in first trimester from May 2009 until November 2010 were included. Exclusion criteria: multiple pregnancy, delivery <20 GA, congenital anomaly, birth weight for gestational age >3 standard deviations from mean, serious infection, second pregnancy of mothers with more than 1 pregnancy during collection, cerclage, and tocolysis. Patients treated for spontaneous onset of preterm delivery by cerclage and tocolysis were excluded. Baseline characteristics NR (only for included predictors PTB not sPTB).	Women attending routine first hospital visit (King's College London) between 11^{+0}-13^{+6} GA from March 2006 and September 2009. Exclusion criteria: multiple pregnancy, major fetal abnormalities, termination, miscarriage or fetal death before 24 weeks and those with an iatrogenic delivery before 34 weeks. Patients may be treated for spontaneous onset of preterm delivery, not specifically reported. Baseline characteristics reported.

Supplementary Table S2.4 (continued) Data extraction of included studies according to the CHARMS checklist[19]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Sample size	Number of participants and number of outcomes/events Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable)	31 cases of sPTB and 3310 participants. 15.5 events/predictor (2 predictors).	31834 pregnancies with 1188 cases. However patients with missing data not taken into account. Development population 17341. Number of cases NR.	153 cases of sPTB and 2699 participants. 25.5 events/predictor (6 predictors).	353 cases of sPTB and 33370 participants. Model 1 20.1 events/predictor (16 predictors). Model 2 (subdividing number previous events of preterm delivery) also 20.1 events/ predictor (16 predictors).
Missing data	Number of participants with any missing value (include predictors and outcomes) Number of participants with missing data for each predictor Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	Missing data NR.	Participants with missing data described (smoking only available for 45% of women, number of women with PAPP-A assay, and number of women with missing outcome data n=693). Complete-case analysis performed.	Missing data predictors NR. 19 women excluded with unreliable birth records. Only participants included with serum collection in first and second trimester (12057 women to 2976 women with two screenings).	Missing outcome data mentioned (n=2005) and those women were excluded. Missing data predictors NR.
Model development	Modelling method (e.g., logistic, survival, neural network, or machine learning techniques) Modelling assumptions satisfied Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome) Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion) Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	Logistic regression analysis. Predictors selected based on significant contribution (p-value <0.05). Shrinkage NR.	Univariate analysis was performed to identify predictors that explained the occurrence of preterm delivery. Then ascending stepwise multivariate logistic regression was performed, variables were left in model at a p-value <0.05. Shrinkage NR.	All covariates potentially related to PTB based on previous studies were screened for entry into a final predictive model and considered for selection at p-value <0.10 using a X^2 test or simple logistic regression. Multivariate logistic regression using a logit link function was used to determine which covariates significantly predicted PTB. The final model was determined using forward, backward, and stepwise selection with Aikike Information Criterion. Shrinkage NR.	Univariate analysis was performed to examine the individual variables contributing significantly to preterm delivery. Logistic regression analysis with backward stepwise elimination of variables was used to develop the model. Shrinkage factor was calculated (0.90 and 0.91).

Supplementary Table S2.4 (continued) Data extraction of included studies according to the CHARMS checklist[19]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Outcome(s) to be predicted	Definition and method for measurement of outcome Was the same outcome definition (and method for measurement) used in all patients? Type of outcome (e.g., single or combined endpoints) Was the outcome assessed without knowledge of the candidate predictors (in example, blinded)? Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)? NA Time of outcome occurrence or summary of duration of follow-up	sPTB defined: sPTB <34 GA (including that following PPROM). Outcome obtained from delivery room database or by contacting patients by telephone. Not mentioned that outcome assessment was blinded.	sPTB defined: sPTB <37 GA. Outcome obtained from registry (not blinded).	sPTB defined: sPTB <37 GA. Women who delivered vaginally with induction or via cesarean section after labour by induction but without preterm PROM (PPROM), tocolysis, or precipitous labor, and women who delivered by cesarean section without labor and no history of cesarean section, tocolysis, or breech presentation were identified as nonspontaneous preterm birth. Outcome obtained from registry (not blinded).	sPTB defined: sPTB <37 GA, which included those with spontaneous onset of labor and those with PPROM. Outcome obtained from the maternity computerized records or the general medical practitioners of the women and was recorded in database. Not mentioned that outcome assessment was blinded.
Candidate predictors	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics) Definition and method for measurement of candidate predictors Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation) Were predictors assessed blinded for outcome, and for each other (if relevant)? Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	Predictors are demographics and patient history obtained from an interview immediately before ultrasound scan early in pregnancy and recorded in a computer database. Predictors were assessed in early pregnancy (before outcome). Predictors are dichotomous variables.	Predictors are demographics and patient history obtained from registry. Predictors and outcome extracted at same time from registry (not blinded). Predictors are dichotomous variables.	Predictors are demographics and patient history obtained from registry. Predictors and outcome extracted at same time from registry (not blinded). Predictors are dichotomous variables.	Predictors are demographics and patient history obtained at screening for aneuploidies (before outcome). Predictors are continues and dichotomous variables.

Supplementary Table S2.4 (continued) Data extraction of included studies according to the CHARMS checklist[19]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Model performance	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	Discrimination and calibration NR. Detection rate 26% at a FPR 8%.	Discrimination AUROC 0.618 (95% CI 0.595-0.641). Calibration NR. Detection rate 23.3% at a FPR 10% with positive predictive value of 7.4% and negative predictive value of 97.2%.	Discrimination AUROC 0.703 (95% CI NR). Calibration NR. Sensitivity 17.5% at specificity of 97.0% and positive and negative predictive values of 26.6% and 95.0%, respectively. Net reclassification improvement for models with biomarkers compared to maternal characteristics.	Discrimination model 1 AUROC 0.668 (95% CI 0.639-0.698), also for nulliparous women AUROC 0.607 (95% CI 0.566-0.649). Discrimination model 1 with obstetric history subdivided according to number of previous preterm deliveries NR. Calibration NR. Detection rate model 1 27.5% at a FPR 10% and for nulliparous the detection rate was 19.5%.
Model evaluation	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators) In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	Development dataset only.	Temporal external validation on a prospective sample using same criteria between 1 December 2011 and 30 June 2012 by same investigators. 2412 pregnancies and 76 cases of sPTB. Sensitivity 18.4% for a specificity of 97.1%. Discrimination and calibration NR. Model not adjusted or updated.	Development dataset only with bootstrapping (1000-fold).	Development dataset only.

Supplementary Table S2.4 (continued) Data extraction of included studies according to the CHARMS checklist[19]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Results	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals) Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance Comparison of the distribution of predictors (including missing data) for development and validation datasets	Odds ratios reported. Intercept NR.	Odds ratios reported. Intercept NR. Full algorithm received from authors by email.	Algorithm and odds ratios not reported. Full algorithm received by email.	Odds ratios reported. Intercept NR.
Interpretation and discussion	Interpretation of presented models (confirmatory, in example, model useful for practice versus exploratory, in example, more research needed) Comparison with other studies, discussion of generalizability, strengths and limitations.	Prior history of preterm delivery and smoking highest risk factors. Indicated 25% of the cases. Doppler and cervical length measurements during first trimester were no useful predictors. Performed comparison with other studies and described strengths and limitations moderately.	Prediction of preterm delivery based on maternal characteristics and obstetric history must be further improved. Tool is better in patients who have already delivered at least once. Further studies necessary to appraise utility of biomarkers and cervical length. Comparison with other studies described. Limitations moderately described.	Combining several readily available risk markers can create a better screening tool than PTB history alone. The model needs verification in other populations. Performed comparison with other studies. Strengths and limitations described in detail.	Despite overall low performance of screening by maternal characteristics and obstetric history, an algorithm combining risk factors can provide patient-specific risks which can be the basis of individualization of subsequent prenatal care. Future studies will define whether the performance can be improved by cervical length measurements and biomarkers. Performed comparison with other studies. Limitations not described.

GA, gestational age in weeks; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; FPR, false positive rate; NR, not reported; (P)PROM, (Preterm) premature rupture of membranes

Supplementary Table S2.5 Risk of bias assessment according to the CHARMS checklist [19] and a study of Smit et al. 2015 [20]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Source of data	Low risk: Prospective cohort study, randomised trials	✓			✓
	High risk: Retrospective cohort study, case-control study		✓	✓	
Participants [20]	Low risk: Selection bias was unlikely Study avoided inappropriate inclusions or exclusions In and exclusion criteria were adequately described Participants were enrolled at a similar presentation of their disease Differences were accounted for by including appropriate predictors in the analysis		✓		✓
	Moderate risk: Not satisfying one of the above or No adequate description of recruitment of study sample No adequate description of the sample for key predictors			✓	
	High risk: Both items were not adequately described	✓			
Predictors [20]	Low risk: Predictor definitions were the same for all participants Predictor measurement was blinded to outcome data All predictors were available at the time the model is intended to be used Predictors were measured with valid and reproducible methods such that misclassification was limited and Predictors were assessed in a similar way for all study participants	✓			✓
	Moderate risk: If one of the criteria was not satisfied		✓	✓	
	High risk: Predictor assessment was not adequately described				
Outcome [20]	Low risk: Outcome was pre-specified, measured with sufficient validity and reproducibility, measured in a similar way for all study participants and if the outcome was assessed independent from assessment of predictors. Note: for easy to obtain predictors such as gender, it is not possible to assess outcome independent of predictor information	✓			✓

Supplementary Table S2.5 (continued) Risk of bias assessment according to the CHARMS checklist [19] and a study of Smit et al. 2015 [20]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Outcome[20]	Moderate risk: Method outcome assessment was described, but does not meet the low risk criteria.		✓	✓	
	High risk: Method for assessment of outcome was not adequately described				
Sample size	Low risk: Sufficient number of participants and number of outcomes Sufficient number events per predictor (≥10 events/predictor)		✓	✓	✓
	Moderate risk: Small population Not adequately described	✓			
	High risk: <10 events per predictor				
Attrition[20]	Low risk: There was no loss-to-follow-up No important differences on key characteristics between included participants and those who were lost-to-follow-up or missing				
	Medium risk: Loss-to-follow-up was lower than 20% and there were no important differences on key characteristics between included participants and those who were lost-to-follow-up or missing OR Loss-to-follow-up was higher than 20% but missing data and loss-to-follow-up were imputed adequately or there were no important differences on key characteristics between included participants and those who were lost-to-follow up or missing		✓	✓	✓
	High risk: Loss-to-follow-up was higher than 20% and/or Important differences on key characteristics between included participants and those who were lost to-follow-up or missing Loss-to-follow-up was not described	✓			

Supplementary Table S2.5 (continued) Risk of bias assessment according to the CHARMS checklist [19] and a study of Smit et al. 2015 [20]

Domain		Parra-Cordero 2014	Sananes 2013	Alleman 2013	Beta 2011
Analysis [20]	Low risk:				
	Relevant aspects of analysis were described allowing to judge the quality of the analysis to be adequate				
	# outcome events per candidate predictor reasonable				
	Missing data handled appropriately or no differences				
	Predictors included independent of p-value				
	Overfitting and optimism accounted for				
	Weights assigned according to regression coefficient				
	Calibration and discrimination assessed				
	Recalibrated or described that it was not needed				
	Moderate risk:		✓		✓
	Relevant aspects of analysis were described allowing to judge the quality of the analysis to be adequate and part or none of the model evaluation items were reported				
	High risk:	✓		✓	
	Not satisfying any/lot of the aspects under low risk of bias				
	High risk of bias quality of the analysis/modeling methods				
Model presentation	Low risk:				
	Complete algorithm (intercept, weighted regression coefficients) reported in the manuscript				
	Medium risk:	✓	✓		✓
	Weighted regression coefficients or odds ratios reported, no intercept given.				
	High risk:			✓	
	Model not reported, impossible to calculate individual probabilities				

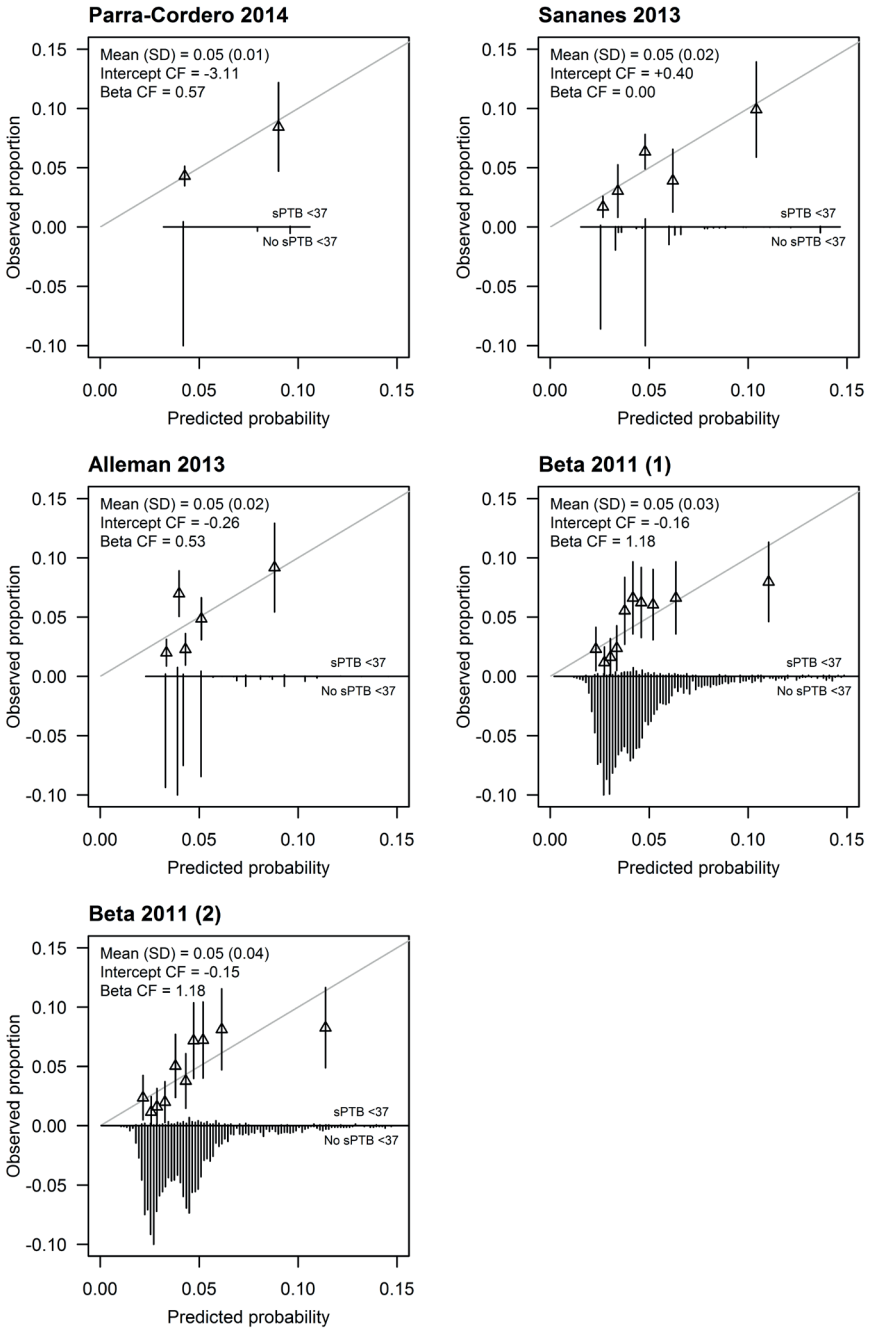
Supplementary Table S2.6 Characteristics of pregnancies in the observed and imputed validation cohort

Characteristics	Missing values, n (%)	Observed Validation cohort Complete cases[a] (n=2502)	Observed Validation cohort Women with missing value(s)[b] (n=38)	Imputed Validation cohort (n=2540)
Age, years	0 (0.0)	30.2 (3.9)	30.9 (4.0)	30.2 (3.9)
Ethnicity	0 (0.0)			
Caucasian		2426 (97.0)	36 (94.7)	2462 (96.9)
Afro-Caribbean		3 (0.1)	0 (0.0)	3 (0.1)
South Asian		3 (0.1)	1 (2.6)	4 (0.2)
East Asian		4 (0.2)	0 (0.0)	4 (0.2)
Other Asian		11 (0.4)	0 (0.0)	11 (0.4)
Hispanic		11 (0.4)	0 (0.0)	11 (0.4)
Mixed		44 (1.8)	1 (2.6)	45 (1.8)
Tertiary level of education	3 (0.1)	1367 (54.6)	13 (34.2)	1380 (54.3)
Height, cm	3 (0.1)	168.8 (6.4)	167.9 (6.7)	168.8 (6.4)
Weight, kg*	5 (0.2)	68.8 (13.0)	71.6 (12.9)	68.9 (13.0)
Body mass index#, kg/m^2	5 (0.2)	24.1 (4.3)	25.3 (4.2)	24.1 (4.3)
Smoking during pregnancy	1 (0.0)	145 (5.8)	4 (10.5)	150 (5.9)
Diabetes mellitus	0 (0.0)	10 (0.4)	0 (0.0)	10 (0.4)
Type 1		8 (0.3)	0 (0.0)	8 (0.3)
Type 2		1 (0.0)	0 (0.0)	1 (0.0)
Other		1 (0.0)	0 (0.0)	1 (0.0)
History of chronic hypertension*	0 (0.0)	23 (0.9)	1 (2.6)	24 (0.9)
Parity	0 (0.0)			
Nulliparous		1280 (51.2)	4 (10.5)	1284 (50.6)
Primiparous		977 (39.0)	26 (68.4)	1003 (39.5)
Multiparous		245 (9.8)	8 (21.1)	253 (9.9)
Conception	0 (0.0)			
Spontaneous		2338 (93.4)	37 (97.4)	2375 (93.5)
Ovulation induction		88 (3.5)	0 (0.0)	88 (3.5)
IVF/ICSI		76 (3.0)	1 (2.6)	77 (3.0)
History of fetal loss <16 weeks of gestation	0 (0.0)	686 (27.4)	16 (42.1)	702 (27.6)
History of recurrent miscarriages (≥3)*	0 (0.0)	46 (1.8)	3 (7.9)	49 (1.9)
Vaginal bleeding (≥2 days)*	0 (0.0)	275 (11.0)	2 (5.3)	277 (10.9)
History of spontaneous preterm delivery	30 (1.2)	75 (3.0)	1 (2.6)	77 (3.0)
16-23 weeks of gestation		3 (0.1)	1 (2.6)	4 (0.2)
24-27 weeks of gestation		7 (0.3)	0 (0.0)	7 (0.3)
28-30 weeks of gestation		2 (0.1)	0 (0.0)	2 (0.1)
31-33 weeks of gestation		13 (0.5)	0 (0.0)	13 (0.5)
34-36 weeks of gestation		52 (2.1)	0 (0.0)	53 (2.1)
History of iatrogenic preterm delivery ≥24 weeks of gestation	29 (1.1)	43 (1.7)	1 (2.6)	44 (1.7)
History of term delivery	29 (1.1)	1128 (45.1)	2 (5.3)	1159 (45.6)
History of live birth	18 (0.7)	1206 (48.2)	15 (39.5)	1239 (48.8)

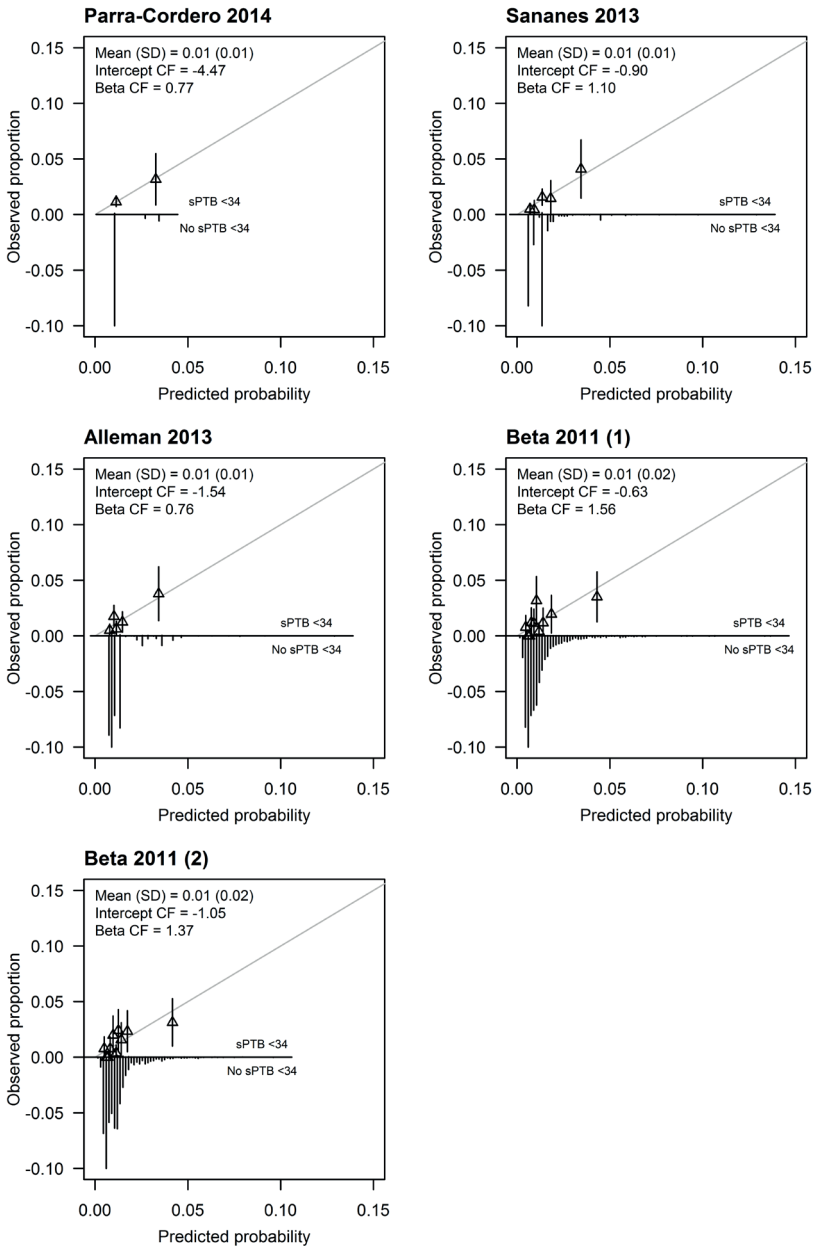
[a]All predictor values of the included models were complete; [b]At least one missing value for a predictor of the included models; *Not a predictor in the included models; #Recoded/calculated on the basis of (imputed) original variables. ICSI, intracytoplasmic sperm injection; IVF, in vitro fertilisation

Supplementary Table S2.7 Baseline characteristics original cohorts and validation cohort

Characteristics	Parra-Cordero 2014[a]		Sananes 2013		Alleman 2013	Beta 2011[a]		Imputed validation cohort					
	sPTB <34 weeks	No sPTB ≥34 weeks	sPTB <37 weeks	No sPTB ≥37 weeks		sPTB <34 weeks	No sPTB ≥34 weeks	All (n=2540)	sPTB <37 weeks (n=118)	No sPTB ≥37 weeks (n=2422)	All (n=2576)	sPTB <34 weeks (n=34)	No sPTB ≥34 weeks (n=2542)
Age, years	30.0	29.0	29.7	30.0		32.6	32.3	30.2 (3.9)	30.1 (3.8)	30.2 (3.9)	30.2 (3.9)	29.9 (3.8)	30.2 (3.9)
≤22 or ≥35			29.6%	26.6%				15.9%	12.7%	15.9%	15.9%	11.8%	15.9%
Ethnicity													
Caucasian						61.5%	72.1%	96.9%	97.5%	96.9%	96.9%	94.1%	96.9%
Afro-Caribbean					5.7%	28.3%	18.7%	0.1%	0.8%	0.1%	0.1%	2.9%	0.1%
South Asian						6.5%	4.3%	0.2%	0.0%	0.2%	0.2%	0.0%	0.2%
East Asian						2.0%	2.0%	0.2%	0.8%	0.1%	0.2%	2.9%	0.1%
Other Asian								0.4%	0.8%	0.4%	0.5%	0.0%	0.5%
Hispanic								0.4%	0.0%	0.5%	0.4%	0.0%	0.4%
Mixed						1.7%	2.9%	1.8%	0.0%	1.9%	1.8%	0.0%	1.8%
Tertiary level of education					48.0%			54.3%	58.5%	54.1%	54.4%	55.9%	54.4%
Height, cm						162.6	164	168.8 (6.4)	167.3 (6.6)	168.9 (6.4)	168.8 (6.4)	166.5 (8.5)	168.8 (6.4)
Weight, kg						66.0	65.5	68.9 (13.0)	65.7 (11.4)	69.0 (13.0)	68.8 (13.0)	67.6 (13.5)	68.9 (13.0)
Body mass index, kg/m^2	23.7	24.4	22.9	23.3				24.1 (4.3)	23.4 (3.8)	24.2 (4.3)	24.1 (4.3)	24.4 (4.6)	24.1 (4.3)
<18.5 kg/m^2					2.3%			3.3%	4.2%	3.3%	3.3%	8.8%	3.3%
≤19.0 kg/m^2			27.8%	21.5%				5.5%	6.8%	5.5%	5.6%	11.8%	5.5%
>40 kg/m^2					6.2%			0.4%	0.0%	0.4%	0.3%	0.0%	0.4%
Smoking during pregnancy	16.1%	9.7%	24.2%	18.7%		12.2%	8.1%	5.9%	6.8%	5.9%	5.9%	8.8%	5.9%
Nulliparous	NR	NR						3.1%	1.7%	3.2%	3.1%	2.9	3.1
Diabetes mellitus					2.1%			0.4%	0.8%	0.4%	0.4%	0.0%	0.4%
Nulliparous	51.6%	49.2%				48.3%		50.6%	65.3%	49.8%	50.7%	64.7%	50.5%
Assisted conception			2.1%	1.2%		6.2%	3.7%	6.5%	3.4%	6.6%	6.6%	5.9%	6.6%
Nulliparous, no previous pregnancies						30.0%	32.2%	40.6%	57.6%	39.8%	40.8%	58.8%	40.5%



Supplementary Figure S2.2 Calibration plots of recalibrated first trimester prediction models for spontaneous preterm birth (sPTB) <37 weeks of gestation. The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF, correction factor



Supplementary Figure S2.3 Calibration plots of recalibrated first trimester prediction models for spontaneous preterm birth (sPTB) <34 weeks of gestation. The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF, correction factor

Chapter 3

Perinatal factors related to pregnancy and childbirth satisfaction: a prospective cohort study

Stéphanie M.P. Lemmens*, Pim van Montfort*, Linda J.E. Meertens, Marc E.A. Spaanderman, Luc J.M. Smits, Raymond de Vries, Hubertina C.J. Scheepers
* Contributed equally

Submitted

Abstract

Background

Satisfaction of pregnancy and childbirth is an important quality measure of maternity care. Satisfaction questionnaires generally result in high scores. However, it has been argued that dissatisfaction relies on a different construct. In response to a worldwide call for obstetric care that is more woman-centred, we identified and described the contributors to suboptimal satisfaction with pregnancy and childbirth.

Methods

A prospective sub cohort of 739 women from a larger cohort (Expect Study I, n= 2,614) received a pregnancy and childbirth satisfaction questionnaire. Scores were transformed to a binary outcome whereby a score <100 points corresponded with less satisfied women. We performed a multiple logistic regression analysis to define independent perinatal factors related to suboptimal satisfaction.

Results

Decreased perceived personal wellbeing, antenatal anxiety, and obstetrician-led care during labour were all independently associated with suboptimal pregnancy and childbirth satisfaction. No difference in satisfaction was found between antenatal care led by a midwife or an obstetrician, but midwife-led antenatal care reduced the odds of suboptimal satisfaction compared to women who were transferred to an obstetrician in the antenatal period. Antenatal anxiety was experienced by 25% of all women and is associated with decreased satisfaction scores.

Discussion

Screening and treatment of women suffering from anxiety might improve pregnancy and childbirth satisfaction, but further research is necessary. Women's birthing experience may improve by reducing unnecessary secondary obstetric care.

Introduction

Satisfaction with care delivered during pregnancy and birth is a topic of increasing interest and is an essential component of quality of obstetric care [1]. In the Netherlands, one in six women has a negative recall of their birth experience [2]. The prevalence of posttraumatic stress disorders resulting from childbirth is estimated at 2.9% [3]. Patient satisfaction and birth experience are important factors influencing short- and long-term outcomes of both mother and child (e.g. postpartum depression, the ability to breast-feed, and child abuse) [1].

Studies of satisfaction with childbirth care are beset by several problems. The role of the healthcare professional is an influential factor shaping a woman's birthing experience [4]. Findings regarding the contribution of several other factors to satisfaction with obstetric care, such as age and pain, are inconsistent [1,5]. Satisfaction questionnaires administered shortly after birth generally result in high satisfaction scores. It has been argued that women may be unable to assess the perceived maternity care properly because they are unaware of other options [6]. Additionally, satisfaction and dissatisfaction are considered to be different constructs rather than a continuum of each other [7]. It may be better to focus on determinants associated with women who are not perfectly satisfied with the obstetric care services received during pregnancy and birth [8]. Focussing on the less satisfied women may result into renewed insights that could improve obstetric care. At present, few studies have focused on determinants of suboptimal care as perceived as such by pregnant women [5].

Antenatal anxiety is related to several adverse pregnancy outcomes (e.g. spontaneous preterm birth, low birth weight [9]) and is associated with a negative subsequent birthing experience [10]. The negative influence of maternal anxiety upon satisfaction levels with received obstetric care services has been reported as well, but mostly for specific subgroups (i.e. women with fear of birth) [11,12].

Women's satisfaction regarding pregnancy and labour is also associated with parity. In general, multiparous women report higher levels of satisfaction as compared to nulliparous women [13,14]. Furthermore, it is likely that multiparous women's expectations concerning their current pregnancy is influenced by their previous experiences with pregnancy, giving birth, and the obstetric care system [11]. These expectations are likely to be more realistic than those of nulliparous women (e.g. prior birth mode is an important prognostic factor for the subsequent mode of birth [15,16]) which expectedly contributes to better satisfaction levels [13].

In this study, we examined the Pregnancy and Childbirth Satisfaction (PCS) of women who recently gave birth in a prospective multicentre cohort. Our objective was to identify factors independently associated with suboptimal PCS and to evaluate the association of maternal anxiety with subsequent PCS in a general population.

Methods

We conducted a cross sectional analysis among a subgroup of a prospective multicentre cohort study, the Expect Study I. The recruitment of this cohort has been described in detail elsewhere [17]. Briefly, women aged 18 years or older were recruited at their first prenatal visit (<16 weeks of pregnancy), in the south region of the Netherlands between 2013 and 2015. Pregnancies ending in a miscarriage (<16 weeks of gestation) or termination before 24 weeks of gestation and women lost-to-follow-up were excluded from the main cohort. Additionally, for this study, we excluded twin pregnancies.

Women were approached for participation in a sub cohort of the Expect Study I after

completion of the first survey (Figure 3.1). Participants in this sub cohort received additional surveys at 24 and 32 weeks of gestation. Moreover, the post-partum survey of the Expect Study, sent 6 weeks after the due date, was extended. The additional questions these women received addressed topics of patient satisfaction, anxiety state, and obstetric care services used. Women who reported preterm birth during the surveys at 24 or 32 weeks were automatically redirected to the post-partum survey.
The medical ethics committee of Maastricht University Medical Centre (MUMC+) evaluated the study protocol and declared that no ethical approval was necessary for this study under Dutch law (METC-17-4-057). All participants gave informed consent.

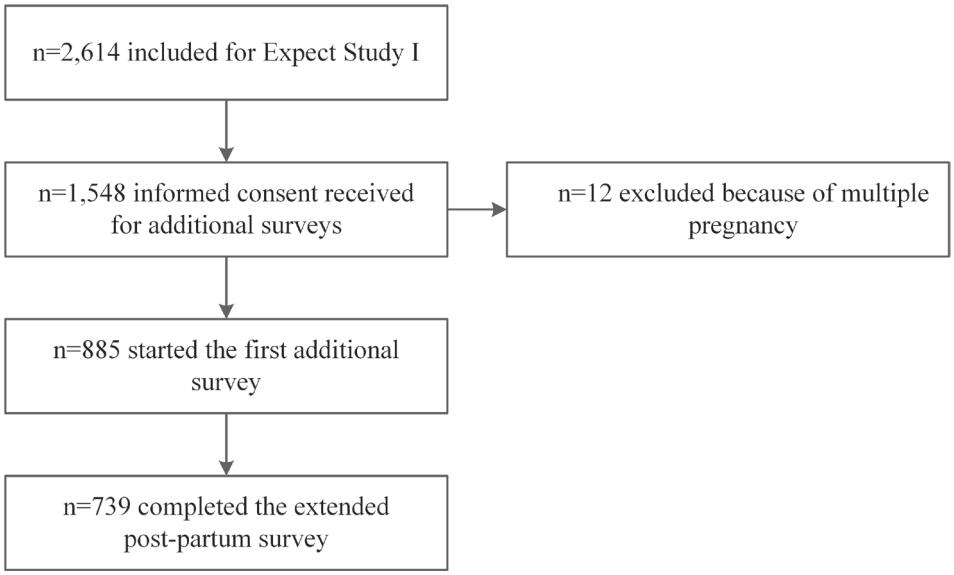


Figure 3.1 Flowchart sub cohort Expect study I

Pregnancy and childbirth satisfaction was measured using the pregnancy and childbirth questionnaire (PCQ). The PCQ is a validated questionnaire measuring perceived quality of care among post-partum women [18]. With 25 questions using a five point Likert scale, it addresses topics specifically related to pregnancy and giving birth. Because the PCQ contains questions addressing childbirth, the PCQ was incorporated in the post-partum questionnaire. PCQ-scores were converted so that higher scores correlates with higher levels of satisfaction. Total scores can range from 25 to 125 points (Cronbach's alpha 0.92). We classified women with a total PCQ score of less than 100 points, mean score < 4 out of 5, less satisfied regarding their childbirth experience. In this study, we classified these women as 'dissatisfied'. Therefore, we will refer to this group from now on as Pregnancy and Childbirth Dissatisfaction (PCD) instead of PCS.
To estimate the strength of the association of independent variables with PCD, we used multiple logistic regression analysis. The independent variables of interest were selected from the literature and consist of maternal factors, neonatal health outcomes, and factors related to the obstetric care received. Additionally, we performed sub-analyses for nulliparous and multiparous women.
Maternal factors included demographic variables such as age, educational degree, and socio-

economic status. Other factors were; antenatal anxiety, parity (nulliparous or multiparous), decrease in perceived personal wellbeing, a neonatal health composite outcome, and a maternal health composite outcome.

Antenatal anxiety levels were measured using the state anxiety items of the State-Trait Anxiety Inventory (STAI), completed at 24 weeks of gestation. The STAI is a validated and commonly used inventory for the measurement of the general anxiety state. Consisting of 20 items using a 4 point Likert scale, STAI scores can range from 20-80. Higher STAI scores represent a higher state of anxiety [19]. We used a threshold of 39 points to identify antenatal anxiety as this cut-off has been suggested to detect clinical significant anxiety symptoms [20].

Socio-economic status was estimated using postal codes and corresponding socio-economic status scores provided by the Dutch government [21].

A decrease in perceived personal wellbeing was defined as a postpartum self-report score (scale 0-100) that was at least 10 points lower than the health status reported at enrollment (<16 weeks of gestation). Personal wellbeing was measured with the Euroqol Visual Analogue Scale [22].

We defined the maternal health composite outcome, a binary outcome, as an occurrence of either pre-eclampsia, gestational diabetes, postpartum haemorrhage (reported blood loss >1000ml), or admission to an intensive or high care unit.

The neonatal composite outcome, a binary outcome defined in Expect Study I [17], was defined as an occurrence of one of the following situations; perinatal death within seven days after birth, asphyxia (Apgar score <7 after 5 minutes), admission to a neonatal intensive care unit (NICU) within 28 days after birth, birthweight <2.3 weight percentile, birth before 32 completed weeks of pregnancy. The birthweight percentile was assessed using Dutch customised birth weight curves which correct for gestational age, ethnicity, gender and parity [23].

Parity and all items of both the neonatal and maternal composite health outcomes, were retrieved from discharge letters, medical records, and the questionnaires. In case of discrepancies, we contacted the corresponding healthcare professional for the final decision.

Independent variables related to the obstetric care services received were: healthcare professional in lead during antenatal care until at least 34 weeks of gestation (categorical variable: autonomous midwife in a primary care setting, obstetrician in a secondary care setting, or both as a result of transfer of care); healthcare professional during labour (categorical variable: midwife, obstetrician, or both as a result of transfer during labour), birth mode (categorical variable: spontaneous vaginal birth, instrumental vaginal birth, or cesarean section), and usage of analgesics (epidural analgesia, intravenous remifentanil) during labour (yes/ no). The variable 'transfer of care' refers to transfer in only one direction, namely from midwife (primary care) to obstetrician (secondary care). In case of antenatal or intrapartum transfer of care after 34 weeks of gestation, we considered the healthcare professional who was in lead until 34 weeks of gestation to be the one in lead during antenatal care.

Missing data for explanatory variables were imputed using stochastic regression imputation with predictive mean matching as the imputation model [24]. Characteristics of the observed cohort were described as mean ± standard deviation (SD) for continuous variables. Categorical variables were expressed as an absolute value with a percentage. We compared the distribution of characteristics in order to evaluate the relatedness of the imputed cohort and the observed cohort.

Table 3.1 Characteristics of the non-responders, the observed cohort and imputed cohort, and women lost to follow-up. Data expressed as mean (standard deviation), or n (%)

Characteristics	Non-response (n=651)	Observed cohort complete cases (n=702)	Missing values n (%)	Imputed cohort (n=739)	Lost to follow-up (n=146)
Maternal characteristics					
Age, years	30.3 (3.9)	30.7 (3.7)	0 (0.0)	30.7 (3.6)	29.9 (3.9)
Tertiary level of education, n (%)	355 (54.5)	454 (64.7)	0 (0.0)	478 (64.7)	68 (46.6)
Socio-economic status score	-0.64 (1.2)*	-0.5 (1.1)	12 (1.6)	-0.5 (1.1)	-0.8 (1.1)
Body mass index, kg/m2	24.2 (4.1)	24.0 (4.2)	0 (0.0)	24.0 (4.2)	24.7 (4.3)
Nulliparous, n (%)	295 (45.3)	360 (51.4)	0 (0.0)	383 (51.8)	65 (44.5)
State anxiety score	-	33.4 (8.7)	2 (0.3)	33.4 (8.7)	34.1 (7.6)
Antepartum anxiety, n (%)	-	173 (24.7)	2 (0.3)	184 (24.9)	32 (21.9)
Neonatal outcomes					
Neonatal composite outcome, n (%)	28 (4.3)	31 (4.4)	0 (0.0)	33 (4.5)	9 (6.2)
Birthweight <2.3 percentile	7 (1.1)	8 (1.1)	0 (0.0)	8 (1.1)	1 (0.7)
Preterm birth <32 weeks of gestation, n (%)	13 (2.0)	4 (0.6)	0 (0.0)	5 (0.7)	2 (1.4)
APGAR score <7, n (%)	10 (1.5)*	10 (1.4)	0 (0.0)	11 (1.5)	4 (2.7)
NICU admission within 28 days of birth, n (%)	16 (2.5)	15 (2.1)	0 (0.0)	16 (2.2)	5 (3.4)
Perinatal death, n (%)	7 (1.1)	2 (0.3)	0 (0.0)	2 (0.3)	1 (0.7)
Maternal outcomes					
Decrease in perceived personal wellbeing, n (%)	-	182 (26.0)	13 (1.78)	186 (26.0)	-
Maternal composite outcome, n (%)	-	83 (11.9)	10 (1.4)	89 (12)	-
Pre-eclampsia, n (%)	12 (1.8)	19 (2.7)	0 (0.0)	21 (2.8)	4 (2.7)
Gestational diabetes, n (%)	10 (1.5)	20 (2.9)	0 (0.0)	22 (3)	8 (5.5)
Postpartum haemorrhage, n (%)	27 (4.1)*	46 (6.6)	0 (0.0)	49 (6.6)	6 (4.1)
Intensive or High Care admission, n (%)	-	6 (0.9)	13 (1.8)	7 (0.9)	0 (0.0)
Healthcare services					-
Antenatal care led by midwife, n (%)	-	451 (64.2)	0 (0.0)	476 (64.4)	91 (62.3)
Antenatal care led by obstetrician, n (%)	-	137 (19.5)	0 (0.0)	145 (19.6)	35 (24.0)
Transfer during antepartum care, n (%)	-	114 (16.2)	0 (0.0)	118 (16.0)	20 (13.7)
Labour led by midwife, n (%)	189 (29.0)	233 (33.2)	0 (0.0)	239 (32.3)	38 (26.0)
Labour led by obstetrician, n (%)	335 (51.5)	369 (52.6)	0 (0.0)	394 (53.3)	81 (55.5)
Transfer during labour, n (%)	127 (19.5)	100 (14.2)	0 (0.0)	106 (14.3)	27 (18.5)
Spontaneous vaginal birth, n (%)	480 (73.7)*	525 (75.0)	0 (0.0)	555 (75.1)	99 (67.8)
Instrumental vaginal birth, n (%)	63 (9.7)*	58 (16.7)	0 (0.0)	63 (8.5)	18 (12.3)
Cesarean section, n (%)	104 (16.0)*	117 (14.3)	0 (0.0)	121 (16.4)	29 (19.9)
Analgesics during labour, n (%)	303 (46.5)*	304 (43.3)	4 (0.5)	326 (44.1)	88 (60.3)

Variables with a minus (-) sign could not be retrieved without the additional surveys.

*These variables had missing data among women who did not complete any of the additional surveys; socio-economic state score (n=13); APGAR score (n=6); postpartum haemorrhage (n=3); mode of giving birth (n=4); analgesics during labour (n=21)

In the Dutch obstetric system, obstetric care is divided in primary, secondary and tertiary care. Autonomous midwives provide care for low-risk pregnant women in primary care independently. Women with high-risk pregnancies receive care by obstetricians in a secondary care (hospital) setting. If women remain low-risk throughout pregnancy, they remain under the supervision of their midwife, including the postpartum period. These women have the option of giving birth at home or in a birthing centre supervised by their midwife, or in a hospital supervised by an obstetrician. Women with a high-risk pregnancy are always supervised by an obstetrician and thus give birth in a hospital. Antenatal, intrapartum or postpartum transfer of care, from midwife to obstetrician, is a result of either an unexpected finding or a complication during pregnancy or labour.

Results

In total 2,614 women were included in the Expect cohort of whom 1,548 (59%) gave informed consent for receiving the additional questionnaires. Twelve participants were excluded because of multiple gestation, which complicates the interpretation of the neonatal composite outcome. After providing informed consent, 885 women eventually participated in the sub cohort by completing the first additional survey. The PCQ was completed by the majority of these women (n=739, 84%), implying a loss to follow-up of 16%.

Table 3.1 displays the characteristics of participants, illustrating the differences between participants lost to follow-up and those who completed the postpartum survey. The differences between these groups were minimal. Women lost to follow-up had a slightly lower socio-economic status, they tended to have a lower level of education, and were more likely to receive analgesics during labour. The postpartum questionnaire is the only instrument with questions regarding a decreased perceived personal wellbeing and admission to a high care or intensive care unit. As a result, differences with respect to these variables between completers and women lost to follow-up cannot be measured.

Overall, total PCQ-scores were high with a mean score of 109.7 out of 125 points for all respondents (SD 12.5). One quarter of all respondents (n= 176) had PCD, with a mean PCQ score of 92.6 points (SD 6.8). As shown in Table 3.2, these women scored lower on all subscales.

Table 3.2 Pregnancy and Childbirth Questionnaire scores. Data expressed as mean (standard deviation)

Scale	All respondents (n=739)	Satisfied respondents (n=563)	Dissatisfied respondents (n= 176)
Total score (25 items)	109.7 (12.5)	115 (8.3)	92.6 (6.8)
Personal treatment during pregnancy (11 items)	49.1 (5.8)	51 (3.9)	41.7 (4.1)
Education information (7 items)	30.3 (4.2)	32 (3.2)	25.3 (3.1)
Personal treatment during labour (7 items)	30.3 (4.5)	32 (3.6)	25.6 (3.9)

In the multivariable logistic regression, several factors were significantly associated with PCD. Results were adjusted for all other factors, as shown in Table 3.3. Statistically significant maternal factors associated with PCD were decreased perceived personal wellbeing (odds ratio: 1.62; 95% CI: 1.09-2.40), and antenatal anxiety (odds ratio: 2.23; 95% CI: 1.50-3.30). Age was borderline significant with younger women tending to be more likely to experience PCD (odds ratio: 0.95; 95% CI: 0.90-1.00).

Regarding factors related to obstetric care services, there was a statistically significant association between PCD and the healthcare professional in charge of antenatal care and during labour. Transfer from primary to secondary care during the antenatal period was

associated with increased PCD. Antenatal transfer of care before 34 weeks of gestation, was significantly associated with PCD when compared to uninterrupted midwife led care (odds ratio: 1.82; 95% CI: 1.10-3.00). Antenatal transfer also increased the odds of PCD (albeit not significantly) when compared to uninterrupted obstetrician led care (odds ratio: 1.62; 95% CI: 0.93-2.83).

If all labour stages were led by an obstetrician (n = 368), the odds ratio for experiencing PCD was 2.33 (95% CI: 1.34-4.08), compared to all labour stages led by a midwife (n = 232). For women who were referred by their midwife to an obstetrician during labour (n = 100), the odds ratio of PCD was 0.80 (95% CI: 0.37-1.65) compared to those who were assisted by their midwife from onset of labour.

We performed sub-analyses for nulliparous and multiparous women. This did not result in material differences except for cesarean section. A cesarean section was significantly correlated with PCD in nulliparous women (odds ratio 2.68; 95% CI: 1.30-5.57), but not in multiparous women (odds ratio 0.61; 95% CI: 0.25-1.47).

Table 3.3 Multiple logistic regression of maternal and healthcare factors related to pregnancy and childbirth discontent

Determinants	Satisfied (n=563) Mean (sd) or n (%)	Dissatisfied (n=176) Mean (sd) or n (%)	Unadjusted odds ratio (95% CI)	Adjusted odds ratio (95% CI)
Patient related factors				
Age (continuous)	30.8 (3.7)	30.4 (3.7)	0.97 (0.93-1.02)	0.95 (0.90-1.00)
Socio-economic status (continuous)	-0.5 (1.1)	-0.5 (1.0)	1.02 (0.87-1.19)	1.08 (0.91-1.29)
Primary or secondary level of education	197 (35.0)	64 (36.4)	1 [Reference]	1 [Reference]
Tertiary level of education	366 (65.0)	112 (63.6)	0.94 (0.66-1.34)	1.05 (0.71-1.55)
Multiparous	271 (48.1)	85 (48.3)	1 [Reference]	1 [Reference]
Nulliparous	292 (51.9)	91 (51.7)	0.99 (0.71-1.40)	0.98 (0.65-1.48)
No neonatal composite outcome	537 (95.4)	112 (63.6)	1 [Reference]	1 [Reference]
Neonatal composite outcome	26 (4.6)	7 (4.0)	0.86 (0.34-1.90)	0.58 (0.22-1.36)
No maternal composite outcome	557 (98.9)	152 (86.4)	1 [Reference]	1 [Reference]
Maternal composite outcome	6 (1.07)	24 (13.6)	1.21 (0.72-1.97)	0.98 (0.56-1.67)
No decreased perceived personal wellbeing	435 (77.3)	118 (67.0)	1 [Reference]	1 [Reference]
Decreased perceived personal wellbeing	128 (22.7)	58 (33.0)	1.70 (1.17-2.45)*	1.62 (1.09-2.40)*
No antenatal anxiety	444 (78.9)	139 (79.0)	1 [Reference]	1 [Reference]
Antenatal anxiety	119 (21.0)	37 (21.0)	2.18 (1.51-3.15)*	2.23 (1.50-3.30)*
Healthcare related factors				
Healthcare professional care during antenatal care				
Antenatal care led by midwife	387 (68.7)	89 (50.6)	1 [Reference]	1 [Reference]
Antenatal care led by obstetrician	103 (18.3)	42 (23.9)	1.77 (1.15-2.71)	1.12 (0.67-1.85)
Transfer during antenatal care	73 (13.0)	45 (25.6)	2.68 (1.73-4.14)*	1.82 (1.10-3.00)*
Healthcare professional during labour				
Labour led by midwife	202 (35.9)	37 (21.0)	1 [Reference]	1 [Reference]
Labour led by obstetrician	269 (47.8)	125 (71.0)	2.54 (1.70-3.86)*	2.33 (1.34-4.08)*
Transfer during labour	92 (16.3)	14 (8.0)	0.83 (0.42-1.58)	0.80 (0.37-1.65)
No analgesics used during labour	324 (57.5)	89 (50.6)	1 [Reference]	1 [Reference]
Analgesics used during labour	239 (42.5)	87 (49.4)	1.33 (0.94-1.86)	0.71 (0.43-1.16)
Mode of giving birth				
Spontaneous vaginal labour	436 (77.4)	119 (67.6)	1 [Reference]	1 [Reference]
Instrumental vaginal labour	47 (8.3)	16 (9.1)	1.25 (0.66-2.23)	1.19 (0.60-2.30)
Cesarean section	80 (14.2)	41 (23.3)	1.88 (1.22-2.87)*	1.53 (0.88-2.63)

sd, standard deviation

Discussion

In general, women were highly satisfied with the obstetric care received during their pregnancy and childbirth period. Women who experienced PCD scored lower on all subscales, indicating that PCD cannot be attributed to one of the PCQ subscales.
Factors statistically significantly and independently related with PCD were antenatal anxiety, decreased perceived personal wellbeing, and labour led by an obstetrician. Antenatal transfer of care significantly increased the odds upon PCD compared to antenatal care led by a midwife, and tends to increase the odds upon PCD compared to antenatal care led by an obstetrician.
The main strengths of our study are the multicentre prospective cohort design, the large sample size, and the completeness of data. Using a multicentre prospective design improves the probability of collecting a representative sample. Furthermore, it enables optimal measurement of outcomes by minimizing recall bias and recording of all independent variables before completion of the PCQ. Additionally, the PCQ, used to assess satisfaction, has been validated among Dutch women and takes the unique features of the Dutch obstetric care system into account [18,25].
A limitation of this study is that our sub cohort may suffer from some selection bias due to non-response rates, particularly since participants were included from a larger cohort [26]. However, differences between the sub cohort and main cohort were minimal. Moreover, the differences between women who agreed to receive additional surveys but never responded them and those who did were minimal as well, as shown in Table 3.1. For women who started with the first additional survey eventually only 16% did not complete the postpartum questionnaire. For women who did complete the postpartum questionnaire we had 98% completeness of data. By imputing independent variables containing missing data, we limited the possibility of biased results and a loss of statistical precision [27].
To obtain sufficient numbers of women with PCD in our analysis, we focused on women who experienced less than perfect obstetric care, using a total PCQ score of less than 100 points as a cut-off. Our study does not have qualitative data regarding the level of satisfaction or dissatisfaction related to the obstetric care services. However, the amount of studies using the PCQ questionnaire is limited and none of these use dissatisfaction as outcome [18,25].
In line with previous reports, our results indicate that most post-partum women are highly satisfied with obstetric care [6,25,28-30]. We found no association between PCD and maternal demographic factors including, socio-economic status, educational level, and parity. These results correspond with the findings by previous reports [14]. We found a borderline association between PCD and maternal age, whereby younger women tend to be more likely to experience PCD. Results of previous studies are inconsistent regarding the influence of maternal age. Some studies report younger women tend to reflect on their childbirth experience more negatively, whereas a recent study, focussing on discontent as well, does not report any effect of age [5,30,31]. Additionally, since age is a non-modifiable factor, its relevance in the reduction of PCD is limited; still it could serve as a risk indicator increasing awareness among healthcare professionals.
Interestingly, the neonatal and maternal composite outcomes, measures of the occurrence of complications, were not correlated with PCD, but there was a significant association between decreased perceived personal wellbeing and PCD. This suggests that it is not the presence or absence of complications, but rather perceived wellbeing that affects the

experience of pregnancy and birth care. It has been reported that the interaction between a woman and her healthcare professional has a greater influence upon women's perceptions of birth than the physical experience of the birth itself [32].

Our analysis discovered antenatal anxiety is highly correlated to PCD. Nearly a quarter of the women met the criteria of clinically relevant anxiety. Taken together, this makes antenatal anxiety an important factor of interest in order to reduce the number of women who experience PCD.

Referral during antenatal care, which results in transfer from primary care to secondary care, was associated with increased odds of PCD. Although several studies discuss the effects of transfer during labour, studies reporting antenatal transfer are limited. This could be due to the unique Dutch setting, which divides obstetric care between primary and secondary care. Women generally go to a midwife for their first antenatal visit, and in case of a healthy woman with an uncomplicated pregnancy, they receive midwife-led care throughout pregnancy, labour and the postpartum period. Due to the nature of this system, transfer of care is a result of either an unexpected finding or a complication during pregnancy or childbirth. This may increase anxiety. In our analysis we adjusted for clinically relevant anxiety, however the increase of anxiety may be more subtle. Another possibility explaining the increased odds of PCD, may be the result of feelings of loss of control [2,32].

We found no association of PCD with either mode of birth or primary (midwife-led) or secondary (obstetrician-led) antenatal care. However, we did find a correlation between the healthcare professional in charge during labour and PCD. Women assisted by a midwife throughout all stages of labour were significantly less likely to experience PCD when compared to women assisted by an obstetrician. This accords with previous literature showing that women receiving continuity of midwifery care are more likely to be satisfied [33]. In contrast with the findings of previous reports [2,29,32,34], transfer during labour was not associated with PCD. The odds of PCD did not differ significantly between women who were transferred during labour and women who continued to receive midwife-led care (adjusted odds ratio 0.80; 95% CI: 0.37-1.65). Furthermore, the odds of PCD was significantly lower for women who were transferred during labour compared to women who received obstetrician-led care during the entire birthing process (adjusted odds ratio 0.34; 95%CI: 0.17-0.66).

Unfortunately, our data do not permit a reliable analysis regarding the reasons for transfer during labour. A woman may be referred for an emergency with varying degrees of urgency (and experienced associated stress) or a woman may be referred as a result of her request of analgesics. In case of a medical emergency, it is reasonable to believe that a woman will feel a loss of control, which has been strongly associated with a traumatic childbirth experience [32]. Because we do not have information on the reasons for transfer of care, we are not able to analyse this with our data, but it is interesting that women who are transferred have lower levels of PCD. This may suggest that, overall, the Dutch system of primary and secondary care works well with respect to women's birthing experience in relation to transfer during labour.

Our sub analysis in nulliparous and multiparous women did not yield any material differences except for cesarean section and level of antenatal care. These two factors increase the odds of PCD only in nulliparous women. A possible explanation for this discrepancy could be a difference in expectations between nulliparous and multiparous women. Unmet expectations have been linked to influence women's satisfaction with pregnancy and childbirth [13]. A substantial proportion of multiparous women may have received obstetrician-led antenatal

care, or a cesarean in any of their previous pregnancies. As a result, their expectations regarding the course of their current pregnancy may have altered.

Implications

At the moment, the Dutch obstetric system is changing, with a movement towards more integrated care [35,36]. The Ministry of Health published a report promoting patient-centred care combined with integrated care and shared decision making as key concepts of the future obstetric care system [37]. As a result, there is increased interest in the use of individual risk-management systems and decision support aids [17,38]. Depending on how it is organized, integrated care has the potential to increase positive collaboration between midwives in a primary care setting and obstetricians in a secondary care setting. Those who design models of integrated care should take note of the positive birthing experiences associated with midwives and find ways to insure that features of midwife-led care are not lost in the transition [39,40].

Antenatal anxiety was the most important factor related to a negative childbirth experience. It is already known that maternal anxiety is related to adverse outcomes, but this study shows that it is an independent factor for the way women experience their childbirth [9].

Current guidelines on anxiety in pregnancy are mostly focused on anxiety or mood disorders and the effects of medication [41,42]. However, they offer little help for women or healthcare professionals who are dealing with the less severe cases. Our study found that almost 25% of women had an anxiety score that was clinically relevant. Post-partum interventions in women with poor mental health have shown to be cost effective [43]. Our work underscores the need for further research on the effects of screening and treatment for anxiety in pregnancy. Similar to somatic diseases like diabetes and hypertension, pregnancy might be a stress test for women's mental health and early identification and treatment is likely to result in an improved birthing experience [44,45]. Decision support aids are reported to reduce anxiety scores and may be effective tools to imply in order to reduce PCD [46].

Conclusions

Decreased perceived personal wellbeing, increased anxiety, transfer of care antenatal, and obstetrician-led birth, were all independently associated with PCD. One in four women experienced general antenatal anxiety. Women's birthing experience may improve by increased awareness regarding women's antenatal anxiety state and reducing the proportion of women unnecessarily receiving obstetric care in a secondary care setting.

References

1. Goodman, P., Mackey, M.C., et al., Factors related to childbirth satisfaction. Journal of advanced nursing, 2004. 46(2): p. 212-219.
2. Rijnders, M., Baston, H., 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): p. 107-116.
3. Grekin, R. and O'Hara, M.W., Prevalence and risk factors of postpartum posttraumatic stress disorder: a meta-analysis. Clinical Psychology Review, 2014. 34(5): p. 389-401.
4. Larkin, P., Begley, C.M., et al., Women's experiences of labour and birth: an evolutionary concept analysis. Midwifery, 2009. 25(2): p. e49-e59.
5. Baas, C.I., Wiegers, T.A., et al., Client-Related Factors Associated with a "Less than Good" Experience of Midwifery Care during Childbirth in the Netherlands. Birth, 2017. 44(1): p. 58-67.
6. Teijlingen, E.R., Hundley, V., et al., Maternity satisfaction studies and their limitations:"What is, must still be best". Birth, 2003. 30(2): p. 75-82.
7. Lee, A.V., Moriarty, J.P., et al., What can we learn from patient dissatisfaction? An analysis of dissatisfying events at an academic medical center. J Hosp Med, 2010. 5(9): p. 514-20.
8. Redshaw, M., Women as consumers of maternity care: measuring "satisfaction" or "dissatisfaction"? . Birth, 2008. 35(1): p. 73-76.
9. Grigoriadis, S., Graves, L., et al., Maternal Anxiety During Pregnancy and the Association With Adverse Perinatal Outcomes: Systematic Review and Meta-Analysis. J Clin Psychiatry, 2018. 79(5).
10. Munstedt, K., von Georgi, R., et al., Wishes and expectations of pregnant women and their partners concerning delivery. J Perinat Med, 2000. 28(6): p. 482-90.
11. Tschudin, S., Alder, J., et al., Previous birth experience and birth anxiety: predictors of caesarean section on demand? J Psychosom Obstet Gynaecol, 2009. 30(3): p. 175-80.
12. Spice, K., Jones, S.L., et al., Prenatal fear of childbirth and anxiety sensitivity. J Psychosom Obstet Gynaecol, 2009. 30(3): p. 168-74.
13. Christiaens, W. and Bracke, P., Assessment of social psychological determinants of satisfaction with childbirth in a cross-national perspective. BMC Pregnancy and childbirth, 2007. 7(1): p. 26.
14. Britton, J.R., The assessment of satisfaction with care in the perinatal period. J Psychosom Obstet Gynaecol, 2012. 33(2): p. 37-44.
15. Verhoeven, C.J., Nuij, C., et al., Predictors for failure of vacuum-assisted vaginal delivery: a case-control study. Eur J Obstet Gynecol Reprod Biol, 2016. 200: p. 29-34.
16. Schoen, C. and Navathe, R., Failed induction of labor. Semin Perinatol, 2015. 39(6): p. 483-7.
17. Meertens, L.J.E., Scheepers, H.C., et al., External Validation Study of First Trimester Obstetric Prediction Models (Expect Study I): Research Protocol and Population Characteristics. JMIR Res Protoc, 2017. 6(10): p. e203.
18. Truijens, S.E., Pommer, A.M., et al., 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: p. 35-40.
19. Ploeg, H.M., Defares, P.B., et al., Handleiding bij de Zelf-Beoordelings Vragenlijst ZBV: een nederlandstalige bewerking van de Spielberger State- Trait Anxiety Inventory STAI-DY. 1980: Swets & Zeitlinger.
20. Julian, L.J., Measures of Anxiety. Arthritis care & research, 2011. 63(0 11): p. 10.1002/acr.20561.
21. SCP. SCP Statusscores. 2017 [cited 2017 24-08]; Available from: http://www.scp.nl/Formulieren/Statusscores_opvragen.
22. EuroQol-Group, EuroQol-a new facility for the measurement of health-related quality of life. Health Policy, 1990. 16(3): p. 199-208.
23. Visser, G.H., Eilers, P.H., et al., New Dutch reference curves for birthweight by gestational age. Early Hum Dev, 2009. 85(12): p. 737-44.
24. Van Buuren, S., Flexible imputation of missing data. 2012: CRC press.
25. Truijens, S.E., Banga, F.R., et al., The Effect of Multiprofessional Simulation-Based Obstetric Team

Training on Patient-Reported Quality of Care: A Pilot Study. Simul Healthc, 2015. 10(4): p. 210-6.
26. Berman, D.M., Tan, L.L., et al., Surveys and Response Rates. Pediatr Rev, 2015. 36(8): p. 364-6.
27. Sterne, J.A.C., White, I.R., et al., Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. The BMJ, 2009. 338: p. b2393.
28. Karlström, A., Nystedt, A., et al., The meaning of a very positive birth experience: focus groups discussions with women. BMC Pregnancy and Childbirth, 2015. 15: p. 251.
29. Hitzert, M., Hermus, M.A., et al., Experiences of women who planned birth in a birth centre compared to alternative planned places of birth. Results of the Dutch Birth Centre Study. Midwifery, 2016. 40: p. 70-8.
30. Jenkinson, C., Coulter, A., et al., Patients' experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care. Quality and safety in health care, 2002. 11(4): p. 335-339.
31. Borjesson, B., Paperin, C., et al., Maternal support during the first year of infancy. J Adv Nurs, 2004. 45(6): p. 588-94.
32. Hollander, M.H., van Hastenberg, E., et al., Preventing traumatic childbirth experiences: 2192 women's perceptions and views. Archives of Women's Mental Health, 2017. 20(4): p. 515-523.
33. Sandall, J., Soltani, H., et al., Midwife-led continuity models versus other models of care for childbearing women. Cochrane Database Syst Rev, 2015(9): p. Cd004667.
34. Geerts, C.C., van Dillen, J., et al., Satisfaction with caregivers during labour among low risk women in the Netherlands: the association with planned place of birth and transfer of care during labour. BMC Pregnancy and Childbirth, 2017. 17: p. 229.
35. Visser, G.H.A., Obstetric care in the Netherlands: relic or example? J Obstet Gynaecol Can, 2012. 34(10): p. 971-975.
36. Boesveld, I.C., Valentijn, P.P., et al., An Approach to measuring Integrated Care within a Maternity Care System: Experiences from the Maternity Care Network Study and the Dutch Birth Centre Study. International Journal of Integrated Care, 2017. 17(2): p. 6.
37. Geboortezorg, E.Z.I., Zorgstandaard Integrale Geboortezorg, 2016, College Perinatale Zorg: Utrecht.
38. Schoorel, E.N., Vankan, E., et al., Involving women in personalised decision-making on mode of delivery after caesarean section: the development and pilot testing of a patient decision aid. Bjog, 2014. 121(2): p. 202-9.
39. Perdok, H., Jans, S., et al., Opinions of maternity care professionals and other stakeholders about integration of maternity care: a qualitative study in the Netherlands. BMC Pregnancy and Childbirth, 2016. 16: p. 188.
40. Hodnett, E.D., Continuity of Caregivers for Care During Pregnancy and Childbirth. Birth, 2000. 27(3): p. 218-218.
41. Yonkers, K.A., Wisner, K.L., et al., The management of depression during pregnancy: a report from the American Psychiatric Association and the American College of Obstetricians and Gynecologists. Obstet Gynecol, 2009. 114(3): p. 703-13.
42. National Collaborating Centre for Mental, H., National Institute for Health and Clinical Excellence: Guidance, in Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance: Updated edition. 2014, British Psychological Society: Leicester (UK).
43. Chojenta, C., William, J., et al., The impact of a history of poor mental health on health care costs in the perinatal period. Arch Womens Ment Health, 2018.
44. Hermes, W., Tamsma, J.T., et al., Cardiovascular risk estimation in women with a history of hypertensive pregnancy disorders at term: a longitudinal follow-up study. BMC Pregnancy Childbirth, 2013. 13: p. 126.
45. Kramer, C.K., Swaminathan, B., et al., Each degree of glucose intolerance in pregnancy predicts distinct trajectories of beta-cell function, insulin sensitivity, and glycemia in the first 3 years postpartum. Diabetes Care, 2014. 37(12): p. 3262-9.

46. Tucker Edmonds, B., Shared decision-making and decision support: their role in obstetrics and gynecology. Curr Opin Obstet Gynecol, 2014. 26(6): p. 523-30.

Chapter 4

Implementation and effects of risk-dependent obstetric care in the Netherlands (Expect Study II): Protocol for an impact study

Pim van Montfort, Jessica P.M.M. Willemse, Carmen D. Dirksen, Ivo M.A. van Dooren, Linda J.E. Meertens, Marc E.A. Spaanderman, Maartje Zelis, Iris M. Zwaan, Hubertina C.J. Scheepers, Luc J.M. Smits

JMIR Research Protocols, 2018

Abstract

Background

Recently, validated risk models predicting adverse obstetric outcomes combined with risk-dependent care paths have been made available for early antenatal care in the southeastern part of the Netherlands. This study will evaluate implementation progress and impact of the new approach in obstetric care.

Objective

The objective of this paper is to describe the design of a study evaluating the impact of implementing risk-dependent care. Validated first-trimester prediction models are embedded in daily clinical practice and combined with risk-dependent obstetric care paths.

Methods

A multicentre prospective cohort study consisting of women who receive risk-dependent care is being performed from April 2017 to April 2018 (Expect Study II). Obstetric risk profiles will be calculated using a Web-based tool, the Expect prediction tool. The primary outcomes are the adherence of healthcare professionals and compliance of women. Secondary outcomes are patient satisfaction and cost-effectiveness. Outcome measures will be established using Web-based questionnaires. The secondary outcomes of the risk-dependent care cohort (Expect II) will be compared with the outcomes of a similar prospective cohort (Expect I). Women of this similar cohort received former care-as-usual and were prospectively included between July 1, 2013 and December 31, 2015 (Expect I).

Results

Currently women are being recruited for the Expect Study II and a total of 300 women are enrolled.

Conclusions

This study will provide information about the implementation and impact of a new approach in obstetric care using prediction models and risk-dependent obstetric care paths.

Trial Registration

Netherlands Trial Registry (NTR): NTR4143

Introduction

Perinatal mortality plays a pivotal role in the quality assessment of perinatal care [1]. In developed countries the main causes of perinatal mortality are small-for-gestational-age infancy (SGA), preterm birth (PTB), and asphyxia [2,3]. Pre-eclampsia (PE) is an important cause for both SGA and induced PTB [4]. Risks of asphyxia and birth injuries are increased among infants that are large-for-gestational-age (LGA) [5], which in turn is strongly associated with gestational diabetes mellitus (GDM) [6]. Thus, PE, GDM, PTB, SGA, and LGA are all directly or indirectly related to perinatal mortality.

A number of interventions have shown to be effective in the prevention of adverse pregnancy outcomes, such as low-dose aspirin treatment in case of PE [7-9], adequate management of GDM [10,11], and progesterone administration in women at risk of spontaneous PTB [12]. Besides calcium supplementation, most of these interventions are not suitable for all pregnant women, because of either possible adverse effects, patient burden, or costs. Early prediction of obstetric risks may therefore help healthcare professionals in designing intervention strategies based on women's individual risks.

Recently, we performed an external validation study of first trimester prediction models predicting the risk of PE, GDM, PTB, SGA and LGA (the Expect Study I) [13,14]. The Expect Study I identified clinically useful prediction models for PE and GDM. The Limburg Obstetric Consortium (LOC), midwives and gynecologists of the southeastern part of the Netherlands developed care pathways, i.e., basic antenatal care for women at low risk and additional risk-dependent care for women with elevated risks of PE, GDM, PTB, SGA, or LGA. The LOC agreed to implement the risk models predicting PE and GDM, in order to identify women at increased risk of these outcomes, and to offer these women risk-dependent care.

The current protocol describes the design of a multicentre prospective cohort study (Expect Study II) evaluating the implementation progress of using these prediction models combined with tailored care paths for PE and GDM.

The primary aims of the Expect Study II are to measure adherence to the new risk-dependent care guidelines by healthcare professionals and compliance of pregnant women who received recommendations. The secondary aims are to evaluate its impact upon patient satisfaction and cost-effectiveness. Secondary aims will be studied by comparing these outcomes of the Expect II cohort with the Expect I cohort.

Methods

Study Design and Recruitment

In April 2017, the Expect prediction tool was introduced. The Expect prediction tool was developed to enable individual risk assessment during early pregnancy regarding the risks of PE, GDM, PTB, SGA, and LGA. Validated models selected by the LOC to predict PE and GDM have been incorporated into this tool (study submitted by Meertens et al). Risk assessment of spontaneous PTB, SGA and LGA is achieved using the revised LOC guidelines [15]. For nulliparous women, the prediction tool comprises 14 variables concerning anthropometric data, relevant medical history, and family history. For multiparous women the tool enquires six more variables, all concerning the women's obstetric history.

The Expect prediction tool is a Web-based form, which calculates the estimated risk profiles. This tool was made available for healthcare professionals via the Expect study website

(https://www.zwangerinlimburg.nl) for implementation in daily obstetric care. Besides the estimated risks of adverse pregnancy outcomes, the tool provides recommendations for tailored antenatal care based on personalized risks (i.e., risk-dependent care). In addition, patient information brochures relevant to the patient's risk profile will be automatically generated. The health care professionals can use this tool during one of the pregnant woman's antenatal visits before 16 weeks of pregnancy. Using a shared decision approach, the appropriate risk-dependent care path with corresponding preventive measures and check-ups will be selected.

In order to implement risk-dependent care successfully, midwives and gynecologists are encouraged to use the Expect prediction tool by representatives of the LOC. The Expect prediction tool is introduced by email to all obstetric healthcare professionals in the region. Furthermore, oral presentations will be given at every hospital and at local midwifery meetings. Additionally, the hospitals and midwifery practices are contacted regularly by phone and in person to evaluate the Expect prediction tool.

The midwives and gynecologists play a central role in enrolling pregnant women into the Expect Study II, by asking women whether they are interested in receiving further information about participating in the Expect Study II. Almost every pregnant woman is eligible for our study. The exclusion criteria are (1) maternal age <18 years, (2) documented multiple pregnancy, and (3) ≥16 weeks of gestation at intake. The eligibility criteria are identical to those of the Expect Study I cohort [13]. Eligible women agreeing to participate are asked to give informed consent and to complete 4 Web-based surveys at enrolment, 24 weeks and 34 weeks of gestation, and 6 weeks after due date.

A personal link to the first online survey will be sent immediately after enrolment. If the survey was not accessed or incomplete, two automatic reminders will be sent by email at 3-day intervals for surveys one to three and at 6-day intervals for the postpartum survey. In case of non-response, women will be contacted by phone (provided that a correct phone number is available). If women report PTB at the beginning of survey two or three, they will automatically be redirected to the postpartum survey.

The medical ethical committee of Maastricht University Medical Centre (MUMC+) evaluated the study protocol and declared that the study did not fall within the scope of the Dutch Medical Research Involving Human Subjects Act (WMO) (METC-17-4-057).

Tailored-Care Paths

The LOC consists of midwives (n=9), gynecologists (n=9), professionals in maternity care (n=2), researchers (n=3), and an independent chairperson. They meet four to five times annually and represent the University medical school, midwifery academy, all hospitals, and roughly 80% (n=90) of the midwives of the province. The midwives and gynecologists of the LOC revised the content of obstetric care. We will briefly describe the most important changes regarding antenatal care compared to former care-as-usual, which has been observed during Expect Study I. All women will receive basic antenatal care. In the new tailored care paths, recommendations about calcium and vitamin D supplementation are emphasized for all women and an additional ultrasound for foetal growth assessment at 32 weeks of pregnancy is introduced as part of basic antenatal care.

An overview of the care pathways is provided in Table 4.1. Additional risk-dependent care for women with a mildly elevated risk of PE comprises the recommendation of preventive

aspirin treatment, 80-100 mg aspirin daily from 12 weeks up to 36 weeks of pregnancy. Obstetric care for women with a substantial risk of PE additionally comprises of extended blood tests, blood pressure measurements every 2 weeks from 14 weeks up to 40 weeks of gestation, and 2 additional ultrasounds for foetal growth measurements.
Women with a history of GDM are advised to have an oral glucose tolerance test (OGTT) at 16 and 26 weeks of pregnancy. Women with a mildly elevated risk are advised to have an OGTT at 24 weeks of pregnancy. Furthermore, in both cases, women will receive two additional ultrasounds for foetal growth measurements in addition to basic antenatal care.

Outcome Measures and Measurement

The primary outcomes are healthcare professionals' adherence to key recommendations and compliance of the women involved in the study. Adherence is defined as the proportion of women that actually received the key recommendations they should have received from their healthcare professional according to the LOC guidelines. Adherence will be analysed regarding recommendations of adequate vitamin D (yes or no) and calcium intake (yes or no) for all women, preventive aspirin treatment (yes or no) for women with elevated PE risks, and OGTT (yes or no) for women with elevated GDM risks.
Compliance is defined as the proportion of women whom comply with the LOC recommendations they have received (yes, no or partially). Compliance will be analysed regarding: adequate vitamin D (10 microgram per day) and calcium (1,000 milligram per day) intake, preventive aspirin treatment, and OGTT.
The secondary outcomes are patient satisfaction and cost-effectiveness. These secondary outcomes of Expect Study II will be compared to the outcomes of Expect Study I.
Patient satisfaction will be measured by validated patient satisfaction questionnaires. The Patient Satisfaction Questionnaire Short Form (PSQ-18) will be incorporated in antepartum surveys two and three. In the postpartum survey, patient satisfaction will be assessed by the Pregnancy and Childbirth Questionnaire (PCQ) [16]. The PCQ is validated for Dutch women who recently gave birth and addresses three topics: women's satisfaction with the healthcare professional during pregnancy, health education, and satisfaction with the healthcare professional during labour. Furthermore, Truijens et al showed the PCQ is sensitive to pick up effects regarding patient satisfaction due to simulation-based obstetric team training [17].
In order to perform cost-effectiveness calculations, we will calculate two incremental cost-effectiveness ratios (ICERs). The first ICER expresses the healthcare costs per one neonatal composite outcome prevented. The neonatal composite outcome is defined as perinatal death within seven days after birth, asphyxia (Apgar score <7 after 5 minutes), admission to a neonatal intensive care unit (NICU) within 28 days after birth, SGA (birthweight <2.3 weight percentile), and very preterm birth (birth before 32 completed weeks of pregnancy) [13]. The second ICER will express the healthcare cost per one maternal gained Quality Adjusted Life Year (QALY).

Data Collection

For the primary outcomes, we will use the data collected for the Expect Study II. For the secondary outcomes, when comparing the effects of risk-dependent care with former care-as-usual, the outcomes of the Expect Study II will be compared with the outcomes of the Expect Study I. For this reason, the survey intervals and the questions regarding the

Table 4.1 Overview of care pathways

Gestational age (weeks)	Basic antenatal care	Additional risk-dependent care				
	All women	Pre-eclampsia (mildly elevated risk)	Pre-eclampsia (high risk)	Gestational diabetes mellitus	Small or large for gestational age infancy	Spontaneous preterm birth
6-10	Intake and risk assessment using the Expect prediction tool and general recommendations (e.g., Calcium and vitamin D intake)					
10-12	Confirmation gestational age (CRL[a] US[b]) and blood tests (e.g., blood type, hemoglobin)	low dose aspirin prophylaxis	low dose aspirin prophylaxis and extended blood tests (e.g., renal function)			Cervical length measurement
14			BP[c] measurement			
16			BP measurement	OGTT[d] (in case of history of GDM[e])		Cervical length measurement and progesterone prophylaxis
18-20	Check-up (e.g., BP and symphysio-fundal height measurements) and US screening for congenital abnormalities					Cervical length measurement
22			BP measurement			
24-26	Check-up		BP measurement	OGTT		Cervical length measurement
27	Additional blood tests (depending on Rhesus genotype)					
28			BP measurement and US foetal biometry	US foetal biometry	US foetal biometry	Cervical length measurement
30	anti-D immunoglobulin prophylaxis (depending on genotype)					

Table 4.1 (continued) Overview of care pathways

Gestational age (weeks)	Basic antenatal care	Additional risk-dependent care		
32	Check-up, blood tests (e.g., hemoglobin), and US foetal biometry			Cervical length measurement
34		BP measurement		
36	Check-up and US foetal position	US foetal biometry	US foetal biometry	
37		BP measurement		
38		BP measurement		
39		BP measurement		
40	Check-up and shared decision regarding induction of labor			
41-42	Check-up			

[a]CRL: Crown rump length; [b]US: ultrasound; [c]BP: blood pressure; [d]OGTT: oral glucose tolerance test according to the International Association of the Diabetes and Pregnancy Study Groups (IADPSG) criteria; [e]GDM: gestational diabetes mellitus

secondary outcomes are similar between the two studies.
In the Expect Study II, data will be collected using the Expect prediction tool, comprising women's personal risk profile, and Web-based patient surveys. A structured overview of patient enrolment and data collection for the Expect Study II is shown in Figure 4.1.
The first survey addresses the following topics: (1) recommendations and information given by health healthcare professionals, (2) women's intention to comply with these recommendations, (3) dietary intake of calcium and vitamin D sunlight exposure, and (4) vitamin and mineral supplement usage.
The second and third surveys are comparable to each other and will address the following topics: (1) patient satisfaction, (2) women's state anxiety, (3) maternal quality of life, (4) changes in vitamin and mineral supplement usage, and (5) healthcare resource use.
In order to document the nature and volume of healthcare resource used, women will be asked to record all visits to midwives, hospitals, and other care institutions. Furthermore, questions related to medication use, hospital admission, diagnostic and medical procedures, and the delivery will be asked. To minimize patient recall problems, information regarding the usage of health care resources will be requested at three intervals (surveys two, three and four) during the study period.
Survey four, the postpartum survey, addresses obstetric outcomes, compliance of healthcare recommendations, and the topics mentioned in survey two and three. Furthermore, this survey also contains questions regarding the healthcare consumption related to the neonate.

Sample size

According to the results of the validation study (Expect I), we expect approximately 30% of women to have an elevated estimated risk of PE, the obstetric complication with the lowest incidence (article submitted by Meertens et al). Furthermore, an adherence of 70% and a compliance of approximately 40% is expected for the recommended aspirin treatment. This will result in approximately 21% and 12% respectively of the general population having an elevated risk of PE. In order to estimate these percentages with a precision of approximately 4% the required sample size is estimated at 400 participants [18].

Statistical Analysis

Missing values will be handled by imputation. Stochastic regression imputation with predictive mean matching as the imputation model will be used to prevent biased results based on complete case analysis only [19].
Adherence will be calculated by the proportion of women who reported to have received the LOC recommendations regarding adequate vitamin D and calcium intake, preventive aspirin treatment, and OGTT. Answers of participants will be linked to their estimated risk profile based on the Expect prediction tool.
Compliance will be analysed by calculating the proportion of women who complied with the recommendations received from their healthcare professional regarding aspirin treatment, OGTT, vitamin D, and calcium intake. Vitamin D is analysed based on supplement intake and sunlight exposure. Calcium intake is determined by calculating the daily intake from diet and supplement use. Dietary intake will be estimated using answers from a selection of questions from the Dutch National Food Frequency Questionnaire tool [20]. These questions address food products that cumulatively cover >80% of the variance in calcium intake

[21]. Total intake of both nutrients will be compared with the recommended intake by the LOC (1000 milligram calcium per day and 10 microgram of vitamin D per day) in order to determine compliance to these recommendations.
The secondary outcomes, patient satisfaction and cost-effectiveness, will be analysed by comparing Expect Study II with the outcomes of former care-as-usual (Expect Study I). Patient satisfaction scores will be analysed using multiple linear regression.
For the economic evaluation, we will use a health care perspective according to the Dutch guidelines for cost calculations [22]. A time horizon of approximately eleven months, from onset of pregnancy up to six weeks post-partum, will be applied. Maternal quality of life will be evaluated using the Euroqol EQ-5D-3L and EQ VAS (Euroqol Visual Analogue Scale) questions, which are incorporated, in the surveys. The EQ-5D-3L and EQ VAS are standardized questionnaires used worldwide to assess quality of life. Maternal QALYs will be calculated using the corresponding utility scores based on the Dutch population [23,24]. All costs will be expressed as 2017 Euros and if necessary cost prices will be transformed to 2017 Euros using the Dutch Consumer Price Index [25]. Bootstrap- and standard sensitivity analyses will be performed to quantify the uncertainty regarding the cost-effectiveness outcomes.

Results

Currently, women are being recruited for the Expect Study II and a total of 300 women are enrolled. We expect to achieve our goal of 400 participants during April 2018 and postpartum data collection will be finished by March 2019. As a result, first study results are expected in 2019.

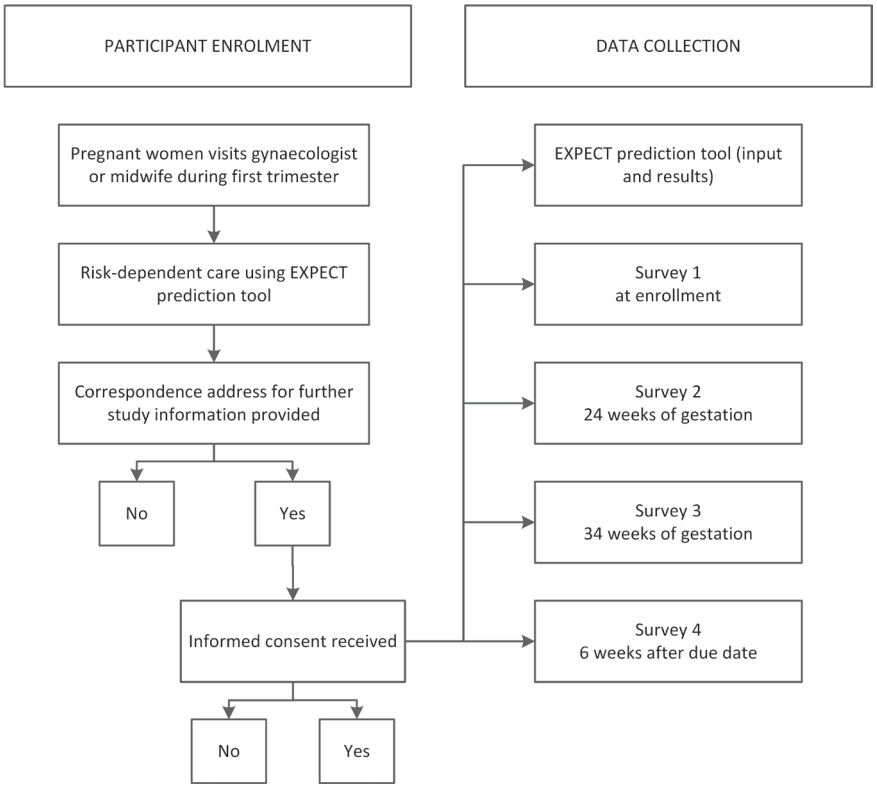


Figure 4.1 Flowchart of participant enrolment and data collection of the Expect Study II. Whether or not a woman participates to the Expect Study II does not affect the health care women receive during their pregnancy

Discussion

This paper describes the protocol of an impact study regarding the implementation of externally validated prediction models combined with risk-based care pathways in obstetric care. Prediction models are becoming increasingly popular in medicine [26]. Although the number of prediction models being published has increased tremendously in recent years, the number of external validation studies remains small [26]. Furthermore, performances of models predicting adverse pregnancy outcomes and the efficacy of preventive interventions for these outcomes are generally documented separately [8,27,28]. Impact studies, describing the effect of using prediction models in daily practice combined with preventive interventions relevant to the estimated risk are nearly non-existent [26]. To the best of our knowledge, no impact studies using prediction models in general obstetric practice have been published.

The strengths of our design are the multicentre prospective data collection and the similarity of both cohorts. Recruitment in multiple centres, hospitals and midwife clinics, improves the probability of collecting a representative sample of the obstetric population. This is essential in the Netherlands, since most pregnant women receive antenatal care by midwives at outpatient clinics [29]. Furthermore, optimal measurement of the outcomes is achieved by prospective data collection [30]. Finally, because the two cohorts are kept as similar as possible, we are able to accurately compare the former care-as-usual with the

new risk-dependent care.

Some limitations of the design must also be noted. First, since the comparison of secondary outcomes of Expect II with those of Expect I is essentially a before-and-after comparison, time trends in the outcomes can theoretically influence results. In the interpretation of the results, we will take such trends into account, e.g., by looking at trends in the studied outcomes from other regions in the Netherlands.

A second possible limitation of our study is that several outcomes will solely be based on participant questionnaires. Potential recall bias, however, is limited due to the prospective design and the usage of four questionnaires at limited intervals. Additionally, questionnaires have been shown to be a valid method of data collection regarding perinatal outcomes and medication exposure during pregnancy [31,32]. In the questionnaires we urge respondents to answer honestly and emphasize that all answers will be treated confidentially and will not influence the care provided by their obstetric health care professional. Furthermore, the additional procedures recommended in the risk-dependent care path are all subject to a shared decision-making process between woman and healthcare professional. As a result, we expect there is currently no taboo regarding the compliance with given recommendations.

We hypothesize that risk-dependent care results in early detection or prevention of obstetric adverse events and can thus reduce prevalence of neonatal adverse events. However, due to low prevalence rates of approximately 5%, large cohorts (approximately two times 6,800 participants) are necessary in order to achieve sufficient power to detect a reduction of at least 20% [18]. Therefore, the influence of risk-dependent care on the incidence of the neonatal composite outcome will be analysed using registry data of the region. Moreover, to achieve the desired effects of risk-dependent care, it first needs to be implemented successfully. Thus, implementation should first lead to behavioural changes for both health care professionals and pregnant women.

References

1. Zeitlin, J., Wildman, K., et al., Selecting an indicator set for monitoring and evaluating perinatal health in Europe: criteria, methods and results from the PERISTAT project. European Journal of Obstetrics & Gynecology and Reproductive Biology, 2003. 111: p. S5-S14.
2. Goldenberg, R.L., Culhane, J.F., et al., Epidemiology and causes of preterm birth. Lancet, 2008. 371(9606): p. 75-84.
3. Ensing, S., Abu-Hanna, A., et al., Trends in birth asphyxia, obstetric interventions and perinatal mortality among term singletons: a nationwide cohort study. The Journal of Maternal-Fetal & Neonatal Medicine, 2015. 28(6): p. 632-637.
4. Koullali, B., Oudijk, M.A., et al., Risk assessment and management to prevent preterm birth. Semin Fetal Neonatal Med, 2016. 21(2): p. 80-8.
5. Horvath, K., Koch, K., et al., Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. Bmj, 2010. 340: p. c1395.
6. Henriksen, T., The macrosomic fetus: a challenge in current obstetrics. Acta Obstet Gynecol Scand, 2008. 87(2): p. 134-45.
7. Meertens, L.J.E., Scheepers, H.C.J., et al., Should women be advised to use calcium supplements during pregnancy? A decision analysis. Matern Child Nutr, 2017.
8. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. American journal of obstetrics and gynecology, 2017. 216(2): p. 110-120. e6.
9. Rolnik, D.L., Wright, D., et al., Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med, 2017.
10. Group, H.S.C.R., Hyperglycemia and adverse pregnancy outcomes. N Engl j Med, 2008. 2008(358): p. 1991-2002.
11. Poolsup, N., Suksomboon, N., et al., Effect of treatment of gestational diabetes mellitus: a systematic review and meta-analysis. PLoS One, 2014. 9(3): p. e92485.
12. Dodd, J.M., Jones, L., et al., Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. Cochrane Database Syst Rev, 2013(7): p. Cd004947.
13. Meertens, E.L.J., Scheepers, C.J.H., et al., External Validation Study of First Trimester Obstetric Prediction Models (Expect Study I): Research Protocol and Population Characteristics. JMIR Res Protoc, 2017. 6(10): p. e203.
14. Meertens, L., van Montfort, P., et al., Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta Obstetricia et Gynecologica Scandinavica, In Print.
15. Lemmens, S.M.P., Spaanderman, M.E.A., et al., Limburg Obstetric Quality System - Zorgpaden (aanbevelingen), 2016, LOQS: Maastricht.
16. Truijens, S.E., Pommer, A.M., et al., 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: p. 35-40.
17. Truijens, S.E., Banga, F.R., et al., The Effect of Multiprofessional Simulation-Based Obstetric Team Training on Patient-Reported Quality of Care: A Pilot Study. Simul Healthc, 2015. 10(4): p. 210-6.
18. Dean, A.G., Sullivan, K.M., et al. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 3.01. 2013/04/06 2018-01-25].
19. van der Heijden, G.J., Donders, A.R.T., et al., Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. Journal of clinical epidemiology, 2006. 59(10): p. 1102-1109.
20. Molag, M.L., de Vries, J.H., et al., Selecting informative food items for compiling food-frequency questionnaires: comparison of procedures. Br J Nutr, 2010. 104(3): p. 446-56.
21. Rossum van, C.T.M., Fransen, H.P., Verkaik-Kloosterman, J., Buurma-Rethans, E.J.M., Ocké, M.C.

Dutch National Food Consumption Survey 2007-2010. 2011.
22. Hakkaart-van Roijen, L., Van der Linden, N., et al., Kostenhandleiding. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. In opdracht van Zorginstituut Nederland. Geactualiseerde versie, 2015.
23. Lamers, L., Stalmeier, P., et al., Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. Nederlands tijdschrift voor geneeskunde, 2005. 149(28): p. 1574-1578.
24. Group, T.E., EuroQol-a new facility for the measurement of health-related quality of life. Health policy, 1990. 16(3): p. 199-208.
25. CBS. Dutch consumer price index. [2018-01-25].
26. Moons, K.G., Kengne, A.P., et al., Risk prediction models: II. External validation, model updating, and impact assessment. Heart, 2012. 98(9): p. 691-8.
27. Kleinrouweler, C.E., Cheong-See, F.M., et al., Prognostic models in obstetrics: available, but far from applicable. American journal of obstetrics and gynecology, 2016. 214(1): p. 79-90. e36.
28. Falavigna, M., Schmidt, M.I., et al., Effectiveness of gestational diabetes treatment: a systematic review with quality of evidence assessment. Diabetes research and clinical practice, 2012. 98(3): p. 396-405.
29. Visser, G.H.A., Obstetric care in the Netherlands: relic or example? J Obstet Gynaecol Can, 2012. 34(10): p. 971-975.
30. Moons, K.G., Royston, P., et al., Prognosis and prognostic research: what, why, and how? Bmj, 2009. 338: p. b375.
31. van Gelder, M., Vorstenbosch, S., et al., Web-based questionnaires to assess perinatal outcome proved to be valid. J Clin Epidemiol, 2017. 90: p. 136-143.
32. Sundermann, A.C., Validation of maternal recall of early pregnancy medication exposure using prospective diary data. 2017. 27(2): p. 135-139.e2.

Part II

Implementation and impact of personalized obstetric care

Chapter 5

Implementing a pre-eclampsia prediction model in obstetrics: cut-off determination and healthcare professionals' adherence

Pim van Montfort, Luc J.M. Smits, Ivo M.A. van Dooren,
Stéphanie M.P. Lemmens, Maartje Zelis, Iris M. Zwaan,
Marc E.A. Spaanderman, Hubertina C.J. Scheepers

Medical Decision Making, 2019

Abstract

Background

Despite improved management, pre-eclampsia remains an important cause of maternal and neonatal mortality and morbidity. Low-dose aspirin (LDA) lowers the risk of pre-eclampsia. Although several guidelines recommend LDA prophylaxis in women at increased risk, they disagree about the definition of high-risk. Recently, an externally validated prediction model for pre-eclampsia was implemented in a Dutch region combined with risk-based obstetric care paths.

Objectives

To demonstrate the selection of a risk threshold and to evaluate the adherence of obstetric healthcare professionals to the prediction tool.

Study Design

Using a survey (n=136) and structured meetings among healthcare professionals, possible cut-off values at which LDA should be discussed were proposed. The prediction model, with chosen cut-off and corresponding risk-based care paths, was embedded in an online tool. Subsequently, a prospective multicenter cohort study (n=850) was performed to analyze the adherence of healthcare professionals. Patient questionnaires, linked to the individual risk profiles calculated by the online tool, were used to evaluate adherence.

Results: Healthcare professionals agreed upon employing a tool with a high detection rate (cut-off: 3.0%; sensitivity 75%, specificity 64%) followed by shared decision between patient and healthcare professional on LDA prophylaxis. Of the 850 enrolled women, 364 women had an increased risk of pre-eclampsia. LDA was discussed with 273 of these women, resulting in an 81% adherence rate.

Conclusion

Consensus regarding a suitable risk cut-off threshold was reached. The adherence to this recommendation was 81%, indicating adequate implementation.

Introduction

Pre-eclampsia (PE) is an important cause of mortality and morbidity for both the mother and the fetus. Although management of PE has improved, a cure that would preserve the pregnancy remains unavailable. Therefore, preventive measures play a pivotal role in decreasing the burden of the disease [1].

In addition to adequate calcium intake, diet, and lifestyle interventions, aspirin treatment receives an increasing amount of attention as a preventive measure [1,2]. Low-dose aspirin (LDA) prophylaxis, in a dosage of 80-150mg daily, has been proven to reduce the risk of pre-eclampsia [3]. Therefore, several professional authorities such as the American College of Obstetricians and Gynecologists (ACOG), the US Preventive Services Task Force, and the National Institute for Health and Clinical Excellence (NICE) recommend LDA prophylaxis in women at increased risk of PE [4-6].

These authorities all recommend LDA to women at increased risk by using a list of separate risk factors (e.g. a history of PE, or chronic hypertension). They however differ in their selection of risk factors and thus their definition of women at increased risk. Universal recommendation for all pregnant women has been proposed as well since LDA is inexpensive, widely available, and appears to be safe in pregnancy beyond the first trimester [7]. However, this view is controversial due to a lack of understanding in the preventive mechanism of LDA, and a lack of proven benefits for women at low risk [7].

Multivariable prediction models estimating the risk of PE weigh several risk factors simultaneously and can assist healthcare professionals in identifying women with increased risk. The results of a recent study comparing several PE prediction models simultaneously in one cohort [8] indicated that some of these models are more efficient compared to a list of single risk factors. For a prediction model to serve as a decision tool, a cut-off has to be determined for the discrimination of low and increased risk.

Recently, the recommendation of LDA prophylaxis was adopted in the regional guidelines in the Southeastern part of The Netherlands [9]. Women with an elevated PE risk are identified using a prediction model. However, dissemination of guidelines or stating recommendations does not automatically result in adherence by healthcare professionals. Implementation of effective preventive interventions often suffers from low adherence rates [10-12]. Despite the increased attention of the role of LDA in the prevention of pre-eclampsia, a recent conference report showed that up to 42% of women considered as high risk according to the NICE guidelines had not been offered LDA [13].

This paper reports on 1) the selection of a cut-off value by healthcare professionals for the identification of women at risk of PE using a prediction model, and 2) results of healthcare professional's adherence to LDA recommendations in the local guidelines.

Methods

Definition of women at risk of pre-eclampsia

The Limburg Obstetric Consortium (LOC) is a committee representing all obstetric health care professionals in the Southeastern part of the Netherlands, which consists of five regions. Every region consists of a hospital providing secondary obstetric care (gynecologists and clinical midwives) and outpatient midwifery practices (autonomous midwives providing primary obstetric care)). Each region provides two to four obstetric healthcare professionals

as LOC representatives. In total, the LOC consists of independent midwives (n= 11), gynecologists (n= 10), maternity care nurses (n=2), and researchers (n=3).
The LOC developed risk-based care pathways that were implemented in 2017. These pathways consist of basic antenatal care for the low-risk group and additional recommendations for women at risk for several pregnancy related complications including PE. The methods of formulating these pathways and their content are reported elsewhere [14,15].
For women at risk of PE additional risk-based care includes the recommendation of LDA prophylaxis (80-100mg) from 12 up to 36 weeks of pregnancy. The LOC agreed to use the prediction model of Syngelaki 2011, externally validated and recalibrated for their specific region by Meertens et al. [8,16]. This model was selected because it was the model with the highest discriminative performance and its predictors are routinely collected in Dutch obstetric practice. Predictors included in the prediction model are age, body mass index, ethnicity, parity, assisted conception treatment, smoking during pregnancy, family history of PE, and medical history (regarding chronic hypertension, PE, and diabetes mellitus). The algorithm of the calibrated model, along with its discriminative performance, is provided in supplementary file 1.
Consensus regarding the PE risk-threshold, the cut-off value at which healthcare professionals discuss the recommendation of LDA, was reached using a three-step procedure. First, all obstetric healthcare professionals of the LOC region received a survey with statements regarding the implementation of a PE prediction model and possible risk-thresholds. Second, using the results of the survey, the preferences of healthcare professionals were discussed in regional meetings with the midwives and gynecologists of the region. Third, the results of both the survey and the regional meetings were discussed with the LOC committee. During a final meeting the decision was made whether the prediction model should be adopted and which risk-threshold was preferred.
In the survey, three possible risk-thresholds were suggested; 1) a threshold with a high sensitivity and low specificity similar to the specificity of the ACOG guideline [4] (risk-threshold 2.85%, sensitivity 79%, specificity 60%); 2) a threshold resulting in a relative risk of 2.0 upon PE for positive results (risk-threshold 3.90%, sensitivity 57%, specificity 80%); and 3) a threshold with a low sensitivity and high specificity similar to the specificity of the NICE guideline [6] (risk-threshold 5.20%, sensitivity 30%, specificity 90%). Each suggested threshold was provided with additional information: sensitivity, specificity, as well as total number of test positives, test negatives, true positives, false positives, true negatives, false negatives, and numbers needed to treat. Data of the external validation study were used to calculate these test characteristics per risk-threshold [8].
Healthcare professionals were asked to answer the statement 'I agree using this cut-off value as threshold determining an elevated PE risk', using a ten-point Likert scale (1 totally disagree – 10 fully agree).
The PE prediction model with corresponding threshold was embedded in the Expect prediction tool, which is available online for healthcare professionals. The LOC strongly encourages midwives and gynecologists to use the Expect prediction tool during the first antenatal visits. This was achieved by oral presentations, e-mails, regular phone calls, and in person evaluations [14].

Data collection of pregnant women

When consensus regarding the threshold was reached, the Expect prediction tool was

implemented. Participants, pregnant women, were enrolled in a multicenter prospective cohort study in the Southeastern part of The Netherlands from April 2017 to August 2018 (Expect Study II). A more detailed description of the study design has been published elsewhere [14]. Briefly, women were recruited at their first prenatal visit (<16 weeks of pregnancy) if their healthcare professional used the Expect prediction tool. In Dutch obstetric care, pregnant women visit either an autonomous midwife (outpatient clinic) or a gynecologist (hospital), both midwives and gynecologists recruited women for the Expect Study.

Women of at least 18 years with a singleton pregnancy were eligible for inclusion. Questionnaires and study information were provided in Dutch only. Eligible women were asked whether they agreed to provide their e-mail address in order to receive information regarding the Expect Study. When women agreed to participate and completed an online informed consent form, they received a personal link by e-mail to the web-based surveys. The first survey, collecting the data used for the analyses in this study, was disseminated at enrolment. Two automatic reminders were sent using 3-day intervals. Women were contacted by phone if no response was received. The survey embedded questions regarding the healthcare services women received from their midwife or gynecologist during the first visits. Women were specifically questioned whether their PE risk was discussed with them (yes, I have an increased risk/ yes, I have an average risk/ yes, I have a low risk/ no, it was not discussed/ I do not recall whether this was discussed). Furthermore, women were asked whether the option of LDA was discussed with them (yes/ no/ I do not recall).

Statistical analysis

We cross-tabulated the proportions of women whom reported to have discussed their PE risk and the option of LDA with respect to the predicted PE risk (low risk / increased risk). Furthermore, we plotted these proportions with respect to the predicted PE risk by categorizing PE risk predictions (≤1.0% to >6.0% using a binwidth of 0.5 percentage points). To analyze possible differences in healthcare professionals' adherence rates to the risk-based recommendations, we plotted LDA discussion rates reported by women using the study duration as a continuous variable. A nonparametric local weighted regression (loess regression) was applied to fit the curves. We analyzed the correlation between the discussion rates and the predicted PE risk for women with a risk exceeding 3.0% by use of logistic regression with predicted PE risk as an independent variable (continuous, percentage).

All statistical analyses were performed using R statistical software version 3.6.0 along with the packages "foreign", "dplyr", "tidyr",, and "ggplot2" [17].

Ethical approval and funding

The Medical Ethical Committee of the Maastricht University Medical Centre evaluated the study protocol and declared that the Expect Study does not fall under the Medical Research Involving Human Subjects Act (METC-17-4-057). All participants gave informed consent.

Financial support for this study was provided entirely by a grant from ZonMw (The Netherlands Organization for Health Research and Development; federal funding). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Results

The survey regarding the risk-threshold preference was sent to 136 healthcare professionals. *(53 midwives, 32 gynecologists, and 51 residents).* In total, 43 (32%) healthcare professionals completed the questions regarding the PE risk-threshold. Response rates per type of healthcare professional were similar: *midwives 30% (n=16), gynecologists 31% (n=10), and residents 33% (n=17)*. The boxplots, displayed in Figure 5.1, indicate that none of the risk-thresholds were clearly rejected, but that there was no evident preference for a certain risk-threshold either.

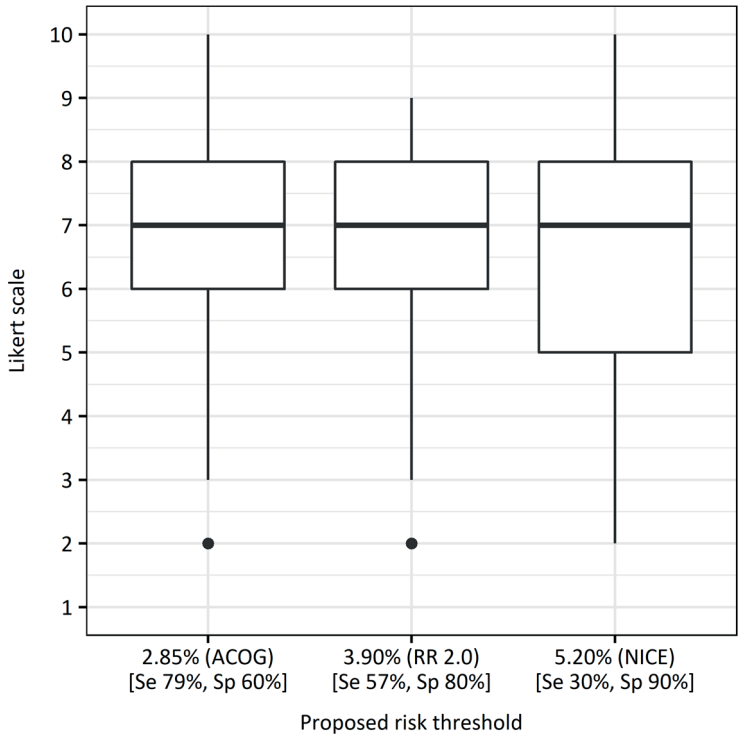


Figure 5.1 Boxplots of preferences of healthcare professionals for given risk-thresholds. Likert scale: 1 totally disagree – 10 fully agree; Se, sensitivity; Sp, specificity; ACOG, the American College of Obstetricians and Gynecologists; RR, relative risk; NICE, the National Institute for Health and Clinical Excellence

During the regional discussions healthcare professionals unanimously stressed that they preferred a prediction tool suitable for shared decision-making. In their opinion, in the case of predicted risk exceeding the chosen cut-off value, the first step should be discussing the LDA recommendation with the pregnant woman. Furthermore, the prediction tool should provide relevant information and insight for both healthcare professional and pregnant woman regarding the predicted risk. When these conditions are met, using the prediction tool as a first step to start the discussion regarding LDA prophylaxis, a threshold with a high sensitivity (high detection rate) was preferred over one with a high specificity (low false positive rate). However, regardless of the detection rate, specificity should be kept at an

acceptable level.
The majority preferred either a threshold of 2.85% or 3.90%. At the same time, healthcare professionals strongly in favor of 5.20% threshold did not agree with 2.85%. It was felt that the number of test positives should not exceed roughly a third of the population. On the other hand, healthcare professionals in favor of the 2.85% threshold stressed that at least everyone with an increased risk should be counselled. The observed incidence rate in external validation study was 2.9% [8]. Thus, it was decided that every woman with a PE risk above the population average should be informed regarding the option of LDA.
During the final LOC meeting, taking all considerations into account, it was decided to that a threshold should be employed and that LDA treatment was to be discussed with the pregnant woman if estimated PE risk was greater than 3.0%. In the external validation study, this threshold corresponded with a sensitivity and specificity of 75% and 64%, respectively [8]. To facilitate the shared decisional approach, the results of the prediction were visualized at a linear scale and provided together with relevant patient brochures.

Table 5.1 Baseline characteristics of the Expect II study cohort (data expressed as mean +/- standard deviation, median (interquartile range), or n (%)

Characteristics	Expect II cohort n=850
Age, years	30.7 +/- 4.0
University, or higher vocational education, n (%)	500 (58.8)
Body mass index, kg/m2	24.8 +/- 4.8
Smoking during pregnancy, n (%)	38 (4.5)
History of chronic hypertension, n (%)	17 (2.0)
Family history of pre-eclampsia (biological mother), n (%)	42 (4.9)
Nulliparous, n (%)	415 (48.8)
Spontaneous conception, n (%)	772 (90.8)
History of pre-eclampsia, n (%)	50 (5.9)
Estimated pre-eclampsia risk percentage, median (interquartile range)	2.7 (1.1-4.3)
Estimated pre-eclampsia risk >3.0%, n (%)	364 (42.8)

In total 866 women provided informed consent, of these 850 (98%) completed the questionnaire at enrolment. Table 5.1 shows the characteristics of the women at enrolment of the Expect Study II, a flow-chart of study enrolment is provided in Figure 5.2, supplementary Figure S5.1 shows the distribution of predicted PE risks of this population. Table 5.2 shows the results of the answers regarding the questions whether PE risk prediction and LDA treatment were discussed during the prenatal visits. A total of 522 women (61%) stated that the results of their estimated PE risk were discussed during the antenatal visits. Estimated risks were not discussed with 265 women (31%), and 63 women (7%) could not recall whether it was discussed.

Table 5.2 Reported rates of discussing pre-eclampsia risk and low-dose aspirin prophylaxis

	Low pre-eclampsia risk n (%)	Increased pre-eclampsia risk n (%)	All women n (%)
Total	486 (100.0)	364 (100.0)	850 (100.0)
Pre-eclampsia risk discussed			
Yes	249 (51.2)	273 (75.0)	522 (61.4)
No	199 (40.9)	66 (18.1)	265 (31.2)
Uncertain	38 (7.8)	25 (6.9)	63 (7.4)
Low-dose aspirin discussed			
Yes	71 (14.6)	294 (80.8)	365 (42.9)
No	400 (82.3)	63 (17.3)	463 (54.4)
Uncertain	15 (3.1)	7 (1.9)	22 (2.6)

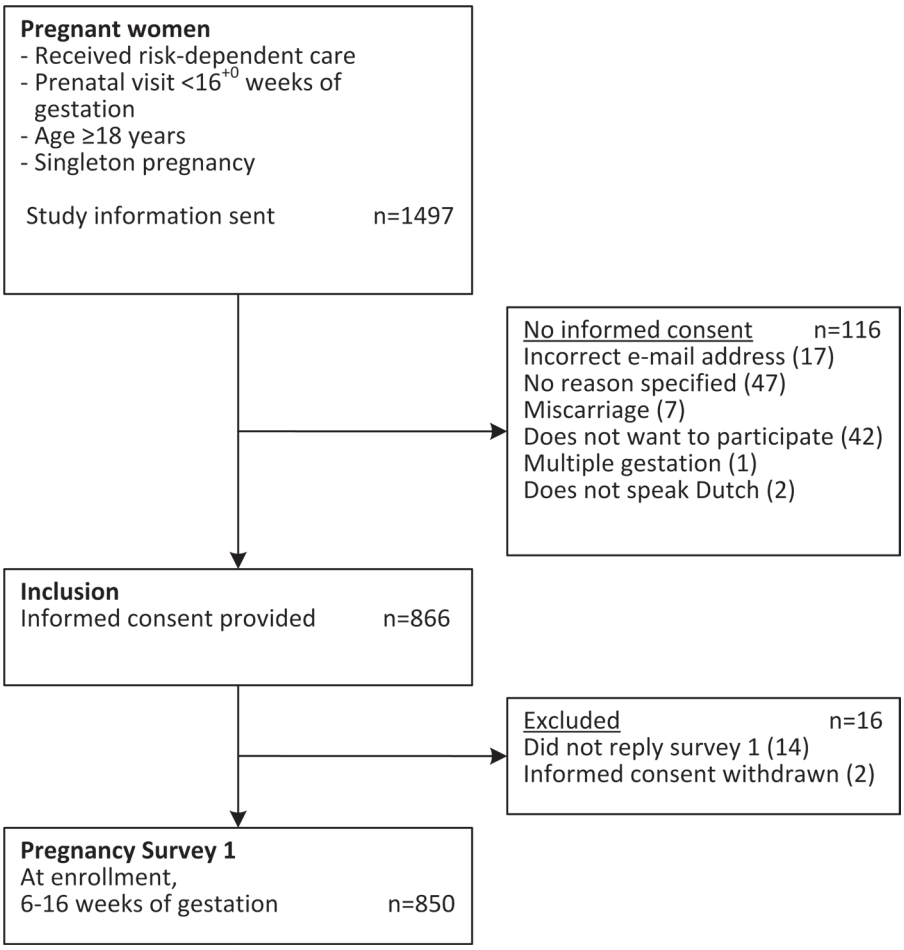


Figure 5.2 Flowchart participant enrollment Expect Study II

An estimated risk exceeding 3.0% was adopted as threshold for discussing LDA. In this subgroup of 364 women with an increased risk, PE risk and LDA prophylaxis were discussed with 273 (75%) and 294 (81%) women, respectively. Figure 5.3 shows the percentages of women who stated their healthcare professionals discussed the PE risk and LDA prophylaxis per risk category. This graph indicates a positive correlation between the predicted PE risk and discussion rates of both PE risk and LDA by healthcare professionals. For women identified with a risk exceeding 3.0%, predicted PE risk was a strong positive determinant of discussing PE risk (odds ratio per percent increase 1.34; 95%CI 1.18-1.56; p<0.01), and of discussing LDA prophylaxis (odds ratio per percent increase 1.28; 95%CI 1.18-1.40; p<0.01). Thus, healthcare professionals are significantly more likely to discuss both the predicted PE risk and LDA recommendation at increased PE risk estimates.

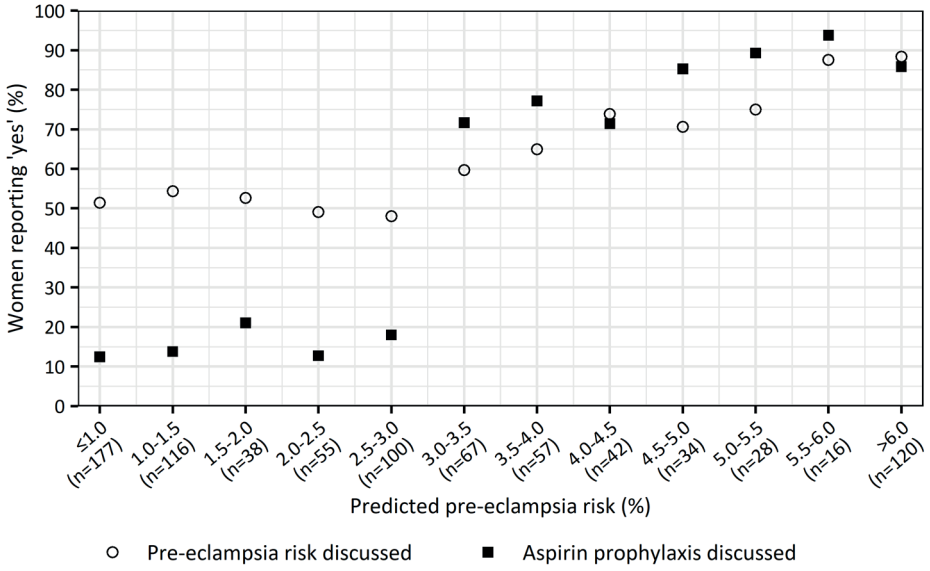


Figure 5.3 Adherence rates of discussing pre-eclampsia risk and low-dose aspirin prophylaxis per estimated risk category

Figure 5.4 shows healthcare professionals' adherence rate throughout the study period. Therefore, we plotted LDA discussion rates reported by women using the study duration as a continuous variable. At the start of our implementation study, adherence rates ranged from 45 to 65% but eventually rose to approximately 85% and remained constant throughout the study period.

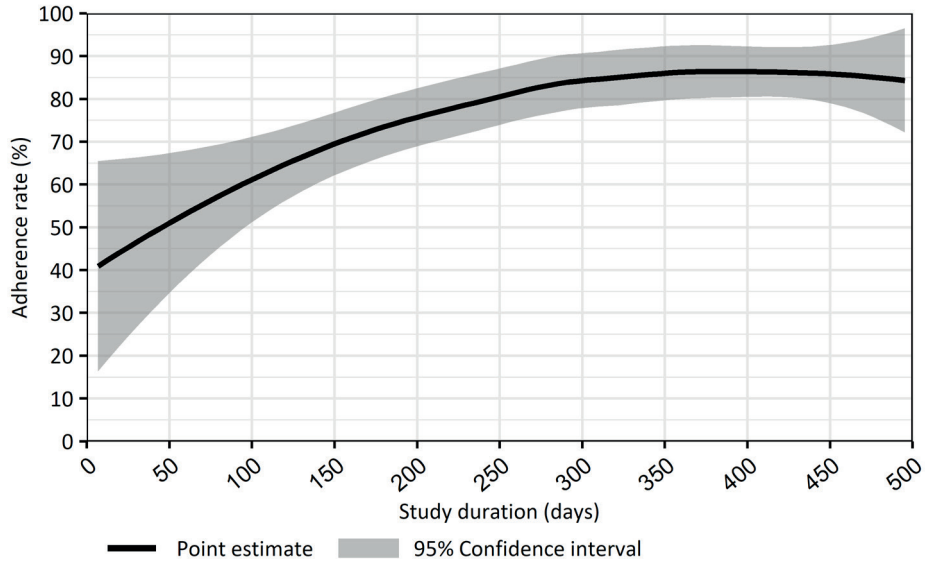


Figure 5.4 Adherence rates of discussing low-dose aspirin prophylaxis during the study period

Discussion

Although, there is an enormous rise in models being published and an increasing amount is externally validated, only a few studies report the implementation of a prediction model [18,19]. To our knowledge, also reported by Kleinrouweler et al. [20], this is the first study to describe the implementation and usage of a prediction model predicting absolute risks for preventive strategies in daily obstetric practice.

Strengths and limitations

Before a prediction model can be used as a basis for clinical decision making, ideally, thresholds should be selected that indicate which risks are considered as an increased risk [21,22]. Although, the publication of prediction models increases rapidly, the amount of models applied in daily practice is still limited. As a result, most healthcare professionals may not be used to interpreting risk estimates. This may explain the low response rate and the lack of consensus in the survey regarding the threshold selection.

In this study, a three-step process was used in order to select suitable risk-thresholds. Reilly et al report the feelings of diminished autonomy by the healthcare professional as one of the potential barriers when applying a decision rule [21]. In the final LOC meeting the shared decisional approach was strongly stressed which may have diminished this potential barrier.

A strength of our study is its prospective multicenter design. Particularly in The Netherlands, recruitment in multiple centers is essential, because most pregnant women receive antenatal care at outpatient midwifery clinics [23]. Furthermore, by using our prediction tool as an inclusion method, we were able to link the received healthcare services to the estimated risk profiles of pregnant women.

The Expect Study II focused on analyzing the impact and results of risk-based care. As a result, only women for whom the prediction tool was used were eligible for inclusion. Usage of our prediction tool as inclusion method enabled us to link the questionnaires completed by women to their individual PE risk prediction. The prediction tool was developed for usage in the general population and was promoted as such [14]. Furthermore, all obstetric healthcare professionals of our region committed themselves to use the prediction tool. Nevertheless, this may have introduced some selection, since pro-active healthcare professionals may be over-represented among the professionals who use our prediction tool. The intensive usage of the prediction tool throughout the region and the multitude of collaborating centers diminishes the amount of selection.

Recommendation of LDA treatment should preferably be based on the PE risk prediction by using a shared decision-making approach. However, for most risk categories more women reported that they discussed LDA prophylaxis than that they discussed their PE risk. Thus, either their PE risk was not discussed or they did not recall the primary reason of discussing LDA prophylaxis. Our data do not allow analyzing possible reasons for this discrepancy. One possibility could be differences in women´s ability to recall both topics since aspirin is an easy, well-known word among non-professionals whereas pre-eclampsia is not. This hypothesis may be supported by the fact that the proportion of women not recalling whether their PE risk was discussed (7.4%) is greater than the proportion of women not recalling whether aspirin was discussed (2.6%).

Interpretation

Discussion of LDA treatment was reported by 81% of women with an elevated PE risk. Compared to previous studies in obstetrics regarding protocol and guideline adherence, this percentage is relatively high [10,12,13]. Additionally, a significant correlation was found between discussing LDA prophylaxis and the predicted PE risk. LDA prophylaxis was discussed more frequently with women having higher PE risk estimates, these women potentially have the highest individual benefit from LDA treatment.

As can be observed in Figure 5.4, the adherence rates tended to increase during the study period. At the start of the implementation of our prediction tool along with the selected threshold, LDA recommendation was at best mediocre and comparable to adherence rates previously reported [13]. However, roughly after nine months of implementation, adherence rates rose up to 85% and remained consistent during the study period.

Recent research emphasized the potential benefit of LDA treatment in women at high risk of PE. The ASPRE trial, a randomized clinical trial towards the effect of LDA treatment in preventing pre-eclampsia, used a prediction model as well to identify the high-risk group [19]. Compared to the model used by the LOC, the ASPRE model has a similar sensitivity but outperforms in specificity. However, the ASPRE model does not solely rely on routinely available predictors and uses biochemical markers as well as the uterine-artery pulsatility index. The addition of these predictors mainly reduces the false positive rate [8]. However, LDA prophylaxis from 12 weeks of gestation is inexpensive and does not result in adverse fetal effects, which reduces the disadvantages of a high false positive rate. As a result, it is arguable whether the costs associated with these additional predictors are proportional to their benefits [24].

Currently, there is no consensus about the best screening method for identifying women at risk of PE. The advantage of a prediction model over a list of risk factors is that it provides both the healthcare professional and the pregnant women with the insight of the absolute risk. Moreover, prediction models weigh several risk factors and their possible inter-relations simultaneously allowing for a more personalized estimation of the absolute risk [25]. This information enables healthcare professionals to use a shared decisional approach. As a result, pregnant women have the opportunity to participate actively in the choices of additional healthcare services aimed at the prevention of PE.

Future research should focus on barriers that hampers the usage of a risk prediction tool by healthcare professionals. Moreover, reasons of non-adherence regarding recommendations provided by the prediction tool should be addressed. Additionally, more insight is needed about the shared decisional approach regarding the choice of LDA prophylaxis. The contradictory results between reporting rates whether PE risk was discussed and whether LDA prophylaxis was discussed (Figure 5.3), suggests that a substantial group of women may not correctly recall or understand the reasons of LDA prophylaxis. In that case, these women are unlikely to be able to make an informed choice.

Conclusion

Consensus regarding a suitable risk cut-off threshold to identify women at risk of PE was reached. Healthcare professionals agreed upon employing a tool with a high detection rate (cut-off: 3.0%, sensitivity 75%, specificity 64%) followed by shared decision between pregnant woman and healthcare professional on LDA prophylaxis. The adherence to this recommendation was 81%, indicating adequate implementation.

References

1. Mol, B.W.J., Roberts, C.T., et al., Pre-eclampsia. Lancet, 2016. 387(10022): p. 999-1011.
2. Meertens, L.J.E., Scheepers, H.C.J., et al., Should women be advised to use calcium supplements during pregnancy? A decision analysis. Matern Child Nutr, 2018. 14(1):p. 1-8.
3. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. Am J Obstet Gynecol, 2017. 216(2): p. 110-120.e6.
4. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. Obstet Gynecol, 2013. 122(5): p. 1122-31.
5. LeFevre, M.L., Low-dose aspirin use for the prevention of morbidity and mortality from preeclampsia: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med, 2014. 161(11): p. 819-26.
6. National Collaborating Centre for Women's Children' Health, National Institute for Health and Clinical Excellence: Guidance, in Hypertension in Pregnancy: The Management of Hypertensive Disorders During Pregnancy. RCOG Press, 2010. Clinical Guideline 107.
7. Mone, F., Mulcahy, C., et al., Should we recommend universal aspirin for all pregnant women? Am J Obstet Gynecol, 2017. 216(2): p. 141.e1-141.e5.
8. Meertens, L.J.E., Scheepers, H.C.J., et al., External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study. Fetal Diagnosis and Therapy, 2019. 45(6): p. 381-393.
9. Smits, L.J.M., van Montfort, P., et al., De Limburgse aanpak: Populatiebrede preventie van pre-eclampsie. NTOG, 2019. 132(3): p. 130-132.
10. Segaar, D., Bolman, C., et al., Identifying determinants of protocol adoption by midwives: a comprehensive approach. Health Educ Res, 2007. 22(1): p. 14-26.
11. Offerhaus, P., Fleuren, M., et al., Guidelines on anaemia: effect on primary-care midwives in The Netherlands. Midwifery, 2005. 21(3): p. 204-11.
12. Zeitlin, J., Manktelow, B.N., et al., Use of evidence based practices to improve survival without severe morbidity for very preterm infants: results from the EPICE population based cohort. Bmj, 2016. 354: p. i2976.
13. Jakes, A., Bhaduri, M., et al., 272. Aspirin prescribing in pregnancy: Are we doing it? Pregnancy Hypertension, 2018. 13: p. S37.
14. van Montfort, P., Willemse, P.P.M.J., et al., Implementation and Effects of Risk-Dependent Obstetric Care in the Netherlands (Expect Study II): Protocol for an Impact Study. JMIR Res Protoc, 2018. 7(5): p. e10066.
15. Lemmens, S.M., Agreement Conform Current Operational Rules and Directives (ACCORD): A Novel Tool to Reach Multidisciplinary Consensus. J Womens Health Gyn, 2019. 5: p. 1-11.
16. Syngelaki, A., Bredaki, F.E., et al., Body mass index at 11-13 weeks' gestation and pregnancy complications. Fetal Diagn Ther, 2011. 30(4): p. 250-65.
17. R Core Team, R: A Language and Environment for Statistical Computing, 2019, R Foundation for Statistical Computing: Vienna, Austria.
18. Melman, S., Schoorel, E.N., et al., SIMPLE: implementation of recommendations from international evidence-based guidelines on caesarean sections in the Netherlands. Protocol for a controlled before and after study. Implement Sci, 2013. 8: p. 3.
19. Rolnik, D.L., Wright, D., et al., Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med, 2017.
20. Kleinrouweler, C.E., Cheong-See, F.M., et al., Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol, 2016. 214(1): p. 79-90.e36.
21. Reilly, B.M. and Evans, A.T., Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of internal medicine, 2006. 144(3): p. 201-209.
22. Grol, R., Dalhuijsen, J., et al., Attributes of clinical guidelines that influence use of guidelines in

general practice: observational study. Bmj, 1998. 317(7162): p. 858-61.
23. Visser, G.H.A., Obstetric care in the Netherlands: relic or example? J Obstet Gynaecol Can, 2012. 34(10): p. 971-975.
24. Sentilhes, L., Azria, E., et al., Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med, 2017. 377(24): p. 2399-400.
25. Steyerberg, E.W., Vickers, A.J., et al., Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology, 2010. 21(1): p. 128-38.

Supplementary file 1

Model algorithm, discriminative performance, and predicted probabilities

Original study	External validation study	Model algorithm after recalibration	AUC (95% CI)
Syngelaki 2011	Meertens 2018	Lp = -5.773 + 0.075 (BMI, kg/m2) + 0.022 (age, years) + 1.125 (if Afro-Caribbean) + 0.804 (if South Asian) + 0.526 (if East Asian) + 0.379 (if Mixed) + 0.289 (if ovulation drugs) + 0.598 (if IVF) - 0.233 (if smoker) + 1.519 (if history of chronic hypertension) + 0.643 (if type 1 diabetes mellitus) - 0.332 (if type 2 diabetes mellitus) - 1.329 (if parous, no history of pre-eclampsia) + 0.743 (if parous, history of pre-eclampsia) + 0.580 (if woman's mother had pre-eclampsia).	0.77 (0.72-0.81)[9]

AUC, area under the curve; CI, confidence interval; Lp, linear predictor; BMI, body mass index; IVF, in vitro fertilization

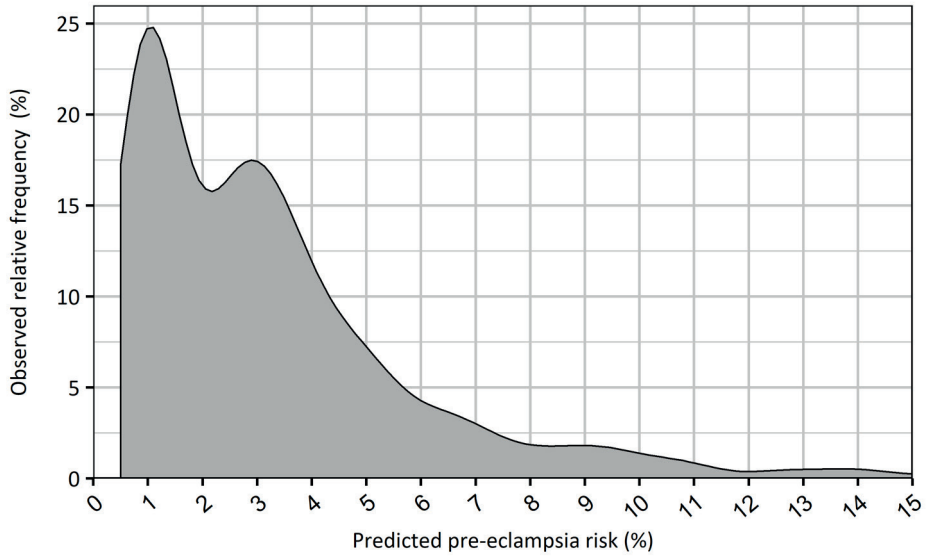


Figure S5.1 Density plot of predicted pre-eclampsia risks

Chapter 6

Low-dose aspirin usage among women with an increased pre-eclampsia risk: a prospective cohort study

Pim van Montfort, Hubertina C.J. Scheepers, Ivo M.A. van Dooren,
Linda J.E. Meertens, Maartje Zelis, Iris M. Zwaan,
Marc E.A. Spaanderman, Luc J.M. Smits

Submitted

Abstract

Background

Low-dose aspirin (LDA) prophylaxis has been shown to reduce women's pre-eclampsia risk. Evidence regarding LDA adherence rates of pregnant women is almost exclusively based on clinical trials, giving a potentially biased picture. Moreover, these studies do not report on determinants of adherence. Since 2017, obstetric healthcare professionals in a Dutch region assess women's pre-eclampsia risk by means of a prediction tool and counsel those with an above population average risk on LDA as a prophylactic measure.

Objective

To assess the rates and determinants LDA usage among women with an increased pre-eclampsia risk in daily practice.

Methods

From 2017 to 2018, 865 women were recruited in multiple centers and prospectively followed using web-based surveys (Expect Study II). Results were compared to findings in a similar cohort from a care-as-usual setting lacking risk-based counseling (Expect Study I, n=2,614).

Results

In total, 306 women had a predicted increased pre-eclampsia risk. LDA usage was higher for women receiving risk-based care as compared to care-as-usual (29.4% vs. 1.5%, RR 19.1; 95%CI 11.2-32.5). Daily LDA usage was positively correlated with both predicted risk and women's concerns regarding pre-eclampsia. Most reported reasons for non- or incomplete use were unawareness of LDA as a preventive intervention, concerns of potential adverse effects, and doubts regarding the benefits.

Conclusion

Risk-based counseling was associated with a higher prevalence of LDA usage, but general usage rates were low. Future research regarding potential factors improving the usage of LDA during pregnancy is necessary.

Introduction

Pre-eclampsia (PE) is an important cause of serious maternal and fetal complications. Despite improved management, curative options preserving the pregnancy remain absent. Preventive measures reducing the risk of PE are therefore an essential part of strategies aimed at decreasing the burden of PE [1].

Besides lifestyle interventions and adequate calcium intake, low-dose aspirin (LDA) treatment is currently one of the key interventions for the prevention of PE [2-4]. Reduction of PE risk has been shown at aspirin dosages between 80 and 150 milligrams per day [5]. The majority of publications on LDA with respect to PE focus on its effectiveness. They mainly differ regarding dosing, gestational window, or target group [2,5]. Published LDA adherence rates are fairly high (66-90%), but mostly measured within clinical trials [2,6]. It is unlikely that women who would not opt for LDA during their pregnancy would be willing to participate in a trial involving LDA usage. Thus, trial-based adherence rates may be seriously biased upwards. Relatively little is known regarding the daily LDL usage rates among pregnant women in daily practice [7].

Several obstetric authorities recommend LDA for women with an increased PE risk, including the American College of Obstetricians and Gynecologists (ACOG), the US Preventive Services Task Force, and the National Institute for Health and Clinical Excellence (NICE, United Kingdom) [8-10]. Nevertheless, 'increased risk of PE' has been defined in different ways and no consensus has yet been achieved. Assessment of PE risk can be performed by using either unweighted or weighted combinations of multiple risk factors. The latter method (i.e., prediction models) has been shown to outperform the use of unweighted risk factors (i.e. NICE criteria) in terms of predictive ability [11,12].

Recently, healthcare professionals in the Southeastern part of the Netherlands implemented an externally validated prediction tool to assess, during the first trimester of pregnancy, the risk of developing PE [11,13,14]. In case of an increased risk, the option of LDA prophylaxis is discussed using a shared-decisional approach. In such an approach, healthcare professionals share the best available evidence with the women in order to make an informed decision together [15]. This observational study reports on LDA usage rates by women with an increased PE risk, as well as on determinants and reasons given for use and non-use.

Methods

Identifying women at increased risk of pre-eclampsia

In 2017, members of the Limburg Obstetric Consortium (located in the Southeastern part of the Netherlands) started to assess women's PE risk during the first antenatal visits by means of a prediction tool. This tool embedded Syngelaki's prediction model, externally validated and recalibrated by Meertens et al [11,16]. This model is based on maternal characteristics (age, BMI, ethnicity, mode of conception, family history, medical history, and obstetric history) and was made available for all healthcare professionals of the region.

A detailed description of the content of risk-based care is reported elsewhere (van Montfort et. al, accepted, [13]). In short, women with a PE risk exceeding the population average risk (>3.0%; sensitivity 75%, specificity 64%) should be counseled regarding the option of LDA-prophylaxis (80-100 milligrams daily) in a shared-decisional approach.

Data collection

All women ≥18 years old with a singleton pregnancy were eligible for inclusion. Women were recruited from 2017 to 2018 at their first prenatal visit (<16 weeks of pregnancy), when their healthcare professional used the prediction tool. Women were recruited from multiple centers; five hospitals and 26 autonomous midwifery practices, all belonging to the geographical area of the LOC.

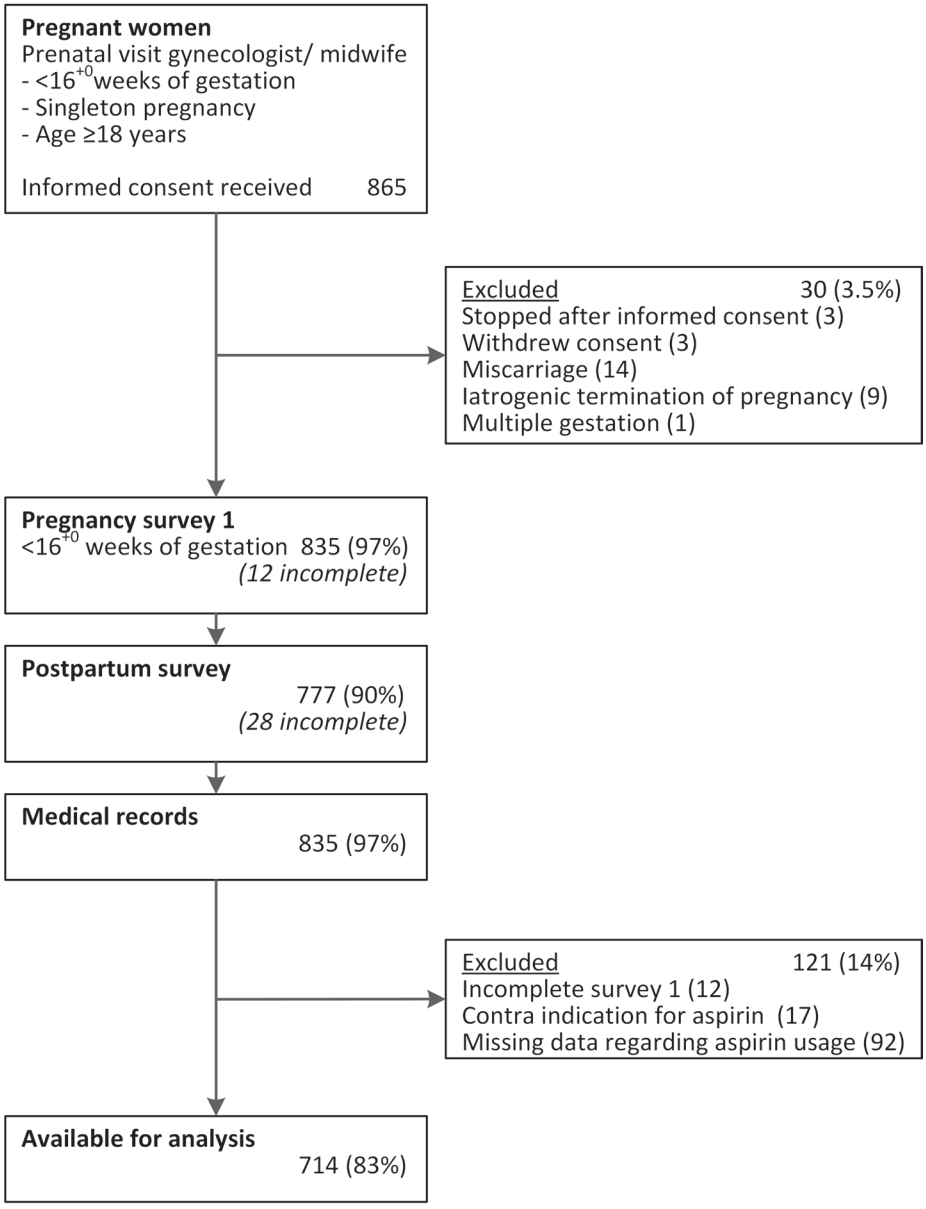


Figure 6.1 Flowchart of participant enrollment of Expect Study II

For the analyses in this paper, women with incomplete data regarding LDA usage, or a contraindication for LDA usage were excluded. A detailed study protocol has been published previously [13]. Briefly, after providing informed consent, the results of the risk assessment were automatically logged. Enrolled women received four online surveys at intervals (at enrolment, at 24 weeks of pregnancy, at 34 weeks of pregnancy, and 6 weeks after the due date). In case of preterm birth, women were automatically redirected to the postpartum questionnaire when completing the questionnaire sent at 24 or 34 weeks of pregnancy. In addition, medical records and discharge letters were retrieved.

The first survey contained questions related to the first antenatal visits. Women were asked whether they were informed regarding LDA and whether they intended to use LDA. Additionally, women were questioned how often they worried about complications related to PE, such as PE itself, small-for-gestational-age (SGA) infancy, and preterm birth (PTB). They could choose from the options not at all, sometimes, regularly, and often. Answers were transformed to a four-point scale (0, not at all; 1, sometimes; 2, regularly; 3, often).

The postpartum survey included questions related to LDA usage throughout the pregnancy. Women who stated to have used LDA received additional questions regarding the gestational window of LDA usage and whether they took it daily. Women stating they did not use LDA received additional questions with respect to their most decisive reason of non-use. Women were able to choose out of predefined options, but were also able to provide a different reason and leave additional remarks.

Statistical analysis

Usage of LDA was analyzed with respect to women's estimated PE risk. Any LDA usage was defined as LDA usage regardless of the numbers of pills taken, duration, or frequency. Per protocol LDA usage was defined as the usage as described in the risk-based care pathways: daily LDA usage from $<16^{+0}$ weeks of gestation up to 36 weeks of gestation or, in case of preterm birth, up to one week before birth. We cross tabulated the proportions of women whom reported to have discussed the option of LDA, any LDA usage, and per protocol LDA usage with respect to the estimated PE risk (low risk/ increased risk).

Data of the Expect Study I (n = 2,614), a similar multicenter prospective cohort study conducted in the same region from 2013 to 2015, were used to represent the care-as-usual approach lacking risk-based recommendations [11,17]. For Expect Study I, a paper and pencil questionnaire was available on request. However, the vast majority of women completed the web-based version of the questionnaires. The data contained information on usage of LDA, but not whether LDA was used in accordance with the risk-based care recommendations. As a result, only the proportions of any LDA usage could be compared between risk-based care and former care-as-usual.

Proportions of any LDA usage by women who received care-as-usual and women who received risk-based care were plotted using the estimated risk as a continuous variable. A nonparametric local weighted regression (loess regression) was applied to fit the curves [18].

For analysis of determinants correlated with per protocol LDA usage, a multiple logistic regression was performed. This analysis was restricted to women with an increased risk whom were informed by their healthcare professional regarding LDA, since only these women are able to make an informed decision. Factors taken into account were estimated PE risk (continuous); reported educational level (tertiary yes/ no); concerns regarding developing PE (continuous); concerns regarding developing complications related to PE

(SGA, continuous; PTB, continuous); and type of healthcare professional responsible for LDA counseling (midwife/ gynecologist). For the continuous determinants, we verified whether assumptions of linearity were not violated using frequency plots. All statistical analyses were performed using R statistical software version 3.6.0 [19].

Ethical approval

The Medical Ethical Committee of the Maastricht University Medical Centre evaluated both Expect Study protocols I and II and declared that both observational studies do not fall under the Medical Research Involving Human Subjects Act (METC-13-4-053 and METC-17-4-057, respectively). Online informed consent was obtained from all participants.

Results

Figure 6.1 displays a flowchart of study enrolment. Informed consent was provided by 865 women. Of these, 30 women were excluded from the study cohort for various reasons. Additionally, 121 women were excluded from the current analysis because of either incomplete data (n=104), or a contraindication for LDA usage (n=17). In total 714 women were available for the analyses. Those excluded (n=121) were more likely to have a primary/ secondary educational level (57.7%) than those included (n=714) in the study (38.2%). Otherwise no differences in characteristics were observed for parity, body mass index, age, ethnicity, unassisted conception, and estimated PE risk (data not presented).

Table 6.1 Baseline characteristics of the Expect Study cohorts I and II

Baseline characteristics <16 weeks of gestation	Expect Study I care-as-usual cohort (n=2,614)	Expect Study II risk-based care cohort (n=714)
Age, years; mean +/- sd	30.2 +/- 3.9	30.8 +/- 4.0
Ethnicity		
Caucasian; n (%)	2533 (96.9)	698 (97.8)
Other; n (%)	81 (3.1)	16 (2.2)
Educational level		
Primary or secondary; n (%)	1194 (45.7)	273 (38.2)
Tertiary level of education; n (%)	1420 (54.3)	441 (61.8)
Body mass index, kg/m2; mean +/- sd	24.2 +/- 4.3	24.8 +/- 4.6
Smoking during pregnancy		
Yes	319 (12.2)	32 (4.5)
No	2137 (81.8)	682 (95.5)
Chronic hypertension	28 (1.1)	16 (2.2)
Conception		
Natural; n (%)	2440 (93.3)	644 (90.2)
Ovulation induction; n (%)	93 (3.6)	35 (4.9)
In vitro fertilization; n (%)	81 (3.1)	35 (4.9)
Obstetric history		
Nulliparous; n (%)	1326 (50.7)	360 (50.4)
Prior PE; n (%)	72 (2.8)	38 (5.3)
No prior PE; n (%)	1216 (46.5)	316 (44.3)
Family history of PE; n (%)	131 (5.0)	36 (5.0)
Counselling of PE risk		
by midwife; n (%)	NA	523 (73.2)
by obstetrician; n (%)	NA	191 (26.8)
Estimated PE risk %; median (IQR)	2.5 (1.0-3.6)	2.7 (1.1-4.2)
Increased PE risk; n (%)	974 (37.2)	306 (42.9)
Estimated PE risk % for women identified with an increased risk; median (IQR)	4.2 (3.4-5.8)	4.7 (3.6-6.8)

PE, pre-eclampsia; sd, standard deviation; IQR, interquartile range; NA, not available

An overview of baseline characteristics for women enrolled in Expect Study I or II (women received care-as-usual and risk-based care respectively) is given in Table 6.1. At baseline, the characteristics of women enrolled for both studies do not substantially differ. However, for Expect Study II relatively more women had a history of PE. As a result the percentage of women identified with an increased PE risk was slightly higher (37.2% vs. 42.9%).

Recommendations of aspirin usage

According to the recommendations of the regional consortium, in risk-based care, women identified with an increased PE risk (risk >3.0%) should be informed regarding LDA usage for the prevention of PE. A large majority of the women (79%, n = 241) reported having discussed LDA with their healthcare provider, indicating a high, but not optimal, adherence rate to regional recommendations by healthcare professionals. Of these women, 94 (39%) intended to use LDA throughout the pregnancy of which 52 eventually used LDA according to protocol resulting in a per protocol usage rate of 22% (Figure 6.2).

Low-dose aspirin usage rates

Postpartum, of all enrolled women 113 (15.8%) reported having used LDA during their pregnancy and 87 (12.2%) used it according to protocol (Table 6.2). Among women with an increased PE risk (>3%), this results in an average usage rate of 29.4% and a per protocol usage rate of 24.8%. Furthermore, a small amount of women (n= 11), used LDA throughout the pregnancy despite not being identified with an increased PE risk.

Table 6.2 Proportions of counseling and usage of low-dose aspirin in relation to predicted pre-eclampsia risk

	All women (n=714)	PE risk ≤3% (n=408)	PE risk >3% (n=306)
Total	714 (100)	408 (100)	306 (100)
Aspirin prophylaxis discussed			
Yes	295 (41.3)	54 (13.2)	241 (78.8)
No	419 (58.7)	354 (86.8)	65 (21.2)
Uncertain	19 (2.7)	13 (3.2)	6 (2.0)
Aspirin used			
Yes	113 (15.8)	23 (5.6)	90 (29.4)
According to protocol	87 (12.2)	11 (2.7)	76 (24.8)
No	601 (84.2)	385 (94.4)	216 (70.6)

PE, pre-eclampsia

The majority of women who started using LDA during their pregnancy in risk-based care, used it according to protocol. Of the 26 women who used LDA, but not according to protocol, three stopped due to complaints they attributed to LDA (diarrhea n=1, nose bleeding n=2). Two women reported they forgot to continue the LDA prophylaxis, and eleven women ended LDA usage at the beginning of their third trimester. Additionally, we could not asses per protocol usage for nine women who did not recall the date they stopped using LDA.

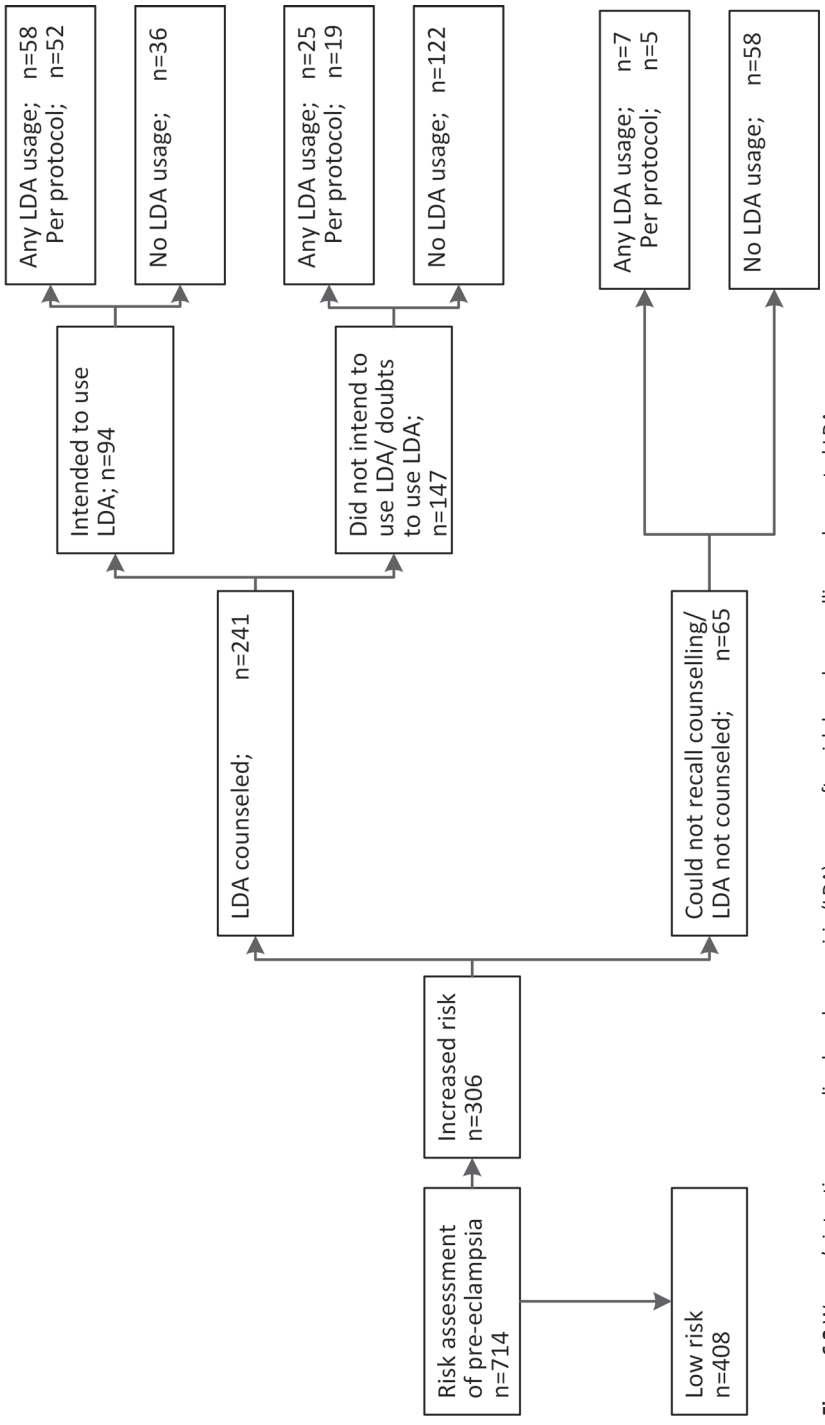


Figure 6.2 Women's intentions regarding low-dose-aspirin (LDA) usage after risk-based counselling and reported LDA usage

For the care-as-usual approach (Expect Study I, 2013-2015), LDA usage was nearly non-existent with only 23 out of 2,614 women reporting to have used LDA (0.9%). We retrospectively calculated the PE risk of these women, resulting in 974 women being classified with an increased PE risk of which 15 (1.5%) used LDA. In risk-based care, women with an increased PE risk estimation were more likely to use LDA (odds ratio 19.1; 95%CI 11.2-32.5). This disparity even rises for higher PE risk estimations.
Supplementary figure S6.1 provides an overview of the distribution of observed PE risk estimates. Figure 6.3 displays the proportions of any LDA usage by estimated PE risk for both risk-based care and the care-as-usual approach. We limited the graph to PE risk estimates of ≤15%, which comprises 99% of the observations. Furthermore, per protocol LDA usage rates are also shown for the for the risk-based care cohort. This graph indicates a positive correlation between estimated PE risk and LDA usage in women receiving risk-based care.

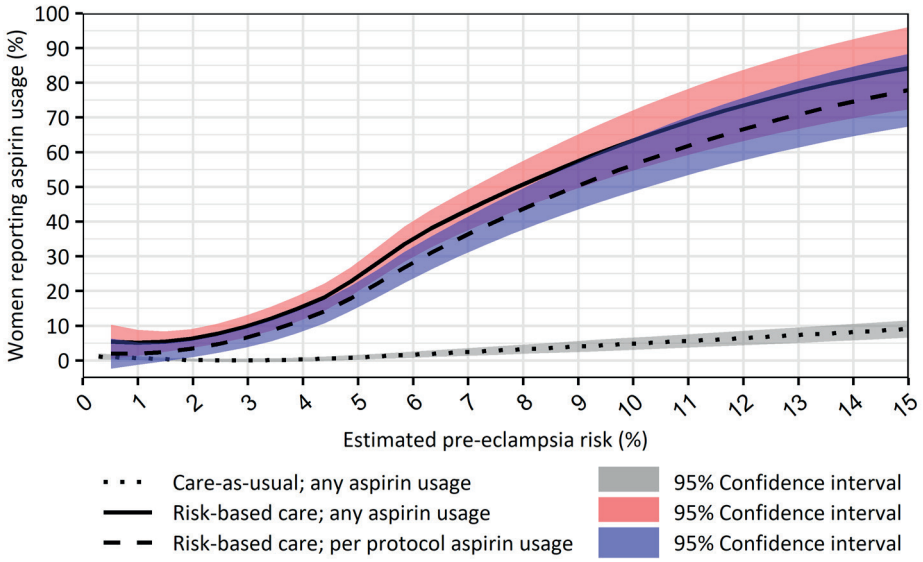


Figure 6.3 Estimated pre-eclampsia risks and low-dose aspirin usage rates by women receiving care-as-usual or risk-based care

Determinants of low-dose aspirin usage

The type of healthcare professional (midwife or obstetrician) informing women about LDA, was significantly correlated with per protocol LDA usage (odds ratio 2.34, indicating higher usage under obstetric-gynecological care; 95%CI 1.32-4.18). However, this association was no longer apparent when correcting for the estimated PE risk (adjusted odds ratio 1.32; 95%CI 0.66-2.60). In the adjusted analysis, Table 6.3, only the degree of women's concerns regarding a pregnancy complicated by PE was statistically significantly associated with per protocol LDA usage when controlling for the estimated PE-risk (adjusted odds ratio 1.99; 95%CI 1.35-2.98).

Table 6.3 Multiple logistic regression of potential determinants of per protocol low-dose aspirin usage among women with an increased risk with whom aspirin usage was discussed

Determinant	No. of participants	No. with per protocol aspirin usage n (%; 95%CI)	Unadjusted odds ratio (95% CI)	Adjusted odds ratio* (95% CI)
All	241	71 (29; 24-36)	-	-
Estimated PE risk			1.23 (1.14-1.35)	1.18 (1.09-1.30)
Educational level				
Primary or secondary	106	30 (28; 21-38)	1 [Reference]	1 [Reference]
Tertiary	135	41 (30; 23-39)	1.10 (0.63-1.94)	1.36 (0.72-2.62)
Concerns regarding PE			2.23 (1.64-3.09)	1.99 (1.35-2.98)
Concerns regarding SGA			1.20 (0.87-1.66)	0.98 (0.65-1.46)
Concerns regarding PTB			1.31 (0.97-1.77)	0.79 (0.52-1.19)
Counselling of PE risk				
by midwife	162	38 (23; 18-31)	1 [Reference]	1 [Reference]
by obstetrician	79	33 (42; 32-53)	2.34 (1.32-4.18)	1.32 (0.66-2.60)

*Odds ratios adjusted for variables listed in left column. PE, pre-eclampsia; SGA, small-for-gestational-age infancy; PTB, preterm birth; CI, confidence interval

Using a semi-qualitative approach, we analyzed women's reasons for not using LDA during the pregnancy. A list of mentioned reasons for not using LDA and their frequencies is shown in Table 6.4. Surprisingly, despite having an increased PE-risk, 92 out of 216 women (43%) reported that they believed that the LDA recommendations were not applicable to their situation. This proportion was similar in subgroups with higher PE risk estimates. This questions whether these women received and understood the information regarding LDA usage. Indeed, 39 of these 92 women reported during the first survey that they were not informed regarding LDA.

Other frequently mentioned reasons for not using LDA were that women felt that either the potential benefit of LDA was too low (n=64; 30%), or that they did not want to use (preventive) medication during their pregnancy (n=27; 13%). In the remarks section, concerns regarding potential adverse effects of LDA and medicalization of the pregnancy were frequently expressed as important reasons for not using LDA. Interestingly, these proportions were not much different among women with high PE risk estimates, or among women with a history of PE.

Table 6.4 Reported reasons for not using low-dose aspirin during pregnancy

Specified reason	PE risk >3% n (%)	PE risk >5% n (%)
It was not applicable to my situation	92 (43.2)	27 (39.1)
It was not recommended by my healthcare professional	14 (6.6)	6 (8.7)
The potential benefit is too low for my situation	64 (30)	17 (24.6)
Because aspirin is a drug	27 (12.7)	8 (11.6)
No clear reason (e.g. forgotten)	8 (3.8)	5 (7.2)
Miscellaneous	5 (2.3)	3 (4.3)
Unknown	6 (2.8)	3 (4.3)
Total	216 (100)	69 (100)

PE, pre-eclampsia

Discussion

Main results

Our prediction tool identified 306 women (43%) with an increased PE risk. The majority of these women (n=241; 79%) reported that their healthcare professional discussed the option of LDA prophylaxis with them, suggesting adequate adherence of healthcare professionals to the risk-based care recommendations. Usage rates of LDA increased as compared to care-as-usual (29.4% vs. 1.5%, RR 19.1; 95%CI 11.2-32.5). Daily aspirin usage was positively correlated with both predicted risk and the degree of women's concerns regarding PE. Most reported reasons for non- or incomplete use were unawareness of LDA as preventive intervention, concerns of potential adverse effects, and doubts regarding the benefits.

Strengths and limitations

This is a large observational study to investigate LDA usage rates by women with an increased PE risk, as well as on determinants and reasons given for use and non-use. Another strength is the multicenter study design. Combined with the broad inclusion criteria this should have ensured an unselected population as possible. Nevertheless, women of Caucasian origin in our cohort are overrepresented and the majority of women are well educated. Since impaired health literacy is correlated with nonadherence [20], usage rates in our study may be somewhat overestimated.

A potential limitation in this paper is that LDA usage was based upon self-report. We were unable to reliably verify LDA usage with medical records or pharmacy registries because LDA is available over-the-counter in the Netherlands. However, there is no clear gold standard available to assess medication use in large-scale studies [21]. It could be possible women answered in a socially acceptable manner resulting in an overestimation of the usage rate [22]. On the other hand, in risk-dependent care, counselling of LDA had the form of a shared decisional process. Usage of medication during pregnancy is not generally perceived as 'good' or 'bad' since women are aware medication may cause adverse effects, but could be beneficial for their health as well [20,23]. Moreover, women were informed that survey results would be processed anonymously and would not be shared with their healthcare professional. The researchers who distributed the web-based surveys were not involved in the care of participants. Therefore, the potential overestimation with respect to the adherence rate due to self-report is probably limited.

Besides socially acceptable answers, self-report of medication usage is also prone to recall biases. However, women reporting non-usage are likely to be telling the truth [22]. Furthermore, underreporting for pregnancy-related medications as well as medication prescribed for a longer period is limited in prospective studies [24].

Interpretation

Women's adherence regarding medication during pregnancy has been studied for several drugs, such as anti-diabetics, medicines for chronic airway conditions, or anti-inflammatory drugs, with varying adherence rates from 40% to 80% [20,23]. However, these drugs are prescribed because of an apparent (chronic) medical condition such as diabetes, asthma, or inflammatory bowel disease. Therefore, these situations likely differ compared to LDA, which is recommended to prevent pre-eclampsia. Most women with an increased PE risk do not have any medical complaints warranting LDA usage, which probably leads to different

risk-benefit evaluations.

Studies of pregnant women's adherence regarding LDA in particular are limited and mostly result from clinical trials [2,6]. These trials indicate high adherence rates (66-90%). However, trial-based adherence rates may be seriously biased upwards, as women who do not want to use any drugs (i.e. LDA), are unlikely to be willing to participate in such a trial. We found one observational study indicating a lower adherence rate (54%) as well, but within a small cohort (n = 42) and restricted to women with high-risk pregnancies [7]. Another observational study, conducted among high-risk women in Iran, did not provide absolute adherence rates [25]. Compared to these reports, the rate of LDA usage of 25% in our cohort is low, but is probably a more realistic estimation of LDA usage in daily practice.

Most guidelines recommend LDA prophylaxis to women with an increased PE risk, but there is no consensus yet as how to identify women with an increased PE risk [8,9,26]. In our study, an externally validated prediction model was used to estimate women's PE risk during the first antenatal visits. Since the risk assessment was used as starting point of the shared decisional process regarding LDA usage, a risk threshold with a relatively high detection rate was used (van Montfort et al., accepted). As a result, women identified with an increased PE risk in our study may have had a lower PE risk on average as compared to other studies. This may have attributed to the lower usage rate. Furthermore, LDA-usage was strongly correlated with the predicted PE risk resulting in high usage rates among women with the highest risks, similar to the rates previously reported.

Despite the lower usage rates in general, LDA usage still improved strongly with an absolute increase of 27.9%. However, during enrollment of the care-as-usual cohort (2013 to 2015) there was no uniform Dutch guideline recommending LDA prophylaxis. Although, many obstetric healthcare professionals were familiar with the NICE guideline for hypertensive disorders [26], especially gynecologists, LDA recommendation depended mainly on the intention of individual healthcare professionals. As a result, the increase of LDA usage may mainly reflect adequate implementation of risk-based-care and uptake of its recommendations by healthcare professionals.

To the best of our knowledge, no studies have yet reported on determinants of LDA usage as well as women's reasons for non-usage of LDA in particular. In the unadjusted analysis, the LDA usage rate was associated with the type of healthcare professional responsible for LDA counselling. However, low-risk women remain primarily under the supervision of autonomous midwives in the Dutch maternity care system. As a result, women's risk should be taken into account. Indeed, when correcting for PE risk at baseline, this effect was no longer apparent. The degree of concern about possible complications related to PE (SGA infancy and PTB) were not significantly linked to the usage rate in the adjusted analysis. However, women may be unaware that PE may result into SGA infancy or (iatrogenic) PTB.

The adjusted analysis also indicates that both the estimated PE risk as well as the level of concern regarding PE are positively correlated with LDA usage. This is in line with previous research, which suggests that women's beliefs about medication and its effectiveness are a crucial factor in determining their adherence [20,23]. This also fits with our finding that most frequent reasons of non-use were concerns regarding potential adverse effects of LDA and doubts regarding the potential benefits resulting from LDA prophylaxis. Moreover, the finding that most women who started using LDA, used it according to protocol suggests those women were conscious about their choice.

Informing women about the low prevalence of effects of LDA, which are also mild, [27-29] may

be a central factor to improve adherence rates. Furthermore, a substantial proportion of women stating that LDA was not applicable to their situation reported LDA had not been discussed with them. Our data do not allow distinguishing whether LDA was not discussed by the healthcare professional, or whether these women could not recall that LDA was discussed. Clear communication of PE risk and adequate counselling regarding potential benefits and harms of LDA may positively influence women's decision regarding LDA usage during pregnancy. Future qualitative research, for example with the aid of focus groups among both healthcare professionals as well as pregnant women, may improve our insight and understanding regarding the key elements at play in the decisional process regarding preventive LDA usage.

Conclusion

Implementation of risk-based care improved LDA usage by pregnant women with an increased PE risk, especially among high-risk women. Nevertheless, general usage rates were relatively low. To improve LDA usage rates, more insight in this decisional process is necessary, which underlines the importance of future (qualitative) research regarding preventive LDA usage by pregnant women.

References

1. Mol, B.W.J., Roberts, C.T., et al., Pre-eclampsia. Lancet, 2016. 387(10022): p. 999-1011.
2. Roberge, S., Bujold, E., et al., Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. Am J Obstet Gynecol, 2018. 218(3): p. 287-293.e1.
3. Hofmeyr, G.J., Lawrie, T.A., et al., Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. Cochrane Database Syst Rev, 2018. 10: p. Cd001059.
4. Allen, R., Rogozinska, E., et al., Effect of diet- and lifestyle-based metabolic risk-modifying interventions on preeclampsia: a meta-analysis. Acta Obstet Gynecol Scand, 2014. 93(10): p. 973-85.
5. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. Am J Obstet Gynecol, 2017. 216(2): p. 110-120.e6.
6. Navaratnam, K., Alfirevic, Z., et al., How important is aspirin adherence when evaluating effectiveness of low-dose aspirin? Eur J Obstet Gynecol Reprod Biol, 2017. 219: p. 1-9.
7. Abheiden, C.N., van Reuler, A.V., et al., Aspirin adherence during high-risk pregnancies, a questionnaire study. Pregnancy Hypertens, 2016. 6(4): p. 350-355.
8. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. Obstet Gynecol, 2013. 122(5): p. 1122-31.
9. LeFevre, M.L., Low-dose aspirin use for the prevention of morbidity and mortality from preeclampsia: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med, 2014. 161(11): p. 819-26.
10. National Collaborating Centre for Women's Children' Health, National Institute for Health and Clinical Excellence: Guidance, in Hypertension in Pregnancy: The Management of Hypertensive Disorders During Pregnancy. RCOG Press, 2010. Clinical Guideline 107.
11. Meertens, L.J.E., Scheepers, H.C.J., et al., External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study. Fetal Diagnosis and Therapy, 2019. 45(6): p. 381-393.
12. Wright, D., Syngelaki, A., et al., Competing risks model in screening for preeclampsia by maternal characteristics and medical history. Am J Obstet Gynecol, 2015. 213(1): p. 62.e1-62.e10.
13. van Montfort, P., Willemse, P.P.M.J., et al., Implementation and Effects of Risk-Dependent Obstetric Care in the Netherlands (Expect Study II): Protocol for an Impact Study. JMIR Res Protoc, 2018. 7(5): p. e10066.
14. DiMatteo, M.R., Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Medical care, 2004: p. 200-209.
15. Elwyn, G., Frosch, D., et al., Shared decision making: a model for clinical practice. Journal of general internal medicine, 2012. 27(10): p. 1361-1367.
16. Syngelaki, A., Bredaki, F.E., et al., Body mass index at 11-13 weeks' gestation and pregnancy complications. Fetal Diagn Ther, 2011. 30(4): p. 250-65.
17. Meertens, L.J.E., Scheepers, H.C.J., et al., External Validation Study of First Trimester Obstetric Prediction Models (Expect Study I): Research Protocol and Population Characteristics. JMIR Res Protoc, 2017. 6(10): p. e203.
18. Cleveland, W.S., LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. The American Statistician, 1981. 35(1): p. 54-54.
19. R Core Team, R: A Language and Environment for Statistical Computing, 2019, R Foundation for Statistical Computing: Vienna, Austria.
20. Lupattelli, A., Spigset, O., et al., Adherence to medication for chronic disorders during pregnancy: results from a multinational study. International Journal of Clinical Pharmacy, 2014. 36(1): p. 145-153.
21. van Gelder, M., de Jong, L.A.A., et al., Assessment of medication use during pregnancy by Web-based questionnaires, pharmacy records and serum screening. Reprod Toxicol, 2019. 84: p. 93-97.

22. George, J., Patient deviations from treatment recommendations: does it matter what we call it? International Journal of Pharmacy Practice, 2010. 18(1): p. 3-5.
23. Sawicki, E., Stewart, K., et al., Medication use for chronic health conditions by pregnant women attending an Australian maternity hospital. Australian and New Zealand Journal of Obstetrics and Gynaecology, 2011. 51(4): p. 333-338.
24. van Gelder, M., Vorstenbosch, S., et al., Using Web-Based Questionnaires to Assess Medication Use During Pregnancy: A Validation Study in 2 Prospectively Enrolled Cohorts. Am J Epidemiol, 2018. 187(2): p. 326-336.
25. Lin, C.-Y., Broström, A., et al., Using extended theory of planned behavior to understand aspirin adherence in pregnant women. Pregnancy Hypertension, 2018. 12: p. 84-89.
26. The National Institute for Health and Clinical Excellence. The management of hypertensive disorders during pregnancy, in Clinical Guideline 107.
27. Norgard, B., Puho, E., et al., Aspirin use during early pregnancy and the risk of congenital abnormalities: a population-based case-control study. Am J Obstet Gynecol, 2005. 192(3): p. 922-3.
28. Roberge, S., Bujold, E., et al., Meta-analysis on the effect of aspirin use for prevention of preeclampsia on placental abruption and antepartum hemorrhage. Am J Obstet Gynecol, 2018. 218(5): p. 483-489.
29. Ahrens, K.A., Silver, R.M., et al., Complications and Safety of Preconception Low-Dose Aspirin Among Women With Prior Pregnancy Losses. Obstet Gynecol, 2016. 127(4): p. 689-98.

Supplementary file 1

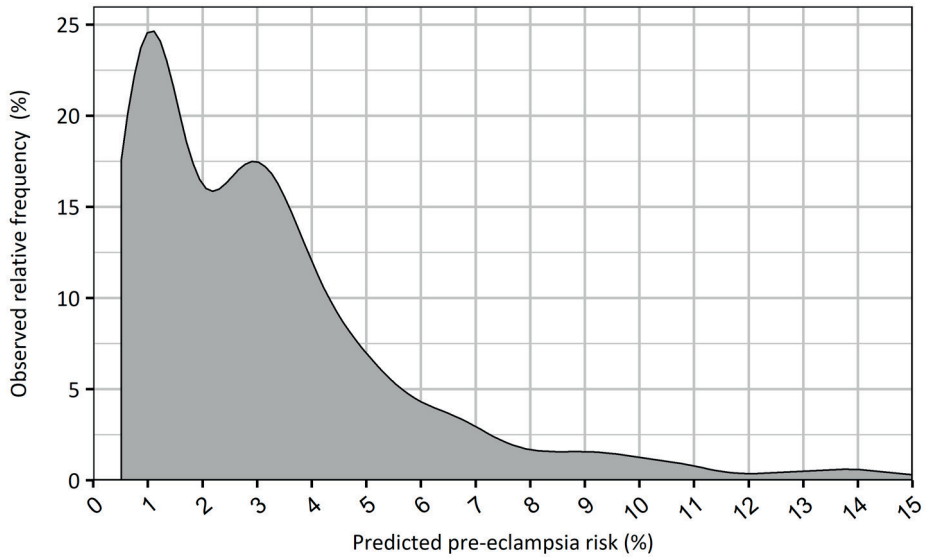


Figure S6.1 Distribution of observed pre-eclampsia risk estimates

Embargo as Requested

Chapter 7

Adherence rates to a prediction tool identifying women with an increased gestational diabetes risk: an implementation study

Pim van Montfort, Hubertina C.J. Scheepers, Ivo M.A. van Dooren, Linda J.E. Meertens, Laure Wynants, Maartje Zelis, Iris M. Zwaan, Marc E.A. Spaanderman, Luc J.M. Smits

Submitted

Chapter 8

Impact on perinatal health and cost-effectiveness of risk-based care in obstetrics: a before after study

Pim van Montfort, Hubertina C.J. Scheepers, Carmen D. Dirksen, Ivo M.A. van Dooren, Linda J.E Meertens, Sander M.J. van Kuijk, Ella J. Wijnen, Maartje Zelis, Iris M. Zwaan, Marc E.A. Spaanderman, Luc J.M. Smits

Submitted

Abstract

Background

Obstetric healthcare relies on an adequate antepartum risk selection. Most guidelines used for risk stratification, however, do not assess absolute risks. In 2017, a prediction tool was implemented in a Dutch region. This tool combines first trimester prediction models with obstetric care paths tailored to the individual risk profile, enabling risk-based care (RBC).

Objective

To assess impact and cost-effectiveness of RBC compared to care-as-usual (CAU) in a general population.

Methods

A before-after study was conducted using two multicenter prospective cohorts. The first cohort (2013-2015) received CAU, the second cohort (2017-2018) received RBC. Health outcomes were 1) a composite of adverse perinatal outcomes and 2) maternal quality adjusted life years (QALYs). Costs were estimated using a healthcare perspective from conception to six weeks after the due date. Mean costs per woman, cost differences between the two groups, as well as incremental cost effectiveness ratios were calculated. Sensitivity analyses were performed to evaluate the robustness of the findings.

Results

In total 3,425 women were included. In nulliparous women there was a significant reduction of perinatal adverse outcomes among the RBC group (aOR 0.56; 95%CI 0.32-0.94), but not in multiparous women. Mean costs per pregnant woman were significantly lower for RBC (mean difference -€2,766, 95%CI -€3,700 – -€1,825). No differences in maternal quality of life, adjusted for baseline health, were observed.

Conclusion

In the Netherlands, RBC in nulliparous women was associated with improved perinatal outcomes as compared to CAU. Furthermore, RBC was cost-effective compared to CAU and resulted in lower healthcare costs.

Introduction

In most developed countries, criteria lists are used to identify women with an increased risk of common adverse pregnancy outcomes (e.g. pre-eclampsia, gestational diabetes mellitus) [1-4]. In the Netherlands an obstetric indication checklist is used to allocate women to either primary care (autonomous midwives) or secondary care (obstetricians) [1]. However, like many other guidelines [2-4], this list is composed of a collection of single risk factors. It does not assess an individual woman's absolute risk and neither does it take a combination of factors into account. Moreover, this guideline does not describe the content of care, but merely indicates the recommended level of healthcare.

Prediction models, weighing several risk factors simultaneously, improve risk assessment of pre-eclampsia (PE) and gestational diabetes mellitus (GDM) (Meertens et al, in press; [5]). If these models are combined with care paths adjusted to the risk profile, obstetric care may transform to a more individual, risk-based approach. The Expect Study was designed to improve risk assessment in pregnant women and to implement clinically beneficial prediction models in daily obstetric practice [6,7]. A prediction tool was developed to facilitate implementation of risk-based care (RBC). This tool assesses women's risks during the first trimester upon PE, (GDM), fetal growth deviation, and spontaneous preterm birth (sPTB). The results of the risk assessment were combined with care paths tailored to the individual risks [7].

RBC comprises basic antenatal care for every woman and specific additional recommendations for women with an increased risk. Due to the different organizational model of RBC, healthcare resources are reallocated. Moreover, RBC is focused at early detection and prevention of pregnancy related complications, which could result in a reduction of complications. For example, in RBC, all women are recommended to assure an adequate calcium intake, which is correlated with a reduction of PE [8]. Furthermore, in case of an increased PE risk, women are counseled regarding low-dose aspirin (van Montfort et al, submitted). Aspirin may improve perinatal outcomes since it is correlated with a reduction of PE, SGA infancy, and sPTB in women at risk of PE [9,11]. Furthermore, screening and diagnosis of GDM improved in RBC (van Montfort et al, submitted), which is also correlated with a reduction of adverse perinatal outcomes [12,13].

Although, studies developing or validating prediction models may result in potentially useful prediction models, clinical impact of a prediction tool in daily practice may vastly differ from the results suggested by these studies. This could be due to, for example, differences in application, or due to an interplay of both healthcare professionals' and women's adherence to the recommendations provided [14].

This is one of the few studies implementing a prediction tool for obstetric care in daily clinical practice [15]. The aim of this study was to investigate the impact of RBC as compared to care-as-usual (CAU) on perinatal health and its cost-effectiveness. A before after analysis has been performed by comparing perinatal outcomes and costs of two successive multicenter prospective cohorts.

Methods

Recruitment and study design

To evaluate the impact of RBC we used a before-after design comprising two successive multiple prospective cohorts. Women enrolled in the validation study (Expect Study I, 2013-2015) received CAU. A subgroup of these women received additional questions related to the cost-effectiveness outcomes. All women enrolled in the implementation study (Expect Study II, 2017-2018) received RBC. All women participating in Expect Study II received the cost questionnaires.

A detailed study protocol for both cohorts has been published previously [6,7]. In short, for both cohorts all women ≥18 years old, with their first prenatal visit before sixteen weeks of pregnancy, were eligible for inclusion. Due to the small number of twin pregnancies (n=4) in the cost-effectiveness cohort of Expect Study I, inclusion for Expect Study II was limited to singleton pregnancies. Furthermore, to assure that all participants received RBC enrollment for Expect Study II was effectuated via the prediction tool (i.e. recruitment was only possible if the prediction tool was used). All hospitals of the region and the majority of autonomous midwifery practices recruited women for both cohorts.

Data collection

Data collection was similar for both cohorts. Women received four online surveys: at enrolment (1), at 24 weeks of pregnancy (2), at 34 weeks of pregnancy (3), and 6 weeks after the due date (4). Additionally, medical records and letters of discharge were retrieved and entered into a predesigned datasheet. For Expect Study II data retrieved by the prediction tool were logged as well.

Surveys two to four embedded the cost questionnaires. The recall periods in the cost questionnaires were approximately 24 weeks (conception – survey two), 10 weeks (survey two – survey three), and 12 weeks (survey three – postpartum survey). The questions covered every possible type of healthcare professional (e.g. general practitioner, midwife, and physiotherapist). A category 'other' was provided in case women felt their particular healthcare professional was not listed. Additional questions were asked to specify the type of contact (e.g. consult, phone call) along with corresponding frequencies.

Questions related to perinatal outcomes were incorporated in the post-partum survey. In case of discrepancies with the medical record, we contacted corresponding healthcare professionals for final decision. With respect to maternal QALYs, the EQ-5D-3L questionnaire was embedded in each survey allowing four time points for the QALY calculation [16].

Risk-based care

The Limburg Obstetric Consortium (LOC), responsible for the maternity care in the Southeastern part of the Netherlands, developed healthcare paths. These paths describe the content of obstetric care in detail for all women (basic care) and additional recommendations for those with an increased risk. The exact content of these care paths is listed in the Expect Study II protocol [7] and is summarized in supplementary Figure S8.1.

To implement RBC, an online prediction tool was developed and made available for all healthcare professionals of the region. The algorithms of the prediction models are provided in supplementary Table S8.1. This tool assesses the risks of PE, GDM, sPTB, and small- and large-for-gestational-age (SGA and LGA) infancy. It embeds the prediction models of

Syngelaki and Van Leeuwen for the risk assessment of PE and GDM respectively, externally validated and recalibrated by our group [5]. Risks of sPTB, SGA, and LGA were assessed with regional guidelines.

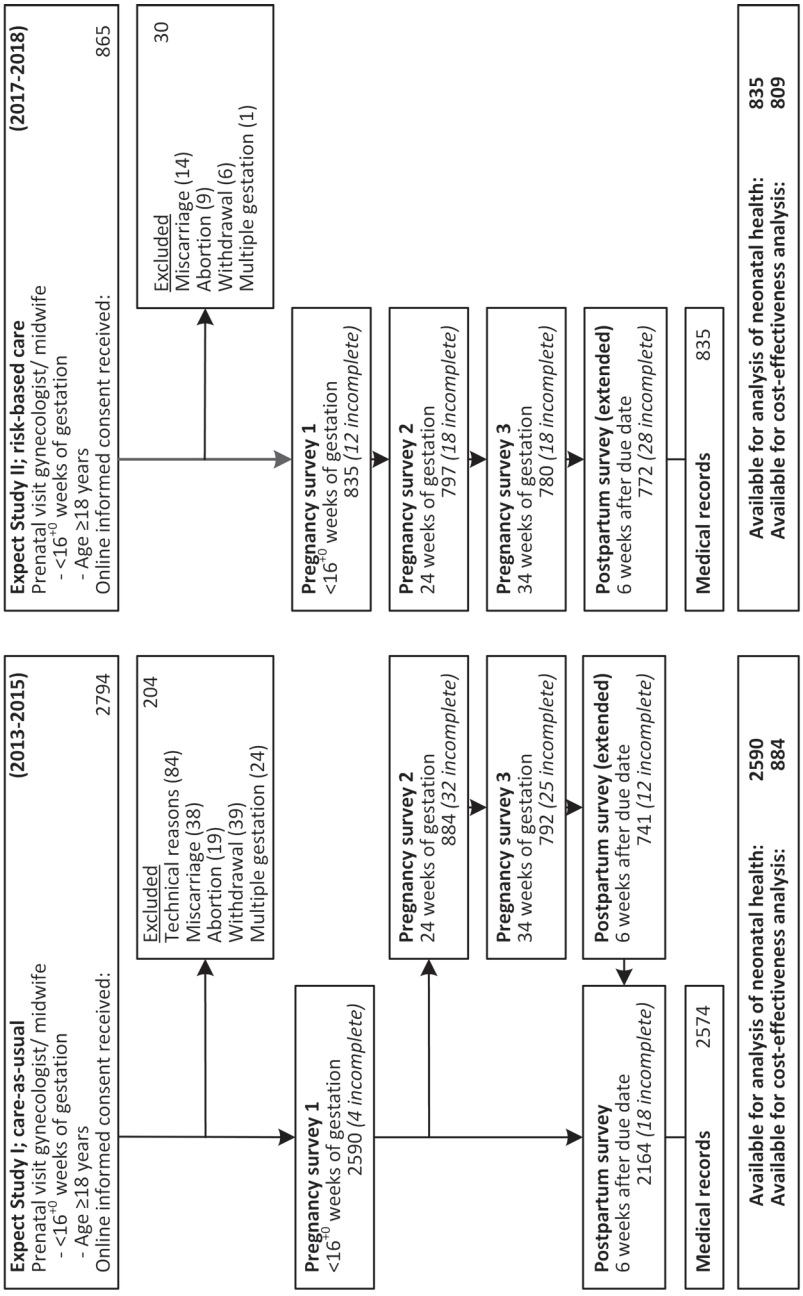


Figure 8.1 Inclusion and data collection of the Expect Study

Ethical approval

The Medical Ethical Committee of the Maastricht University Medical Centre declared that no ethical approval was necessary for Expect Study I and II (MEC-13-4-053 and MEC-17-4-057, respectively). All participating women gave informed consent.

Costs

Unit costs of healthcare resources were obtained from the Dutch manual for costing in health economic evaluations [17]. In case unit costs were unavailable, they were retrieved from the Dutch Healthcare Authority Tariffs, or a recently published Dutch cost-effectiveness study in obstetrics [18,19]. Costs of medication were retrieved from the Dutch Pharmacotherapeutic Register [20]. Using the Dutch Consumer Price Index all costs were expressed in Euros (2017 value) [21].

Perinatal health

To assess perinatal health, we prospectively defined a composite outcome [7]. The composite outcome consists of at least one of the following situations: stillbirth or neonatal death within seven days after birth, asphyxia (Apgar score <7 after 5 minutes), admission to a neonatal intensive care unit (NICU) within 28 days after birth, birthweight <2.3 weight percentile, and birth before 32 completed weeks of pregnancy. Birthweight percentiles were calculated using Dutch reference curves, corrected for gestational age, parity, fetal sex, and ethnicity [22]. Only the first survey (or, in case of RBC, data of the prediction tool) combined with either the postpartum survey or medical record were necessary to evaluate this outcome. For this reason, we used the data of all participants of both cohorts to assess the impact upon perinatal health.

Statistical analysis

The organization of maternity care (CAU vs. RBC) was used as independent variable in the logistic regression. To account for differences at baseline, we also performed a multiple logistic regression adjusting for: maternal baseline health utility (continuous), PE risk (continuous), GDM risk (continuous), obstetric history (nulliparous, multiparous with prior sPTB <34 weeks or with prior SGA infancy <10th percentile, multiparous without prior sPTB <34 weeks and without prior SGA infancy <10th percentile), level of healthcare received at recruitment (primary care vs. secondary care).

For the economic evaluation we used a healthcare perspective, comprising all healthcare services received by the woman or her child, over a time horizon of approximately eleven months (conception – six weeks after the due date). Women who did not complete any of the cost questionnaires (surveys 2-4), were excluded from the cost-effectiveness analysis. Missing data were imputed using stochastic regression imputation with predictive mean matching (average amount of missing data per variable was 5%) [23]. We compared the observed cohort and the imputed cohort by comparing the distribution of imputed variables. Two incremental cost-effectiveness ratios (ICERs) were calculated. The first ICER expresses the incremental costs per perinatal composite outcome prevented. Since the nature of the perinatal composite outcome is strongly correlated with neonatal admission, costs of neonatal admission are not taken into account for this ICER. For cost-effectiveness calculations, outcomes are usually coded so that the highest score represents the best

health outcome. Therefore, for this ICER, we converted the perinatal composite score: 1 corresponds with non-occurrence and 0 with occurrence of the outcome.
The second ICER expresses the incremental costs per incremental maternal Quality Adjusted Life Year (QALY). Health-related quality of life was evaluated by means of the standardized Euroqol EQ-5D-3L questionnaire using corresponding health utility scores based on the Dutch population [16,24].
To determine the 95% confidence interval (CI), we applied non-parametric bootstrapping using 10,000 replications with replacement from the original data and calculated the mean costs, effects and ICERs. Confidence intervals were obtained by calculating the bias-corrected and accelerated (BCa) bootstrap interval [25]. Uncertainty regarding these results was visualized by plotting the cost-effectiveness plane and the cost-effectiveness acceptability curve. All statistical analyses were performed using R statistical software version 3.6.0 [26].

Sensitivity analysis and subgroup analysis.

To analyze the influence of parity and level of healthcare at recruitment on both costs and health outcomes, we performed a subgroup analysis for nulliparous and multiparous women, and for women recruited in primary care and women recruited in secondary care.
For the first sensitivity analysis we used the Hoftiezer birthweight percentile curves. These new curves describe birthweight more accurately [27], but lack a 2.3rd percentile. Therefore, we used the 3rd percentile and adapted our perinatal composite for this analysis.
To examine the influence of differences between healthcare professionals recruiting women for the two cohorts, we performed a sensitivity analysis with data restricted to women enrolled by obstetric centers that recruited women for both cohorts.
To account for possible trends over time we applied a linear and a logistic regression to the CAU cohort for healthcare costs and the perinatal composite outcome, respectively. The duration of Expect Study I (days, continuous) was used as an independent variable while correcting for the same baseline characteristics as in our primary analysis.

Results

Data of 3,425 women were available for the analysis of the adverse perinatal outcome; 2590 women received CAU and 835 received RBC. For the economic evaluation, data of 1,693 women were available: 884 and 809 women receiving CAU and RBC, respectively. Figure 8.1 provides a flowchart of the participant enrollment. Baseline characteristics of both cohorts, as well as the cost-effectiveness sub-cohorts are tabulated in Table 8.1.
The cohorts did not substantially differ for the distributions of age, BMI, as well as the proportion of nulli- and multiparous women. The RBC cohort, however, contains a slightly larger proportion of women recruited in secondary care, compared to the CAU cohort. Additionally, women of the RBC cohort had a slightly lower health utility score at baseline, and relatively less often conceived naturally.

Table 8.1 Baseline characteristics of Expect Study cohort I and II

Baseline characteristics <16+0 weeks of gestation	CAU, all participants; n=2,590	CAU, available for cost-effectiveness analysis; n=884	RBC, all participants; n=835	RBC, available for cost-effectiveness analysis; n=809
Age, years; mean +/- sd	30.2 +/- 3.9	30.6 +/- 3.7	30.7 +/- 4.0	30.7 +/- 4.0
Ethnicity				
Caucasian; n (%)	2,509 (96.9)	872 (98.6)	817 (97.8)	791 (97.8)
Other; n (%)	81 (3.1)	12 (1.4)	18 (2.2)	18 (2.2)
Educational level				
Primary or secondary; n (%)	1,183 (45.7)	339 (38.3)	337 (40.8)	324 (40.4)
Tertiary level of education; n (%)	1,407 (54.3)	545 (61.7)	488 (59.2)	478 (59.6)
Body mass index, kg/m2; mean +/- sd	24.2 +/- 4.3	24.1 +/- 4.2	24.8 +/- 4.7	24.8 +/- 4.7
Smoking during pregnancy				
Yes	314 (12.1)	81 (9.2)	38 (4.6)	37 (4.6)
No	2,276 (87.9)	803 (90.8)	797 (95.4)	772 (95.4)
Medical history				
Pre-existent hypertension	27 (1.0)	18 (2.0)	16 (1.9)	16 (2.0)
Pre-existent diabetes mellitus	12 (0.5)	7 (0.8)	10 (1.2)	9 (1.1)
Health utility score; mean +/- sd	0.93 +/- 0.13	0.94 +/- 0.12	0.91 +/- 0.13	0.91 +/- 0.13
Conception				
Natural; n (%)	2,419 (93.4)	810 (91.6)	759 (90.9)	734 (90.7)
Ovulation induction; n (%)	92 (3.6)	41 (4.6)	36 (4.3)	36 (4.4)
In vitro fertilization; n (%)	79 (3.1)	33 (3.7)	40 (4.8)	39 (4.8)
Obstetric history				
Nulliparous; n (%)	1,315 (50.8)	448 (50.7)	421 (50.4)	409 (50.6)
Multiparous; n (%)	1,275 (49.2)	436 (49.3)	414 (49.6)	400 (49.4)
Prior PE; n (%)	72 (2.8)	31 (3.5)	50 (6.0)	48 (5.9)
Prior GDM; n (%)	14 (0.5)	5 (0.6)	19 (2.3)	19 (2.3)
Prior SGA; n (%)	110 (4.2)	42 (4.8)	44 (5.3)	43 (5.3)
Prior LGA; n (%)	168 (6.5)	59 (6.7)	44 (5.3)	42 (5.2)
Prior sPTB <34 weeks; n (%)	29 (1.1)	11 (1.2)	11 (1.3)	11 (1.4)
Risk assessment				
Increased PE risk; n (%)	965 (37.3)	349 (39.5)	359 (43.0)	350 (43.3)
Increased GDM risk; n (%)	1,394 (53.8)	478 (54.1)	408 (48.9)	400 (49.4)
Recruited in				
Primary care (midwife); n (%)	2,113 (81.6)	680 (76.9)	616 (73.8)	593 (73.3)
Secondary care (obstetrician); n (%)	477 (18.4)	204 (23.1)	219 (26.2)	216 (26.7)

CAU, care-as-usual; RBC, risk-based-care; sd, standard deviation; PE, pre-eclampsia; GDM, gestational diabetes mellitus; SGA, small-for-gestational-age infancy (<10th percentile); LGA, large-for-gestational-age infancy (>90th percentile); sPTB, spontaneous preterm birth; IQR, inter quartile range

Perinatal and maternal health outcomes

Table 8.2 displays the perinatal and maternal health outcomes. No statistically significant difference was observed regarding the adverse perinatal composite outcome between the RBC and CAU group (4.3% vs. 5.2% respectively). Taking differences at baseline into account, the adjusted odds ratio (aOR) was 0.76 (95%CI 0.51-1.11; Table 8.3). Subgroup analysis regarding parity, Table 8.4, revealed that for nulliparous women in RBC the risk of adverse perinatal outcomes was strongly and statistically significantly reduced (aOR 0.56; 95%CI 0.32-0.94), while no meaningful association showed in multiparous women (aOR 1.15; 95%CI 0.64-1.97).

Table 8.2 Health outcomes

Health outcomes	CAU, all participants; n=2,590	CAU, available for cost-effectiveness analysis; n=884	RBC, all participants; n=835	RBC, available for cost-effectiveness analysis; n=809
Neonatal				
Perinatal composite outcome	135 (5.2)	42 (4.8)	36 (4.3)	32 (4.0)
Birth <32 weeks	26 (1.0)	7 (0.8)	11 (1.3)	8 (1.0)
NICU admission	54 (2.1)	20 (2.3)	12 (1.4)	11 (1.4)
Birth percentile <2.3	48 (1.9)	9 (1.0)	11 (1.3)	11 (1.4)
APGAR <7 after 5 minutes	43 (1.7)	15 (1.7)	12 (1.4)	11 (1.4)
Stillbirth or neonatal death <7 days	14 (0.5)	4 (0.5)	7 (0.8)	4 (0.5)
Maternal				
Maternal QALYs	-	0.89 +/- 0.11	0.87 +/- 0.12	0.87 +/- 0.12
Health utility at baseline	0.93 +/- 0.13	0.94 +/- 0.12	0.91 +/- 0.13	0.91 +/- 0.13
Health utility at 24 weeks	-	0.85 +/- 0.17	-	0.84 +/- 0.16
Health utility at 34 weeks	-	0.81 +/- 0.18	-	0.79 +/- 0.18
Health utility postpartum	0.94 +/- 0.12	0.94 +/- 0.12	0.91 +/- 0.14	0.91 +/- 0.14

Data expressed as n (%) or mean +/- standard deviation. CAU, care-as-usual; RBC, risk-based-care; NICU, neonatal intensive care unit; QALY, quality adjusted life year

Table 8.3 Analysis of perinatal composite score

	No. of participants	No. with perinatal composite outcome n (%; 95%CI)	Odds ratio (95% CI)	P-value
All	3,425	171 (5.0; 4.3-5.8)	-	-
Unadjusted analysis				
Risk-based-care				
No (CAU)	2,590	135 (5.2; 4.4-6.1)	1 [Reference]	
Yes (RBC)	835	36 (4.3; 3.1-5.9)	0.82 (0.55-1.18)	0.30
Adjusted analysis				
Risk-based-care				
No (CAU)	2,590	135 (5.2; 4.4-6.1)	1 [Reference]	
Yes (RBC)	835	36 (4.3; 3.1-5.9)	0.76 (0.51-1.11)	0.17
Baseline health utility			0.99 (0.98-1.00)	0.09
Estimated PE risk			1.00 (0.95-1.04)	0.92
Estimated GDM risk			1.01 (0.98-1.03)	0.64
Obstetric history				
Nulliparous	1,736	105 (6.0; 5.0-7.3)	1 [Reference]	
Prior sPTB <34 weeks or SGA infancy	186	18 (9.7; 6.2-14.8)	1.50 (0.85-2.51)	0.14
No prior sPTB <34 weeks or SGA infancy	1503	48 (3.2; 2.4-4.2)	0.52 (0.36-0.74)	0.00
Recruited in				
Primary care (midwife)	2,729	120 (4.4; 3.7-5.2)	1 [Reference]	
Secondary care (obstetrician)	696	51 (7.3; 5.6-9.5)	1.61 (1.12-2.29)	0.01

CAU, care-as-usual; RBC, risk-based-care; PE, pre-eclampsia; GDM, gestational diabetes mellitus; sPTB, spontaneous preterm birth; SGA, small-for-gestational-age; CI, confidence interval

Table 8.4 Subgroup analysis of perinatal composite score in nulli- and multiparous women

	Nulliparous women			Multiparous women		
	No. of women	**No. with perinatal composite outcome n (%; 95%CI)**	**Odds ratio (95% CI)**	**No. of women**	**No. with perinatal composite outcome n (%; 95%CI)**	**Odds ratio (95% CI)**
All	1,736	105 (6.0; 5.0-7.3)	-	1,689	66 (3.9; 3.1-4.9)	-
Unadjusted analysis						
Risk-based-care						
No (CAU)	1,315	88 (6.7; 5.5-8.2)	1 [Reference]	1,275	47 (3.7; 2.8-4.9)	1 [Reference]
Yes (RBC)	421	17 (4.0; 2.5-6.4)	0.59 (0.33-0.97)	414	19 (4.6; 3.0-7.1)	1.26 (0.71-2.13)
Adjusted analysis						
Risk-based-care						
No (CAU)	1,315	88 (6.7; 5.5-8.2)	1 [Reference]	1,275	47 (3.7; 2.8-4.9)	1 [Reference]
Yes (RBC)	421	17 (4.0; 2.5-6.4)	0.56 (0.32-0.94)	414	19 (4.6; 3.0-7.1)	1.15 (0.64-1.97)
Baseline health utility			1.00 (0.98-1.01)			0.98 (0.97-1.00)
Estimated PE risk			1.00 (0.90-1.08)			1.00 (0.93-1.05)
Estimated GDM risk			1.01 (0.97-1.04)			1.00 (0.96-1.04)
Obstetric history						
Prior sPTB <34 weeks or SGA infancy	-	-	-	186	18 (9.7; 6.2-14.8)	2.79 (1.52-4.91)
No prior sPTB <34 weeks or SGA infancy	-	-	-	1,503	48 (3.2; 2.4-4.2)	1 [Reference]
Recruited in						
Primary care (midwife)	1,414	78 (5.5; 4.4-6.8)	1 [Reference]	1,315	42 (3.2; 2.4-4.3)	1 [Reference]
Secondary care (obstetrician)	322	27 (8.4; 5.8-11.9)	1.61 (0.99-2.57)	374	24 (6.4; 4.3-9.4)	1.63 (0.92-2.80)

CAU, care-as-usual; RBC, risk based care; PE, pre-eclampsia; GDM, gestational diabetes mellitus; sPTB, spontaneous preterm birth; SGA, small-for-gestational-age; CI, confidence interval

Maternal health utility scores at enrolment were high in both groups and declined slightly during pregnancy (Figure 8.2). A small, but statistically significant difference in maternal QALYs was observed. However, the difference was largely attributable to a lower health utility at baseline in the RBC group and effectively disappeared after adjustment for baseline health utility (adjusted β = -0.002, 95%CI -0.008; 0.004, p=0.54) [28].

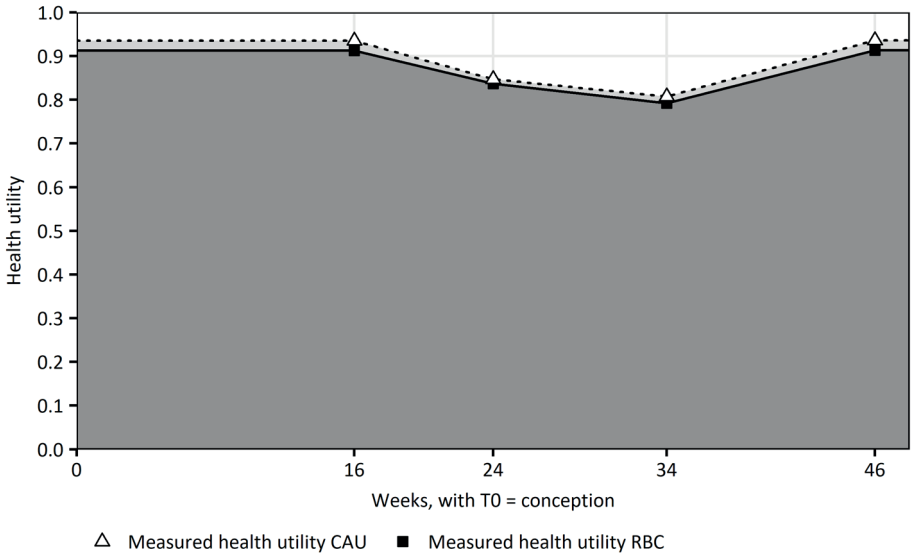


Figure 8.2 Health utility scores in care-as-usual and risk-based-care cohort. Area under curve represents the quality-adjusted-life-years

Costs and cost-effectiveness

Table 8.5 provides an overview of mean observed costs as well as the mean cost differences between RBC and CAU. Mean costs per pregnant woman were lower for RBC (mean difference -€2766; 95% BCa –€3703; –€1794). This difference was mainly driven by the difference in costs generated by maternal hospitalization and secondary care (healthcare services provided by obstetricians). With the exception of costs attributable to labour or alternative healthcare services, costs of all components were lower in RBC.

Results of bootstrapped data were plotted in a cost-effectiveness plane (Figure 8.4). Figure 8.3 shows the cost-effectiveness acceptability curve. Regarding the perinatal composite outcome, the ICER indicates that RBC dominates CAU, as costs are lower and perinatal outcomes are better for RBC. Regarding maternal QALYs, the ICER point estimate was €170,390. Furthermore, 95% of the bootstrapped QALY ICERs are in the quadrant where RBC is less costly but also slightly less effective. The probability that RBC was cost-effective compared to CAU ranged from 97-100%, assuming an ICER ceiling ratio from €10,000-€80,000 per QALY in accordance with the Dutch Health Insurance Board [29].

The subgroup analysis with respect to parity, supplementary Table S8.2, showed a discrepancy between nulli- and multiparous women regarding the ICER of the perinatal composite outcome. In nulliparous women the ICER indicates that RBC dominates CAU for nulliparous women, as costs and perinatal outcomes are better for RBC. For multiparous

women, the ICER was €203,402, since most bootstrapped ICERS are in the quadrant where RBC is less costly but also slightly less effective (see Figure 8.5 and supplementary Table S8.3).

Table 8.5 Costs per pregnant woman

Costs°	CAU; mean +/- sd	RBC; mean +/- sd	Mean difference* (CAU – RBC)
Total [95%CI]**	11,478 +/- 10,994	8,712 +/- 8,811	-2766 [-3700 – -1825]
Total, without neonatal admission [95%CI]**	8,969 +/- 8,687	6,562 +/- 7,290	-2406 [-3233 – -1719]
Primary care	835 +/- 481	813 +/- 459	-22
Midwifery	579 +/- 320	578 +/- 325	-1
Secondary care	1,176 +/- 1,507	658 +/- 919	-517
Gynecology	1,070 +/- 1,420	584 +/- 836	-486
Delivery	1,273 +/- 462	1,347 +/- 445	74
Hospitalization	2,828 +/- 5,447	1,468 +/- 2,980	-1360
Miscellaneous	746 +/- 435	562 +/- 333	-185
Diagnostics	659 +/- 310	517 +/- 292	-142
Medication	64 +/- 207	16 +/- 106	-48
Alternative healthcare	23 +/- 70	28 +/- 79	5
Maternity care	2,135 +/- 1,128	2,008 +/- 629	-127
Neonatal care	2,486 +/- 7,214	1,856 +/- 7,344	-630
Hospitalization	2,054 +/- 7,110	1,662 +/- 7,315	-392

CAU, care-as-usual; RBC, risk-based-care. All costs are expressed in 2017 Euro's.
*Costs of CAU cohort (n) minus costs of RBC cohort (n)
**For the mean difference, confidence interval based on bias-corrected and accelerated (BCa) bootstrap interval.
°Supplementary Table S8.2 provides a full overview of unit costs used for the economic evaluation

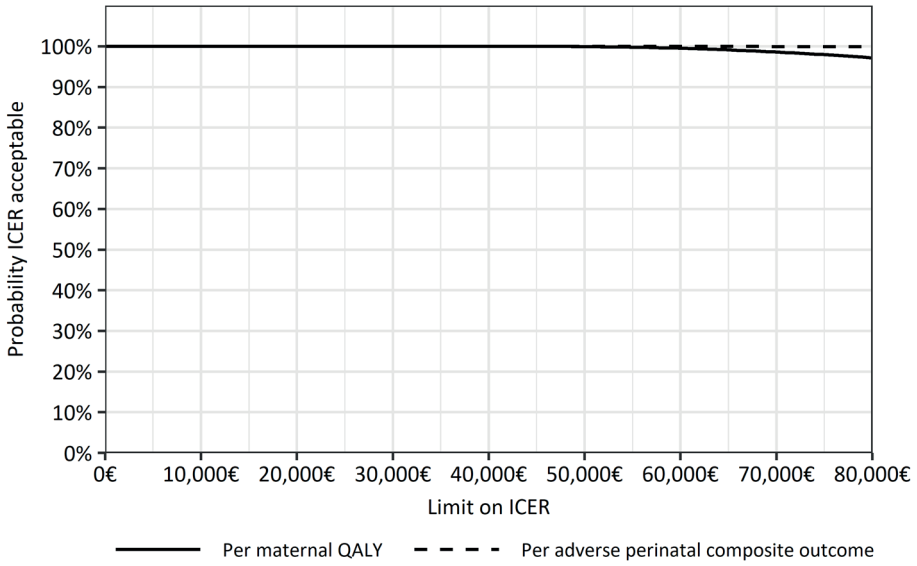


Figure 8.3 Cost-effectiveness acceptability curve for the incremental costs gained from a healthcare perspective per incremental maternal QALY or per incremental adverse perinatal composite outcome. QALY, quality-adjusted lifer year. ICER, incremental cost-effectiveness ratio

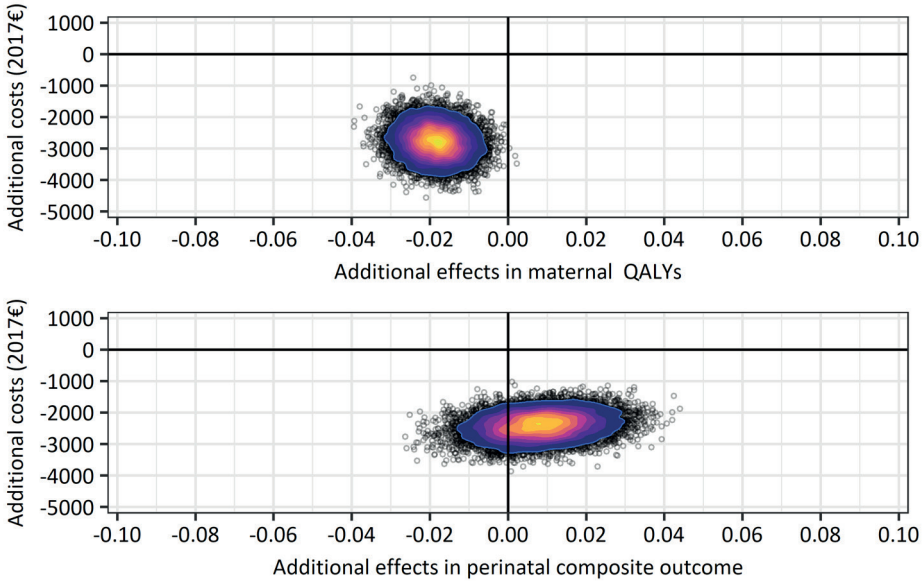


Figure 8.4 Cost-effectiveness density plane showing the incremental costs from a healthcare perspective (y-axis) and incremental effects (x-axis; maternal QALYs or prevented adverse perinatal composite outcome, top and bottom figure, respectively). Each data point represents one bootstrapped estimate of incremental costs and effects. QALY, quality-adjusted life year

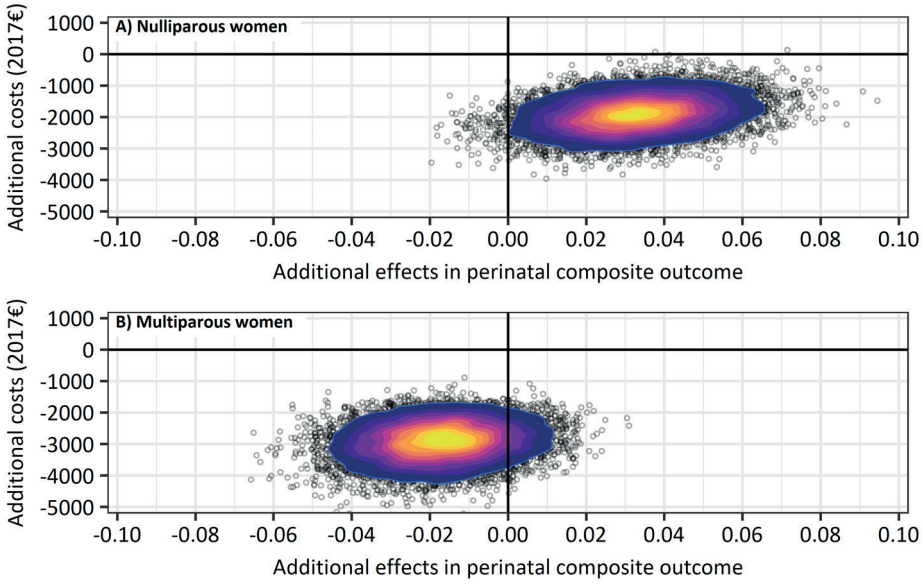


Figure 8.5 Cost-effectiveness density plane showing the incremental costs from a healthcare perspective (y-axis) and incremental effects (x-axis; prevented adverse perinatal composite outcome) in nulliparous (A) and multiparous (B) subgroups. Each data point represents one bootstrapped estimate of incremental costs and effects

Discussion

Main findings

For the population as a whole, the risk of an adverse perinatal outcome did not decrease statistically significant in RBC as compared to CAU. However, a statistically significant reduction of perinatal adverse outcomes was observed in nulliparous women (reduction: 44%; 95%CI 6%-68%), whereas in multiparous women no clear difference was observed. A small difference in QALYs was observed between women receiving RBC and women receiving CAU. This difference in maternal QALYs was no longer apparent when adjusting for health utility at enrolment. Furthermore, RBC resulted in lower costs and the ICERs indicate RBC was cost-effective as compared to CAU.

Strengths and limitations

Strengths of this study were its large sample size and its prospective, multicenter design. The use of multiple web-based surveys, a user-friendly method of data collection, provided high data quality by reducing potential recall biases and the numbers of missing data [30-32]. Furthermore, the diversity of participating midwifery centers, as well as hospitals, combined with the broad inclusion criteria, results in a low probability of selection bias. Nevertheless, the majority of enrolled women have a tertiary level of education and are of Caucasian (native) origin, which may have resulted in generally healthy women with above average health literacy skills [33].

Next to our primary outcome (a perinatal composite score) we used maternal QALYs as a secondary outcome. Ideally, the QALY calculation would take both maternal and perinatal outcomes into account. Yet, combining QALYs is challenging and literature describing how to achieve this is limited [34,35]. Furthermore, long-term outcomes should preferably be taken into account as well, but our study design only allowed for follow-up up to six weeks after the due date.

To ensure women received RBC, inclusion of the RBC cohort was achieved by our prediction tool. The prediction tool was developed for usage in the general population and was promoted as such [7]. All obstetric healthcare professionals of our region of interest committed themselves to provide RBC. Nevertheless, it could have been the case that for the RBC cohort women in particular were recruited by enthusiastic, above-averagely adherent healthcare professionals. On the other hand, the widespread use of our prediction tool, as well as the fact that most women receive obstetric care from multiple professionals during their pregnancy, minimize the possibility of this effect. Additionally, inclusion criteria for both cohorts were identical. Nevertheless, subtle differences are apparent at baseline between both cohorts. Per characteristic, differences were small, but together they yielded a less favorable risk profile among women in RBC as compared to CAU (e.g. lower health utility at baseline, a higher proportion recruited in secondary care, and a more often complicated obstetric history).

Although we adjusted for prognostic important baseline characteristics, residual confounding remains possible. Residual confounding may still result in women having a more untoward risk profile in the RBC group compared to the CAU group. However, this would rather result in an underestimation than an overestimation of the positive effects correlated with RBC. Moreover, we performed a sensitivity-analysis restricted to data of women enrolled by obstetric centers that recruited women for both cohorts. This did not yield substantially

different results and neither did the subgroup analysis to level of care received by women at enrolment (primary/ secondary care). This reduces the likelihood that our results are solely attributable to a difference in involved healthcare professionals (supplementary Table 8.2).
In essence, our study design used to assess the impact of RBC compared to CAU, is a 'before-after-analysis'. Theoretically, despite both cohorts succeeding each other in a relatively short time-span (~1.5 years), outcomes may have been affected by external trends over time (e.g. reduction of neonatal deaths due to improved healthcare, or a reduction in healthcare expenditures). However, the analyses taking into account the study period were not suggestive of a decreasing trend regarding the adverse perinatal composite outcome during the CAU cohort (aOR 1.02 95%CI 0.98-1.05). Neither did we find an association between the study period and the costs (adjusted β -2.0, 95%CI -8.0; 4.0, p=0.51). Moreover, nationwide statistics of Dutch health expenditures per capita suggest an increase rather than a decrease over time [36]. Therefore, we conclude that the cost reduction and improved perinatal outcomes in the RBC group are unlikely to be solely attributable to trends over time.

Interpretation

This is one of the few studies implementing a prediction tool for obstetric care in daily clinical practice [15]. To our knowledge, there are no other studies reporting an economic evaluation of obstetric care based on risk assessments provided by a prediction tool.
We found no differences in maternal QALYs between women receiving RBC and women receiving CAU after adjusting for health utility at enrolment. Overall, the measured health utilities of both groups were high and close to the perfect health state of '1'. This could be due to the fact that both cohorts represent a general, young population, with low proportions of women suffering from complications.
Our study indicates RBC is associated with a considerable cost reduction without a negative impact on maternal QALYs and improved perinatal outcomes in nulliparous women. In observational studies, like ours, interpretation of possible causal relationships should be done with caution. Moreover, in RBC usage of several preventive measurements improved (e.g. low-dose aspirin usage, GDM screening, (van Montfort et.al, submitted)). All these factors may have attributed to the improved outcomes in nulliparous women.
From a larger perspective, differences between RBC and CAU can be summarized by a different strategy assessing obstetric risks, combined with specific recommendations in case of an increased risk. Both the cost reduction as well as the improved perinatal outcomes may be attributable to the availability of clear instructions and standardizing care. Protocols, checklists, and triggers are known to improve health outcomes and efficiency [37-39]. The prediction tool may merely have worked as a triggering system regardless whether the risk assessment and usage of preventive measurements actually improved.
We found a significant reduction of adverse perinatal outcomes in nulliparous women. Interestingly, we did not find a similar beneficial effect in multiparous women. We hypothesize that the differences between RBC and CAU primarily affect obstetric care of nulliparous women due to differences in risk assessment. Prediction models take into account the weighted risk of multiple factors and possible inter-relations between them, allowing for a more personalized estimation of the absolute risk. However, in multiparous women, irrespective of the method to assess risks, risk assessment is strongly influenced by the obstetric history. In case of a complicated obstetric history (e.g. prior PE, prior GDM,

prior sPTB) healthcare professionals and pregnant women are probably already aware of any increased risks and their corresponding recommendations. For nulliparous women, the risk assessment may be less straightforward, as less information is available. Therefore, the improved risk assessment in RBC may be more pronounced in nulliparous women. This could have resulted that nulliparous women who would not have been identified with an increased risk with CAU, were identified as such with RBC. As a result, these nulliparous women may have received additional recommendations and (preventive) interventions relatively more often in RBC. Furthermore, both healthcare professionals' and pregnant women's awareness towards clinical symptoms of possible complications may be improved for these nulliparous women. This would particularly explain the reduction of the perinatal adverse composite outcome in nulliparous women.

Conclusion

RBC, as compared to CAU, resulted in a significant reduction of perinatal adverse outcomes in nulliparous women, but not in multiparous women. Apparently, in nulliparous women, transparent personalized risk estimations followed by tailored care may increase awareness amongst all involved. Moreover, RBC was cost-effective and resulted in lower costs without a negative impact on maternal health outcomes when adjusted for baseline health utility.

References

1. Commissie Verloskunde van het College voor zorgverzekeringen, Dutch obstetric indication list; Verloskundige indicatielijst (VIL), 2003.
2. NVOG [Dutch association for obstetrics and gynecology], Guideline Diabetes Mellitus and pregnancy (3.0). 2018.
3. National Institute for Health and Clinical Excellence, Diabetes in pregnancy: management from preconception to the postnatal period. Guideline NG3. 2015.
4. Euro-Peristat Project with SCPE and EUROCAT, European Perinatal Health Report. The Health and care of pregnant women and babies in Europe in 2010. 2013: p. 113.
5. Meertens, L.J.E., Scheepers, H.C.J., et al., External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study. Fetal Diagnosis and Therapy, 2019. 45(6): p. 381-393.
6. Meertens, L.J.E., Scheepers, H.C., et al., External Validation Study of First Trimester Obstetric Prediction Models (Expect Study I): Research Protocol and Population Characteristics. JMIR Res Protoc, 2017. 6(10): p. e203.
7. van Montfort, P., Willemse, J.P., et al., Implementation and Effects of Risk-Dependent Obstetric Care in the Netherlands (Expect Study II): Protocol for an Impact Study. JMIR Res Protoc, 2018. 7(5): p. e10066.
8. Hofmeyr, G.J., Lawrie, T.A., et al., Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. Cochrane Database Syst Rev, 2014(6): p. Cd001059.
9. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. Am J Obstet Gynecol, 2017. 216(2): p. 110-120.e6.
10. van Vliet, E.O., Askie, L.A., et al., Antiplatelet Agents and the Prevention of Spontaneous Preterm Birth: A Systematic Review and Meta-analysis. Obstet Gynecol, 2017. 129(2): p. 327-336.
11. Rolnik, D.L., Wright, D., et al., Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med, 2017. 377(7): p. 613-622.
12. Crowther, C.A., Hiller, J.E., et al., Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med, 2005. 352(24): p. 2477-86.
13. Landon, M.B., Spong, C.Y., et al., A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med, 2009. 361(14): p. 1339-48.
14. Kappen, T.H., van Klei, W.A., et al., Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagnostic and Prognostic Research, 2018. 2(1): p. 11.
15. Kleinrouweler, C.E., Cheong-See, F.M., et al., Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol, 2016. 214(1): p. 79-90.e36.
16. The EuroQol Group, EuroQol-a new facility for the measurement of health-related quality of life. Health policy, 1990. 16(3): p. 199-208.
17. Kanters, T.A., Bouwmans, C.A.M., et al., Update of the Dutch manual for costing studies in health care. PLoS One, 2017. 12(11): p. e0187477.
18. NZa [Dutch Health Authority], Dutch Health Authority Tariffs. 2017.
19. van Baaren, G.-J., Vis, J.Y., et al., Cost-effectiveness of diagnostic testing strategies including cervical-length measurement and fibronectin testing in women with symptoms of preterm labor. Ultrasound in Obstetrics & Gynecology, 2018. 51(5): p. 596-603.
20. Pharmacotherapeutic compass., 2018, Dutch Health Insurance Board; Zorginstituut Nederland.
21. CBS [Central Bureau of Statistics]. Dutch consumer price index. 2018 25-10-2018].
22. Visser, G.H., Eilers, P.H., et al., New Dutch reference curves for birthweight by gestational age. Early Hum Dev, 2009. 85(12): p. 737-44.
23. Van Buuren, S., Flexible imputation of missing data. 2012: CRC press.
24. Lamers, L., Stalmeier, P., et al., Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. Nederlands tijdschrift voor geneeskunde, 2005. 149(28): p. 1574-1578.

25. Barber, J.A. and Thompson, S.G., Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med, 2000. 19(23): p. 3219-36.
26. R Core Team, R: A Language and Environment for Statistical Computing, 2019, R Foundation for Statistical Computing: Vienna, Austria.
27. Hoftiezer, L., Hof, M.H.P., et al., From population reference to national standard: new and improved birthweight charts. Am J Obstet Gynecol, 2019. 220(4): p. 383.e1-383.e17.
28. Manca, A., Hawkins, N., et al., Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ, 2005. 14(5): p. 487-96.
29. Cost-effectiveness in practice; Kosteneffectiviteit in de praktijk, Dutch Health Insurance Board; Zorginstituut Nederland: Diemen.
30. van Gelder, M., Vorstenbosch, S., et al., Web-based questionnaires to assess perinatal outcome proved to be valid. J Clin Epidemiol, 2017. 90: p. 136-143.
31. van Gelder, M., Vorstenbosch, S., et al., Using Web-Based Questionnaires to Assess Medication Use During Pregnancy: A Validation Study in 2 Prospectively Enrolled Cohorts. Am J Epidemiol, 2018. 187(2): p. 326-336.
32. van Gelder, M.M., Schouten, N.P., et al., Using Web-Based Questionnaires and Obstetric Records to Assess General Health Characteristics Among Pregnant Women: A Validation Study. J Med Internet Res, 2015. 17(6): p. e149.
33. Berkman, N.D., Sheridan, S.L., et al., Low health literacy and health outcomes: an updated systematic review. Ann Intern Med, 2011. 155(2): p. 97-107.
34. Teune, M., Van Wassenaer, A., et al., Long-term child follow-up after large obstetric randomised controlled trials for the evaluation of perinatal interventions: a systematic review of the literature. BJOG: An International Journal of Obstetrics & Gynaecology, 2013. 120(1): p. 15-22.
35. Kreimeier, S. and Greiner, W., EQ-5D-Y as a Health-Related Quality of Life Instrument for Children and Adolescents: The Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. Value Health, 2019. 22(1): p. 31-37.
36. CBS [Central Bureau of Statistics]. National health expenditure. 2019 21-06-2019].
37. Arora, K.S., Shields, L.E., et al., Triggers, bundles, protocols, and checklists--what every maternal care provider needs to know. Am J Obstet Gynecol, 2016. 214(4): p. 444-451.
38. Fausett, M.B., Propst, A., et al., How to develop an effective obstetric checklist. Am J Obstet Gynecol, 2011. 205(3): p. 165-70.
39. Burgansky, A., Montalto, D., et al., The safe motherhood initiative: The development and implementation of standardized obstetric care bundles in New York. Semin Perinatol, 2016. 40(2): p. 124-31.
40. Meertens, L.J.E., Personalizing obstetric care, 2018, Maastricht University: Enschede.
41. Meertens, L.J.E., Scheepers, H.C.J., et al., External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study. Fetal Diagn Ther, 2018: p. 1-13.
42. van Leeuwen, M., Vijgen, S., et al., Costs and effects of screening for gestational diabetes mellitus with a prediction model based on patient characteristics, in Faculty of Medicine2012, University of Amsterdam: Amsterdam.

Supplementary Files

Supplementary Figure S8.1 Overview of risk-based care, adapted from [41]

Table S8.1 Algorithms of prediction models applied in risk-based care

Outcome	Original study	External validation study	Model algorithm after recalibration	AUC (95% CI)
Pre-eclamspia risk	Syngelaki 2011	Meertens 2018	Lp = -5.773 + 0.075 (BMI, kg/m2) + 0.022 (age, years) + 1.125 (if Afro-Caribbean) + 0.804 (if South Asian) + 0.526 (if East Asian) + 0.379 (if Mixed) + 0.289 (if ovulation drugs) + 0.598 (if IVF) - 0.233 (if smoker) + 1.519 (if history of chronic hypertension) + 0.643 (if type 1 diabetes mellitus) - 0.332 (if type 2 diabetes mellitus) - 1.329 (if parous, no history of pre-eclampsia) + 0.743 (if parous, history of pre-eclampsia) + 0.580 (if woman's mother had pre-eclampsia)	0.77 (0.72-0.81)
Gestational diabetes mellitus risk	Van Leeuwen 2010	Meertens 2018 (in press)	Lp = -6.28 + 0.83 (if non-Caucasian ethnicity) + 0.57 (if positive family history of DM) – 0.67 (if multipara without history of GDM) + 0.5 (if multipara with history of GDM) + 0.13 (BMI, kg/m2)	0.74 (0.70 0.79)

AUC, area under the curve; CI, confidence interval; Lp, linear predictor; BMI, body mass index; IVF, in vitro fertilization

Supplementary Table S8.2 Sensitivity analyses regarding perinatal health effects and cost-effectiveness of care-as-usual versus risk-based care.

	adverse perinatal composite outcome odds ratio (95% CI)*	Mean cost difference (95% BCa°)	Mean differential effects perinatal composite outcome (95% BCa°)	Mean differential maternal QALYs (95% BCa°)	ICER perinatal composite outcome €	ICER maternal QALYs €
Base case analysis						
All women	0.82 (0.55-1.18)	-2766 (-3700; -1825)	0.008 (-0.011; 0.028)	-0.018 (-0.029; -0.008)	Dominant	170,390
Nulliparous women	0.59 (0.33-0.97)	-3,386 (-4,814; -2,182)	0.033 (0.005; 0.063)	-0.019 (-0.034; -0.005)	Dominant	223,024
Multiparous women	1.26 (0.71-2.13)	-2,128 (-3,287; -533)	-0.017 (-0.033; -0.002)	-0.017 (-0.044; 0.007)	203,402	133,384
Adapted composite outcome (Hoftiezer birthweight curves with SGA <3rd percentile						
All women	0.73 (0.50; 1.04)	Equal to base case analysis	0,011 (-0.010; 0.032)	Equal to base case analysis	Dominant	Equal to base case analysis
Nulliparous women	0.59 (0.36; 0.93)		0.034 (0.001; 0.067)		Dominant	
Multiparous women	1.06 (0.57; 1.85)		-0.012 (-0.038; 0.011)		99,737	
Women enrolled by obstetric centers that recruited women for both cohorts						
All women	0.70 (0.46; 1.03)	-2,787 (-4,875; -2,172)	0.011 (-0.011; 0.034)	-0.020 (-0.032; -0.008)	Dominant	157,177
Nulliparous women	0.55 (0.31; 0.91)	-3,153 (-4,454; -1,926)	0.036 (0.004; 0.071)	-0.026 (-0.042; -0.009)	Dominant	156,720
Multiparous women	1.07 (0.57; 1.93)	-2,410 (-3,914; -588)	-0.013 (-0,0425; 0,0173)	-0.015 (-0,033; 0,003)	232,052	192,374
Women recruited from primary care						
All women	0.84 (0.54; 1.27)	-2,432 (-3,760; -1,699)	0.008 (-0.013; 0.028)	-0.023 (-0.035; -0.012)	Dominant	113,565
Women recruited from secondary care						
All women	0.64 (0.31; 1.22)	-4,727 (-7,311; -2,648)	0.015 (-0.032; 0.064)	0.005 (-0.020; 0.030)	Dominant	Dominant

*Odds ratios based on full cohorts; cost-effectiveness results based on cost-effectiveness sub cohorts.

°Confidence interval based on bias-corrected and accelerated (BCa) bootstrap interval.

Supplementary Table S8.3 Unit costs of healthcare resources

Item	Costs per unit (2017 €)	Source
Primary care		
Consultation (regular)	33.67	Dutch costing guideline [16]
Consultation (out of hours)	79.02	Dutch Health Authority Tariff [17]
Home visit (regular)	51.01	Dutch costing guideline [16]
Home visit (out of hours)	118.52	Dutch Health Authority Tariff [17]
Phone call (regular)	17.34	Dutch costing guideline [16]
Phone call (out of hours)	25.34	Dutch costing guideline [16]
Secondary care		
Consultation (regular)	92.85	Dutch costing guideline [16]
Consultation (out of hours)	264.25	Dutch costing guideline [16]
Phone call	17.34	Dutch costing guideline [16]
Maternity care		
Intake	65.78	Dutch costing guideline [16]
Maternity care (hour)	47.60	Dutch Health Authority Tariff [17]
Diagnostics		
Ultrasound (fetal dating)	44.37	Dutch Health Authority Tariff [17]
Counselling of screening for fetal abnormalities	44.22	Dutch Health Authority Tariff [17]
Ultrasound (fetal abnormalities screen)	167.17	Dutch Health Authority Tariff [17]
Ultrasound (fetal biometry)	36.99	Dutch Health Authority Tariff [17]
Oral glucose tolerance test	25.87	Van Leeuwen [40]
Laboratory testing, high-risk cases of pre-eclampsia°*	41.56	Dutch costing guideline, Dutch Health Authority Tariff [16,17]
Laboratory testing, pre-eclampsia diagnosed*	83.12	Dutch costing guideline, Dutch Health Authority Tariff [16,17]
Hospitalization		
Maternal		
General ward (day)	485.65	Dutch costing guideline [16]
Intensive care (day)	2,055.87	Dutch costing guideline [16]
Neonatal		
General ward (day)	639.72	Dutch costing guideline [16]
Neonatal intensive care unit (day)	1,664.30	Apostel I [18]
Delivery		
Home (vaginal, spontaneous)	536.76	Dutch Health Authority Tariff [17]
Birthing center (vaginal, spontaneous)	1,093.57	Dutch Health Authority Tariff [17]
Hospital (vaginal, spontaneous)	1,212,14	Dutch Health Authority Tariff [17]
Hospital (vaginal, instrumental)	1,431.85	Apostel I [18]
Hospital (cesarean)	2,137.69	Apostel I [18]
Medication**		
Tocolysis (treatment)*	55.33	Dutch Pharmacotherapeutic Register [19]
Corticosteroids (treatment)	25.73	Dutch Pharmacotherapeutic Register [19]
Magnesium sulfate (treatment)	16.01	Dutch Pharmacotherapeutic Register [19]

°For baseline values for women receiving RBC with an estimated pre-eclampsia risk ≥ 5.1%

*The mean of several methods is presented

**Costs of miscellaneous medication (e.g. antibiotics, antimycotics, anti-hypertensive drugs, antidepressants, antiemetics) are not shown

Chapter 9

General discussion

General discussion

Obstetric healthcare relies on an adequate antepartum risk selection. Risk selection in obstetric care is the process of quantifying and judging a woman's risk of an adverse pregnancy outcome. The methods used to identify women at increased risk of adverse outcomes varies greatly among countries. The common aim of the Expect Study and the Limburg Obstetric Consortium is to improve obstetric healthcare. In order to achieve this goal, the Expect Study focused at improving the risk selection of pregnant women, whereas the consortium focused at standardizing obstetric care and the development of healthcare pathways tailored to individual risk assessments. By combining prediction models with the risk-based care (RBC) pathways, healthcare professionals became able to perform individual risk assessments and discuss risk-based recommendations using a shared-decisional approach.

The Expect Study consists of two parts: a validation study (Expect Study I); and an implementation and impact study (Expect Study II). The validation study evaluated external validity of models for the prediction of pre-eclampsia (PE) [1], gestational diabetes mellitus (GDM) (Meertens et al., in press), fetal growth deviations [2], and spontaneous preterm birth (sPTB) during the first trimester [3]. To make implementation of any models feasible and suitable for the general population, the study was restricted to models using predictors that were non-invasive and easily obtained in Dutch obstetric practice (i.e. maternal characteristics, medical history). Expect Study II evaluated the implementation of risk-based care (RBC) and its impact on perinatal health outcomes, maternal quality of life, and healthcare costs.

The first part of this thesis reports on preparatory studies necessary to implement RBC and to facilitate the study of its impact. The second part reports on studies evaluating implementation and impact of a prediction tool in obstetric care, the Expect Calculator.

The current chapter gives an overview of the main findings, followed by a number of methodological considerations, clinical implications, and recommendations for future research.

Framework conditions for implementing risk-based obstetric healthcare

Prediction of spontaneous preterm birth

Our systematic search identified 2,018 articles, which resulted in five models predicting sPTB risks based on maternal characteristics. After excluding women with multiple pregnancies or iatrogenic preterm birth, data of 2,540 women were available for the external validation. In the general population, external validation showed poor to average discriminative performance of the models (area under curve 0.54 to 0.70). A subgroup analysis showed that the models discriminated poorly among nulliparous women (area under curve 0.51-0.56). Additionally, decision curve analyses indicated low clinical benefit, even for the best performing model. These results indicated that the prediction models were unable to adequately predict sPTB, or are at least unable to improve current clinical practice. Therefore, the Limburg Obstetric consortium decided that in the Expect Calculator, sPTB risk-assessment should not be performed by a prediction model, but remains to be based on a list of single risk factors [4].

Perinatal factors related to pregnancy and childbirth satisfaction

Most women receiving care-as-usual (CAU) were highly satisfied with the obstetric healthcare

services they received. However, satisfaction questionnaires generally result in high scores and some investigators have argued that dissatisfaction relies on a different construct [5,6]. For this reason, we focused on the less satisfied women to retrieve new insights that could improve obstetric care.

Our analyses indicated that antenatal anxiety, obstetrician-led care during labor and a decrease in perceived personal wellbeing were independently associated with satisfaction scores. No difference in satisfaction scores was found between antepartum care led by either a midwife or an obstetrician, but midwife-led antepartum care reduced the odds of reduced satisfaction compared to transfer of antenatal care.

The Expect Calculator

If a prediction model is to be used as a basis for clinical decision making, thresholds should be selected that indicate which risks are considered as increased [7]. Risk thresholds for the Expect calculator were determined by use of the ACCORD methodology [8]. Recommendations provided by the Expect Calculator are not normative, but are meant to trigger a process of counselling and shared-decision making. The Expect Calculator was introduced to all obstetric healthcare professionals of the region in 2017.

Implementation and impact of risk-based care

In total, 865 women were recruited for Expect Study II. Using multiple web-based surveys, these women were questioned regarding the shared decision making with their healthcare professional and the services they eventually received. Outcomes considered for the implementation and impact study were guideline adherence by caregivers, uptake of risk-based recommendations by pregnant women, as well as maternal quality of life, perinatal health outcomes and healthcare costs.

Adherence to guidelines and uptake of risk-based recommendations

Pre-eclampsia

Low-dose aspirin (LDA) was discussed with 81% of women with an increased PE risk, indicating adequate implementation by healthcare professionals. This rate tended to further increase over time during the study period. As compared to CAU, LDA usage vastly increased in RBC (RR 19.1; 95%CI 11.2-32.5). Yet, just 25% of the women with an increased PE risk in the RBC group reported daily LDA usage. Aspirin usage was positively correlated with both the predicted PE risk and women's concerns regarding development of PE. As a result, the LDA usage rate increased to a more acceptable level in high-risk women. Most important reasons for non-use were unawareness of LDA as preventive intervention, concerns of adverse effects, and doubts regarding the benefits.

Gestational diabetes mellitus

The majority of women (78%) reported their healthcare professional discussed their GDM-risk. Furthermore, an oral glucose tolerance test (OGTT) was performed within the recommended gestational window in 59% of women with an increased GDM risk estimation. Predicted GDM risks were positively correlated with the probability of performing an OGTT, resulting in high adherence rates among high-risk women. The majority of women who did not have an OGTT within the gestational window reported never having discussed an OGTT with their healthcare professional. Notably, a quarter of the women experienced discomfort

from the OGTT (Likert score 6-10, with 10 being extremely unpleasant).

Health outcomes and healthcare costs

To evaluate the impact on health outcomes and cost-effectiveness of RBC care as compared to CAU, we conducted before-after analyses. For these analyses, we used data of two successive multicenter prospective cohorts: Expect Study I (CAU group) and Expect Study II (RBC group). In total 3,425 women were included; 2590 women received CAU and 835 women received RBC.

After adjusting for health utility at baseline, we observed no differences in maternal quality of life between both groups. Overall, in RBC as compared to CAU, the risk of an adverse perinatal outcome did not decrease statistically significant (aOR 0.76; 95%CI 0.51-1.11). However, a statistically significant reduction was found among nulliparous women (aOR 0.56; 95%CI 0.32-0.94), whereas in multiparous women no clear difference was observed (aOR 1.15; 95%CI 0.64-1.97). Using a healthcare perspective, RBC was cost-effective and mean costs per woman were significantly lower for RBC compared to CAU (mean difference -€2,766, 95% CI -€3,700 – -€1,825).

Methodological considerations

In this paragraph, the most important methodological considerations of the research described in this thesis are discussed alongside with their potential influence upon the results.

Study population and data collection

Data of both the validation study (Expect Study I) as well as the impact study (Expect Study II) were used for the research described in this thesis. The diversity of participating midwifery centers, as well as hospitals, combined with the broad inclusion criteria should have ensured a population as unselected as possible. Nevertheless, women of Caucasian origin were overrepresented and the majority of women are well educated. Since impaired health literacy is correlated with non-adherence and impaired health outcomes [9,10], results in our study with respect to these outcomes may be somewhat overestimated. The use of multiple web-based surveys, a user-friendly method of data collection in today's digital era, provided high data quality by reducing potential recall biases and the numbers of missing data [11-13].

Recruitment of women was similar for both cohorts. However, to assure women participating in the impact study received RBC, only women for whom the Expect Calculator was used were eligible for inclusion. The Expect Calculator was developed for usage in the general population and was promoted as such. Still, this may have introduced a selection bias, since pro-active healthcare professionals may have been over-represented among the professionals who used our prediction tool. The intensive usage of the prediction tool throughout the region and the multitude of collaborating centers diminishes the potential influence of selection bias. Additionally, inclusion criteria for both cohorts were identical. Nevertheless, subtle differences were apparent at baseline between both cohorts. Per characteristic, differences were small, but together they yielded a less favorable risk profile among women in RBC as compared to CAU (e.g. lower health utility at baseline, a higher proportion recruited in secondary care, and a more often complicated obstetric history).

Although we adjusted for prognostic important baseline characteristics, residual confounding remains possible. Residual confounding may still result in women having a more untoward risk profile in the RBC group compared to the CAU group. However, this would rather result in an underestimation than an overestimation of the positive effects correlated with RBC.

External validation of sPTB prediction models

To our knowledge, this is the first systematic review of studies reporting non-invasive prediction models for the risk of sPTB. For the validation study, we enrolled 2,614 women receiving care-as-usual, of which 2,540 women were available for the external validation of sPTB. Although there is no golden rule available for the required sample size of external validations studies, a general rule of thumb is a minimum of 100 events (i.e. spontaneous preterm birth) [14,15]. An inadequate sample size decreases the precision of external validation measures [14,15]. Our sample included 118 women with a sPTB <37 weeks of gestation. Furthermore, the data were very complete with a maximum of only 1.2% of missing values. Our cohort might suffer from treatment bias to a small extent since we did not exclude women who had received treatment such as a cerclage or tocolysis. This may have resulted into the prevention of sPTB and thus an underestimation of model discrimination and calibration [16].

The outcome sPTB was obtained from a combination of the medical record and the postpartum survey. Combination of these two data sources, ensured for a reliable evaluation of the cause of preterm birth. In case of discrepancies, healthcare professionals were contacted.

Pregnancy and childbirth satisfaction

The usage of a multicenter prospective study design improved the probability of collecting a representative sample. Furthermore, it enabled optimal measurement of outcomes by minimizing recall bias and recording of all independent variables before completion of the patient satisfaction questionnaire.

To obtain a sufficient number of women in our analysis, we focused on women who experienced less than perfect healthcare. Our study does not have qualitative data regarding the level of satisfaction or dissatisfaction related to the obstetric healthcare services. However, the amount of studies using the validated pregnancy and childbirth questionnaire is limited and none of these used dissatisfaction as outcome [17,18]. Focusing on the less satisfied women may result into renewed insights that could improve obstetric care.

Assessing the usage of risk-based interventions

In this study usage of the risk-based interventions recommended by the Expect Calculator are mainly based on self-report. Women may have answered in a socially acceptable manner resulting in an overestimation of the usage rate [19]. However, women reporting non-use are likely to be telling the truth [19]. The potential overestimation of usage rates due to self-report is probably limited since all risk-based recommendations were subject to a shared decisional process.

Besides socially acceptable answers, self-report is also prone to recall problems. By using multiple surveys, strategically timed (e.g. shortly after the antenatal intake) and with relatively short intervals, the influence of recall problems was minimized.

With respect to the LDA recommendations, we were unable to reliably verify LDA usage with medical records or pharmacy registries since LDA is available over-the-counter in the Netherlands. Regarding the recommendations of an oral glucose tolerance test (OGTT), OGTT dates were retrieved from the medical record if women did not complete the postpartum survey, or when they did not recall the gestational age at the time of the OGTT. Given the nature of the OGTT (i.e. a specific appointment at a diagnostic center, for which women need to fast and drink a concentrated glucose solution), it is unlikely women would incorrectly recall whether they had undergone an OGTT.
We deliberately chose not to examine other interventions recommended to women with an increased PE or GDM risk (e.g. extra fetal biometry, extra blood pressure measurements). Women are possibly unable to distinguish between additional risk-based care and general basic care. With respect to the medical record, provided all interventions are registered reliably, it would be hard to determine whether the additional interventions were initiated as part of additional risk-dependent care, or due to other reasons (i.e. initiated due to clinical symptoms arisen during pregnancy). Therefore, we concluded usage rates of these interventions could not be determined reliably.

Evaluating impact and cost-effectiveness of risk-based care

In essence, our study design used to assess the impact of RBC compared to CAU, is a 'before after comparison'. Theoretically, despite both cohorts succeeding each other in a relatively short time-span (~1.5 years), outcomes may be affected by external trends over time (e.g. reduction of neonatal deaths due to improved obstetric care, or a reduction in healthcare expenditures). To detect such trends, we performed analyses taking into account the study period regarding healthcare related costs and perinatal health outcomes in the CAU cohort. These analyses did not point to a decreasing trend regarding the perinatal composite outcome. With respect to the costs, we did not find a trend over the study period. Moreover, nationwide statistics of Dutch health expenditures per capita suggest an increase rather than a decrease over time [20]. As a result, we conclude that the cost reduction and improved perinatal outcomes in the RBC group are unlikely to be attributable to trends over time.
Despite the fact that all healthcare professionals of the region committed themselves to RBC, it could have been the case that for the RBC cohort women in particular were recruited by enthusiastic, above-averagely adherent healthcare professionals. The widespread use of the Expect Calculator, on the other hand, as well as the fact that most women receive obstetric care from multiple professionals during their pregnancy, limit the possibility of this effect. Moreover, results did not essentially differ after restriction of the analysis to women enrolled by obstetric centres that recruited women for both cohorts. This reduces the likelihood of our results being influenced by a difference in healthcare professionals involved.

Clinical implications and future directions

The studies covered in this thesis provide useful insights into the clinical utility of a prediction tool in obstetric care 'outside the realm of research'. Prediction model development studies can provide us with potentially useful models and validation studies may improve our confidence in model's estimated discriminative performance. Nevertheless, an adequate discriminate performance does not guarantee a prediction model has a positive clinical

impact in daily practice. This could be due to several reasons, for example differences in application (e.g. using the model in a specific subgroup of women opposed to the general population), or due to an interplay of both healthcare professionals' and women's adherence to the recommendations provided [21]. The results described in this thesis may act as a starting point to improve the utilisation of the prediction tool and its recommendations, as well as implementation in other regions.

Spontaneous preterm birth and risk-based care

Unfortunately, our external validation of sPTB prediction models indicated that these models are unable to reliably predict the occurrence of sPTB. For this reason, assessing sPTB risk is still performed with the aid of lists of single risk factors. Currently, a large meta-analysis using individual patient data (IPD) from a large number of studies is being performed (using Expect Study data as well). Such a study has the advantage that results are more robust and that relevant subgroup analyses can be carried out such as preterm births in nulliparous women. Hopefully, the IPD study can help improve the prediction of sPTB risks and pave the way for better RBC with respect to preterm birth.

Pregnancy and childbirth satisfaction

In general, women were highly satisfied with the healthcare received during their pregnancy and childbirth period. Referral during antepartum care, which results in transfer from primary care to secondary care, was associated with suboptimal satisfaction. Furthermore, antenatal anxiety was experienced by 25% of all women and was associated with decreased satisfaction scores. Screening and treatment of women suffering from anxiety might improve pregnancy and childbirth satisfaction, but further research is necessary. Women's birthing experience may improve by reducing unnecessary secondary obstetric healthcare.

Utilization of risk-based recommendations

Despite the vast increase of preventive measurements used by women identified with in an increased risk, the potential clinical benefit of RBC is currently not fully utilized. The majority of women with an increased PE risk estimation reported their healthcare professional discussed the option of LDA. Yet, most women opted not to use LDA during their pregnancy due to concerns regarding the effectiveness or possible adverse effects. However, no serious adverse effects of LDA have been reported and it appears to be safe for the neonate, thus the risks of adverse effects likely outweigh the risks of harmful effects caused by PE [22-26]. Future qualitative research, for example with the aid of focus groups, is warranted to further explore women's decisional process and attitude regarding LDA usage. This will increase our insight how women weigh competing risks (i.e. PE-risk vs. risks upon adverse effects), whether the information currently offered is clear and sufficient, and how the shared decisional process may be improved. Such studies may provide us with suggestions how to increase the LDA usage rate among high-risk women.

There is also room for improvement regarding the utilisation of RBC in women with respect to GDM. The majority of women with an increased GDM risk stated their healthcare professional offered the option of an OGTT, and most women eventually had an OGTT in the recommended gestational window. Still, 58% of the women who did not had an OGTT within the recommend gestational window, reported the option of an OGTT was not discussed

with them. This indicates healthcare professionals likely fulfil a key role. Future qualitative research, exploring the reasons why healthcare professionals not always offer an OGTT in women with an increased GDM risk estimate, is therefore necessary. Moreover, the possible barriers responsible for a less than optimal adherence rate are likely to be of interest as well if a universal screening approach is considered.

Universal versus risk-based recommendations

For both the OGTT as well as LDA usage, recommendation of these preventive measurements to all women has been advocated as well [27,28]. A universal approach has the advantage that it simplifies the guidelines for healthcare professionals and the options for pregnant women. Moreover, such an approach would yield the highest clinical benefit at population level, because every prediction tool or guideline that targets specific risk groups will inevitably result in cases being missed (false-negatives), since they generally do not have a 100% sensitivity rate.

On the other hand, a universal approach may have several disadvantages. The results in this thesis indicate that both performance of an OGTT and LDA-usage were strongly correlated with the predicted risks of GDM and PE, respectively. This may suggest that in case of lower risk estimates, healthcare professionals and pregnant women deliberately chose not to use these preventive measurements. It is questionable, whether these women and their healthcare professionals would feel comfortable with a universal recommendation and would adhere to it.

A universal approach will increase the number women being recommended an OGTT or LDA enormously, especially low-risk women. At the same time, low-risk women are least likely to benefit from these preventive measurements. Additionally, a universal approach does not provide a specific argument for an individual woman. High-risk women may remain unaware of their risk, which deprives them of an extra argument compared to average-risk women. As a result, even though a universal approach may enhance the average adherence rate, it may result in reduced adherence rates among high-risk women when compared to a selective approach.

Another disadvantage, perhaps the most important one, is that universal recommendations bypass women's feelings and thoughts regarding these decisions. By using a prediction tool, absolute risks can be calculated which empowers women to make an informed decision together with their healthcare professional. It enables women to weigh the possible advantages and disadvantages for their individual situation. Moreover, previous reports indicated that decision tools and a shared decisional approach are likely to reduce women's anxiety 29, which, according to the research in this thesis, is correlated with patient satisfaction scores.

In general, universal approaches have been compared with a selective approach relying on an 'opt-in' strategy. For example, in case of the existence of any listed risk factors (e.g. BMI ≥30) an OGTT or LDA-usage is recommended [30,31]. This usually results in much more stringent strategies with a remarkably lower detection rate [32]. When a universal approach is considered, it should also be compared with a selective approach relying on an 'opt-out' strategy. Taking GDM as an example, this could mean recommending an OGTT to all pregnant women unless she meets specific exclusion criteria (i.e. multiparous women with an uncomplicated obstetric history). The likelihood these women develop GDM in a subsequent pregnancy is minimal [33], while such a strategy would reduce the amount of

OGTT's substantially. We suggest an alternative: to exclude women identified as low-risk by a prediction model with a high detection rate.
The trade-off between a universal approach (more true-positives, but also more false-negatives) versus a selective approach (more true-negatives, but also more false-positives), differs per topic (GDM risk and performing an OGTT vs. PE risk and using LDA) due to differences in advantages and disadvantages. Furthermore, it differs per country due to differences in the organization of obstetric care, but also because the trade-off depends on the incidence rate and thus a populations' a priori risk. Countries with an a priori high-risk population (e.g. due to a high obesity prevalence), potentially have more to gain with a universal approach. Eventually, the choice between a universal versus a selective approach is one that needs to be made by all stakeholders together (policymakers, obstetricians, midwives, and pregnant women).

Impact of risk-based care

Regardless of the approach preferred, it is important to utilize the full potential clinical benefit of the chosen approach. Our impact analysis indicated that RBC, being a more pro-active form of obstetric care, as compared to CAU may improve Dutch obstetric care. Although RBC did not lead to any clinically relevant difference in maternal quality of life or a statistically significant decrease in adverse perinatal outcomes, sub group analysis showed a clinically relevant and statistically significant reduction of the adverse perinatal composite outcome in nulliparous women (reduction: 44%; 95%CI 6%-68%).
The interpretation of the nature of the reduction of adverse perinatal outcomes is somewhat complicated since we used a composite outcome. Still, the clinical significance of this reduction is clear since all components of the composite outcome are important determinants of child mortality and morbidity. In the Netherlands, 44% of singletons are born to nulliparous women. Together, they give birth to roughly 71,000 children [34]. Assuming a 6% prevalence rate of the perinatal composite outcome, equal to the observed rate among nulliparous women in Expect Study I, a 44% reduction would mean that nationwide implementation of RBC in the Netherlands would prevent 1,874 newborns having an adverse perinatal composite outcome annually. Moreover, our cost-effectiveness analysis showed that RBC, as compared to CAU, is cost-effective and correlated with a substantial cost reduction. This, taken together with the improved perinatal outcomes in nulliparous women, implies that nationwide implementation of RBC is likely to make Dutch obstetric care cheaper and more effective.

Conclusions

In conclusion, the research presented in this thesis provide useful insights into the implementation and impact of a prediction-based first trimester decision tool in daily obstetric practice in the Netherlands. Soon after its introduction, obstetric care providers started using the tool and discussed estimated risks with a large majority of pregnant women, indicating adequate adherence. Furthermore, usage of preventive measurements strongly increased in comparison to previous care as usual, particularly in high-risk women. In comparison to CAU, RBC resulted in a significant reduction of perinatal adverse outcomes in nulliparous women, but not in multiparous women. Possibly, in nulliparous women, transparent personalized risk estimations followed by tailored care increased awareness

amongst all involved, resulting in better outcomes. Moreover, RBC was cost-effective and resulted in lower costs without a negative impact on maternal health outcomes when adjusted for baseline health utility. Besides, women's birthing experience may improve when risk-based care reduces unnecessary secondary obstetric healthcare.

Nevertheless, the potential clinical benefit of RBC is currently not fully utilized. Both LDA-usage as well as the performance of OGTTs remained suboptimal. Future qualitative research is necessary to identify factors that positively or negatively influence healthcare professionals' adherence and women's decisions regarding risk-based recommendations.

References

1. Meertens, L.J.E., Scheepers, H.C.J., et al., External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study. Fetal Diagnosis and Therapy, 2018.
2. Meertens, L., Smits, L., et al., External validation and clinical usefulness of first-trimester prediction models for small- and large-for-gestational-age infants: a prospective cohort study. Bjog, 2019. 126(4): p. 472-484.
3. Meertens, L.J.E., van Montfort, P., et al., Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta Obstetricia et Gynecologica Scandinavica, 2018. 97(8): p. 907-920.
4. Lemmens, S.M.P., Spaanderman, M.E.A., et al., Limburg Obstetric Quality System - Zorgpaden (aanbevelingen), 2016, LOQS: Maastricht.
5. Lee, A.V., Moriarty, J.P., et al., What can we learn from patient dissatisfaction? An analysis of dissatisfying events at an academic medical center. J Hosp Med, 2010. 5(9): p. 514-20.
6. Redshaw, M., Women as consumers of maternity care: measuring "satisfaction" or "dissatisfaction"?. Birth, 2008. 35(1): p. 73-76.
7. Reilly, B.M. and Evans, A.T., Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of internal medicine, 2006. 144(3): p. 201-209.
8. Lemmens, S.M., Agreement Conform Current Operational Rules and Directives (ACCORD): A Novel Tool to Reach Multidisciplinary Consensus. J Womens Health Gyn, 2019. 5: p. 1-11.
9. Lupattelli, A., Spigset, O., et al., Adherence to medication for chronic disorders during pregnancy: results from a multinational study. International Journal of Clinical Pharmacy, 2014. 36(1): p. 145-153.
10. Berkman, N.D., Sheridan, S.L., et al., Low health literacy and health outcomes: an updated systematic review. Ann Intern Med, 2011. 155(2): p. 97-107.
11. van Gelder, M., Vorstenbosch, S., et al., Web-based questionnaires to assess perinatal outcome proved to be valid. J Clin Epidemiol, 2017. 90: p. 136-143.
12. van Gelder, M., Vorstenbosch, S., et al., Using Web-Based Questionnaires to Assess Medication Use During Pregnancy: A Validation Study in 2 Prospectively Enrolled Cohorts. Am J Epidemiol, 2018. 187(2): p. 326-336.
13. van Gelder, M.M., Schouten, N.P., et al., Using Web-Based Questionnaires and Obstetric Records to Assess General Health Characteristics Among Pregnant Women: A Validation Study. J Med Internet Res, 2015. 17(6): p. e149.
14. Vergouwe, Y., Steyerberg, E.W., et al., Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. Journal of clinical epidemiology, 2005. 58(5): p. 475-483.
15. Collins, G.S., Ogundimu, E.O., et al., Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Statistics in medicine, 2016. 35(2): p. 214-226.
16. Pajouheshnia, R., Peelen, L.M., et al., Accounting for treatment use when validating a prognostic model: a simulation study. BMC Medical Research Methodology, 2017. 17(1): p. 103.
17. Truijens, S.E., Banga, F.R., et al., The Effect of Multiprofessional Simulation-Based Obstetric Team Training on Patient-Reported Quality of Care: A Pilot Study. Simul Healthc, 2015. 10(4): p. 210-6.
18. Truijens, S.E., Pommer, A.M., et al., 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: p. 35-40.
19. George, J., Patient deviations from treatment recommendations: does it matter what we call it? International Journal of Pharmacy Practice, 2010. 18(1): p. 3-5.
20. CBS [Central Bureau of Statistics], National health expenditure, 2019, Centraal Bureau voor de Statistiek: kerncijfers zorguitgaven.

21. Kappen, T.H., van Klei, W.A., et al., Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagnostic and Prognostic Research, 2018. 2(1): p. 11.
22. Norgard, B., Puho, E., et al., Aspirin use during early pregnancy and the risk of congenital abnormalities: a population-based case-control study. Am J Obstet Gynecol, 2005. 192(3): p. 922-3.
23. Roberge, S., Bujold, E., et al., Meta-analysis on the effect of aspirin use for prevention of preeclampsia on placental abruption and antepartum hemorrhage. Am J Obstet Gynecol, 2018. 218(5): p. 483-489.
24. Ahrens, K.A., Silver, R.M., et al., Complications and Safety of Preconception Low-Dose Aspirin Among Women With Prior Pregnancy Losses. Obstet Gynecol, 2016. 127(4): p. 689-98.
25. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. Am J Obstet Gynecol, 2017. 216(2): p. 110-120.e6.
26. Askie, L.M., Duley, L., et al., Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. Lancet, 2007. 369(9575): p. 1791-1798.
27. ACOG Practice Bulletin No. 190: Gestational Diabetes Mellitus. Obstet Gynecol, 2018. 131(2): p. e49-e64.
28. Hod, M., Pretty, M., et al., Joint position statement on universal screening for GDM in Europe by FIGO, EBCOG and EAPM. Eur J Obstet Gynecol Reprod Biol, 2018. 228: p. 329-330.
29. Vlemmix, F., Warendorf, J.K., et al., Decision aids to improve informed decision-making in pregnancy care: a systematic review. Bjog, 2013. 120(3): p. 257-66.
30. NVOG [Dutch association for obstetrics and gynecology], Guideline Diabetes Mellitus and pregnancy (3.0). 2018.
31. National Institute for Health and Clinical Excellence, Diabetes in pregnancy: management from preconception to the postnatal period. RCOG press, 2015. Guideline NG3.
32. Benhalima, K., Van Crombrugge, P., et al., Risk factor screening for gestational diabetes mellitus based on the 2013 WHO criteria. Eur J Endocrinol, 2019. 180(6): p. 353-363.
33. Young, C., Kuehl, T.J., et al., Gestational diabetes screening in subsequent pregnancies of previously healthy patients. Am J Obstet Gynecol, 2000. 182(5): p. 1024-6.
34. Perined, Perinatale Zorg in Nederland 2017, 2019: Utrecht.

ADDENDUM

Summary

Dutch obstetric care is divided into primary care provided by autonomous midwives and secondary care provided by obstetricians. In this system, risk selection plays a pivotal role. Nevertheless, the obstetric guideline used for the risk assessment, and thus the assignment of healthcare level, is merely a checklist of several single risk factors. This list does not assess women's absolute risk nor does it take a combination of multiple factors into account. Furthermore, Dutch obstetric care typically involves reacting to complications when they already exist or are imminent. Especially in obstetrics, prevention is better than cure, since therapeutic options are often limited due to the relatively short time window and sometimes potentially adverse effects for the foetus. A number of interventions has been shown to reduce the risk of adverse pregnancy outcomes, but most of these interventions are not suitable for all pregnant women.

In the past years, a number of prediction models have been developed that estimate the risk of pregnancy related complications including pre-eclampsia, gestational diabetes, preterm birth, and foetal growth deviations. Prediction models weigh several risk factors simultaneously and consider their possible interrelations, thereby potentially improving risk assessment. During Expect Study I, published prediction models relying on predictors readily available in Dutch obstetric care were externally validated. At the same time, the Limburg obstetric consortium focused at standardizing obstetric care by developing healthcare pathways tailored to individual risk assessments. Risk-based care (RBC) was designed by combining the results of the external validation study with the obstetric care pathways. Subsequently, the Expect Study II was performed, focusing on implementation and impact of RBC. The preparations and framework conditions necessary for the implementation of RBC are covered in the first part of this thesis. The second part of this thesis reports on the implementation and impact of RBC as compared to care-as-usual (CAU).

For Expect Study I, from 2013 to 2015, 2,614 pregnant women were enrolled in a multicenter prospective cohort in the Southeastern part of the Netherlands. These women received CAU, completed multiple web-based surveys, and allowed collection of their medical record. Of these women, 884 participated in a sub cohort by completing additional surveys. This sub cohort was used to evaluate healthcare related costs and patient satisfaction associated with CAU.

Results of the external validation of models predicting spontaneous preterm birth (sPTB) are covered in chapter 2 of this thesis. Five models were retrieved from the literature. Most studies suffered from a moderate to high risk of bias. Models' discriminative performance ranged from 0.54 to 0.70 in the general population, but was poor in a subgroup composed of nulliparous women (0.51-0.56). Decision curve analyses indicated low clinical benefit, even for the best performing model.

Chapter 3 evaluates women's satisfaction regarding the obstetric care services they received in the CAU situation. In general, women were highly satisfied. However, satisfaction questionnaires often result in high scores. For this reason, determinants related to sub optimal satisfaction scores were analyzed. Antenatal anxiety and antenatal transferal from healthcare level were both significantly related to reduced satisfaction scores. Moreover, antenatal anxiety was experienced by 25% of the pregnant women.

Chapter 4 covers the strategy used to implement RBC as well as the methods used to evaluate its impact as compared to CAU. Chapter 5 describes the methods used to decide

on the risk-threshold to discriminate between low and increased risk of PE. Participants in the decision process stressed that the threshold to be selected should be a starting point for a shared-decisional process regarding management of PE risk, rather than a compulsory ground for advising low-dose aspirin (LDA). As a result, an above-population-average PE risk was selected as threshold (>3.0%; sensitivity 75%, specificity 64%) to start discussing the option of using LDA with the pregnant woman. General adherence of care professionals to this recommendation was high: 81% of women identified with a PE risk >3.0% reported that the option of LDA usage was discussed with them.

Chapter 6 evaluates LDA usage-rates of pregnant women receiving RBC and compares it to the usage rates reported by women whom received CAU. LDA usage by women with an elevated risk increased strongly as compared to CAU (29.4% vs. 1.5%, RR 19.1; 95%CI 11.2-32.5). However, the general per protocol usage rate of LDA in RBC, 25%, remained moderate. In RBC, daily LDA usage was positively associated with both predicted PE risk and women's concerns regarding PE. Most reported reasons for non- or incomplete use were unawareness of LDA as a preventive intervention, concerns for potential adverse effects, and doubts regarding the benefits.

The consortium achieved consensus regarding a suitable GDM risk-threshold using a similar procedure as for the selection of a PE risk threshold. A predicted risk ≥3.5% was used as cut-off value to identify women at increased risk of GDM (sensitivity 80%, specificity 51%) and to discuss the option of an oral glucose tolerance test (OGTT) with these women. The adherence rate to risk-based GDM care is covered in chapter 7. Of all women, 78% reported their healthcare professional discussed their GDM-risk with them. In case of an increased risk, 59% of women received an OGTT within the recommended gestational window. Predicted GDM risks were positively correlated with the performance of an OGTT. The OGTT was experienced as uncomfortable by 25% of women who had an OGTT. Therefore, a selective screening strategy based on a prediction model with a high detection rate may be an interesting alternative to universal screening. Furthermore, a selective screening strategy relying on a prediction model enables women to make an informed decision together with their healthcare professional.

The impact of RBC on perinatal outcomes and healthcare related costs are described in chapter 8. Data of 3,425 women were available for the analysis of the adverse neonatal outcome; 2,590 women received CAU and 835 received RBC. No statistically significant difference was observed regarding the adverse neonatal composite outcome between the RBC and CAU group. However, subgroup analysis regarding parity showed a significant reduction of neonatal adverse outcomes among the RBC group in nulliparous women (aOR 0.56; 95%CI 0.32-0.94). We think that the differences between RBC and CAU mostly affect obstetric care for nulliparous women. For multiparous women, irrespective of care being RBC or CAU, health care professionals' judgment of risk is strongly influenced by the available information on obstetric history. For nulliparous women, the risk assessment may be less straightforward, as less information is available. As a result, improvement of the risk assessment would mainly effect these nulliparous women.

For the economic evaluation, data of 1,693 women were available: 884 and 809 women receiving CAU and RBC, respectively. Healthcare related costs per pregnant women were statistically significantly lower for RBC (mean difference -€2,766, 95%CI -€3,700 – -€1,825). Moreover, the incremental cost effectiveness ratios (ICERs) indicated RBC was highly cost-effective), while no differences in maternal quality of life, adjusted for baseline health, were

observed.
Chapter 9 discusses the evidence presented in this thesis. We conclude that RBC, as developed and implemented in our region, increases the usage of preventive measurements, but also that there remains room for improvement. We also conclude that RBC results in lower costs and, in nulliparous women, improves neonatal outcomes. Nationwide implementation of RBC is likely to have a positive impact on the obstetric care in the Netherlands. Future qualitative research is necessary to improve our insights regarding the shared decisional process between pregnant women and healthcare professionals, in order to improve usage rates of preventive measurements.

Samenvatting

Het Nederlandse verloskundige zorgsysteem is onderverdeeld in 1e lijns zorg (verloskundigen in zelfstandige praktijken) en 2e lijns zorg (gynaecologen in het ziekenhuis). In dit systeem speelt risicoselectie, het beoordelen van de zwangerschapsrisico's, een belangrijke rol. Immers, aan de hand van de verloskundige indicatielijst, de richtlijn die doorgaans wordt gebruikt voor de risicoselectie, wordt een zwangere vrouw al dan niet doorverwezen naar de 2e lijn. De verloskundige indicatielijst is echter slechts een checklist van losstaande risicofactoren. Middels deze lijst kan niet het absolute risico van een vrouw worden bepaald en ook is het niet mogelijk om een combinatie van factoren gelijktijdig te wegen in de daadwerkelijke risicoselectie.

De Nederlandse verloskundige zorg worden medische interventies doorgaans toegepast op het moment dat er complicaties (dreigen te) ontstaan. Juist in de verloskunde geldt echter het adagium van 'voorkomen is beter dan genezen'. In een zwangerschap worden de therapeutische mogelijkheden beperkt door mogelijke foetale bijwerkingen en het relatief korte tijdsbestek waarin een effect zou moeten optreden. Van een aantal interventies en maatregelen is bekend dat zij complicaties voorkomen of het risico daarop verkleinen. Het merendeel van deze interventies is echter niet geschikt om aan alle vrouwen aan te bieden. In de afgelopen jaren zijn diverse predictiemodellen ontwikkeld die het risico op zwangerschapscomplicaties voorspellen, bijvoorbeeld: pre-eclampsie, diabetes gravidarum, vroeggeboorte en afwijkende foetale groei. Predictiemodellen zijn in staat om meerdere risicofactoren simultaan te wegen en nemen daarbij ook eventuele onderlinge verbanden mee in de voorspelling. In de Expect Studie I zijn gepubliceerde predictiemodellen die gebruik maken van voorspellers die eenvoudig beschikbaar zijn in de Nederlandse verloskunde, extern gevalideerd. Tegelijkertijd heeft het Limburgs obstetrisch consortium zich gericht op het standaardiseren van de obstetrische zorg middels het ontwikkelen van risico zorgpaden. Door de resultaten van de externe validatie studie te combineren met de ontwikkelde zorgpaden ontstaat risico-gebaseerde zorg. Middels risico-gebaseerde zorg is het mogelijk om vrouwen met een verhoogd risico te counselen omtrent preventieve maatregelen die het risico verkleinen. De Expect Studie II richtte zich op de implementatie van risico-gebaseerde zorg en het meten van de impact daarvan.

De voorbereidende werkzaamheden alsmede de basiscondities die nodig waren voor het slagen van de implementatie zijn beschreven in het eerste deel van dit proefschrift. Het tweede deel van dit proefschrift beschrijft de implementatie en impact van risico-gebaseerde zorg ten opzichte van het voormalige verloskundige systeem (standaardzorg).

Gedurende 2013-2015 is ten behoeve van Expect Studie I in Limburg een multicenter prospectief cohort gevormd bestaande uit totaal 2.614 vrouwen. Deze vrouwen hebben allen de standaardzorg ontvangen tijdens hun zwangerschap. Verder hebben zij meerdere online vragenlijsten beantwoord en toegang tot hun medisch dossier verleend. Van deze groep heeft 884 vrouwen deelgenomen aan een subcohort door extra vragenlijsten te beantwoorden. Dit subcohort is gebruikt voor de evaluatie van zorgkosten en patiënttevredenheid voor de standaardzorg.

De resultaten van de externe validatie van modellen die spontane vroeggeboorte voorspellen staan beschreven in hoofdstuk 2 van dit proefschrift. In totaal werden vijf modellen geselecteerd uit de literatuur. De meeste studies van deze modellen hadden een redelijk tot hoog risico op vertekende resultaten. Het onderscheidende vermogen van de

modellen in de algemene populatie was matig tot redelijk (0,54-0,70), echter in nulliparae was dit vermogen lager (0,51-0,56). *Decision curve analysis* toonde aan dat de modellen waarschijnlijk niet in staat zijn de huidige klinische praktijk te verbeteren.
Hoofdstuk 3 evalueert de tevredenheid van zwangere vrouwen over de medische hulpverlening die zij tijdens de zwangerschap en de geboorte hebben mogen ontvangen. Over het algemeen waren vrouwen erg tevreden, maar tevredenheidsvragenlijsten binnen het verloskundige domein resulteren vaak in hoge scores. Om deze reden zijn de analyses gericht geweest op factoren die bijdragen tot een suboptimale tevredenheid. Antenatale angst alsook een antenatale overname van de zorg waren significant geassocieerd met verminderde tevredenheid. Antenatale angst werd door 25% van de zwangere vrouwen ervaren.
De strategie en methoden toegepast om risico-gebaseerde zorg te implementeren en de impact ten opzichte van de standaardzorg te evalueren zijn beschreven in hoofdstuk 4. In hoofdstuk 5 wordt beschreven welke methode is gehanteerd om een geschikt afkappunt te selecteren. Op basis van dit afkappunt wordt de mogelijkheid tot preventieve aspirine-inname besproken met de zwangere vrouw om zo het risico op pre-eclampsie te reduceren. Zorgverleners kwamen overeen om een afkappunt te kiezen waarbij laagdrempelig het gebruik van aspirine besproken zou worden. Er werd echter benadrukt dat dit afkappunt als startpunt dient voor de gezamenlijke besluitvorming omtrent preventief aspirine gebruik. Dit resulteerde in het feit dat het risico van de algemene populatie als grenswaarde is gekozen (grenswaarde >3,0%; sensitiviteit 75%, specificiteit 64%). De naleving van deze aanbeveling was over het algemeen hoog: 81% van de vrouwen met een pre-eclampsie risico >3,0% gaf aan dat de optie om aspirine in te nemen met hen besproken was.
In hoofdstuk 6 wordt geëvalueerd hoeveel vrouwen aspirine hebben gebruikt gedurende hun zwanger, daarbij wordt de risico-gebaseerde zorg met de standaardzorg vergeleken. Het aspirine gebruik nam tijdens de risico-gebaseerde zorg sterk toe in vergelijking tot de standaardzorg (29.4% vs. 1.5%, RR 19.1; 95%BI 11.2-32.5). Desondanks was ook gedurende de risico-gebaseerde zorg het percentage vrouwen dat conform de aanbevelingen aspirine gebruikte, met 25%, relatief laag. Het aspirine gebruik in de risico-gebaseerde zorg was positief gecorreleerd met het voorspelde pre-eclampsie risico als ook de mate van bezorgdheid van de vrouw omtrent pre-eclampsie. De meest genoemde redenen voor het niet innemen van aspirine waren onwetendheid over het preventieve effect, zorgen omtrent mogelijke bijwerkingen en twijfels over de voordelen.
Middels een vergelijkbare strategie als toegepast bij het pre-eclampsie model, bereikte het consortium ook consensus omtrent een afkappunt voor het diabetes gravidarum predictiemodel. Een geschat risico ≥3,5% (sensitiviteit 80%, specificiteit 51%) werd geselecteerd als drempelwaarde om vrouwen met een verhoogd diabetes gravidarum-risico op te sporen. Bij een risico ≥3,5% wordt middels gezamenlijke besluitvorming een keuze gemaakt om gedurende de zwangerschap een orale glucosetolerantie test (OGTT) uit te voeren. De naleving van deze aanbevelingen staan verslagen in hoofdstuk 7. Van alle vrouwen gaf 78% aan dat de zorgverlener het diabetes gravidarum-risico met hen had besproken, in geval van een verhoogd risico was bij 58% van de vrouwen de OGTT tijdig uitgevoerd. Het voorspelde diabetes gravidarum-risico was daarbij positief gecorreleerd met het tijdig uitvoeren van een OGTT. De OGTT werd door 25 van de vrouwen als een erg onaangename test ervaren. Mede om die reden is een selectieve screeningsprocedure gebaseerd op een predictiemodel met een hoge detectiegraad wellicht een interessant alternatief vergeleken

met een universele screeningsprocedure. Daarnaast biedt selectieve screening middels een predictiemodel zwangere vrouwen de mogelijkheid om samen met hun zorgverlener een weloverwogen besluit te nemen (zgn. *shared decision making*).

De impact van risico-gebaseerde zorg met betrekking tot perinatale uitkomsten en zorgkosten staat beschreven in hoofdstuk 8. Voor deze analyse werden gegevens van in totaal 3.425 vrouwen gebruikt, daarvan hebben 2.590 vrouwen de standaardzorg ontvangen en 835 risico-gebaseerde zorg. Er was geen statistisch significant verschil tussen beide groepen met betrekking tot de neonatale uitkomstmaat. Subgroepanalyses lieten echter een statistisch significante reductie van negatieve neonatale uitkomsten zien onder nulliparae (gecorrigeerde OR 0.56; 95%BI 0.32-0.94). Mogelijk hebben de verschillen tussen risico-gebaseerde zorg en standaardzorg met name een effect hebben op nulliparae. Bij multiparae wordt de boordeling van risico's sterk bepaald door informatie over de obstetrische voorgeschiedenis, ongeacht de methode van risicoselectie die men toepast. Voor nulliparae is de risicoselectie wellicht minder eenduidig, omdat er minder informatie beschikbaar is. Om die reden zal een mogelijke verbetering van de risicoselectie met name de zorg voor nulliparae beïnvloeden.

Voor de economische evaluatie waren de gegevens van 1.693 vrouwen beschikbaar: 884 vrouwen die standaardzorg ontvingen en 809 vrouwen die risico-gebaseerde zorg kregen. Zorgkosten per zwangere vrouw waren statistisch significant lager bij risico-gebaseerde zorg (gemiddelde verschil -€2.766, 95%BI -€3.700 – -€1.825). Na correctie voor de gezondheidsscore bij aanvang van de zwangerschap werd er geen verschil met betrekking tot maternale kwaliteit van leven waargenomen tussen beide groepen. Bovendien impliceerde de incrementele kosteneffectiviteit ratio's (ICERs) dat risico-gebaseerde zorg overduidelijk kosteneffectief was.

In hoofdstuk 9 wordt de onderzoeksresultaten gepresenteerd in dit proefschrift bediscussieerd. Wij concluderen dat risico-gebaseerde zorg, zoals ontwikkeld en geïmplementeerd in onze regio, leidt tot een toename in het toepassen van preventieve maatregelen. Er blijft echter ruimte voor verbetering. Verder concluderen wij dat risico-gebaseerde zorg leidt tot lagere zorgkosten en kosteneffectief is. Bovendien verbeteren de neonatale uitkomsten bij nulliparae. Landelijke implementatie van risico-gebaseerde zorg zal daarom zeer waarschijnlijk een positieve impact hebben op de Nederlandse obstetrische zorg. Toekomstig kwalitatief onderzoek is noodzakelijk om ons inzicht met betrekking tot het proces van gezamenlijke besluitvorming tussen de zwangere vrouw en haar zorgverlener te verbeteren. Op die manier worden mogelijk handvatten aangedragen om het gebruik van preventieve maatregelen verder te laten toenemen.

Valorisation

This chapter discusses the societal and economic relevance of this thesis. Valorisation has been defined by the Dutch National valorisation committee as *'the process of value creation from knowledge by making knowledge suitable for either economical or societal utilization and by translating knowledge into new products, services, processes, or business'* [1].

Relevance

The unique Dutch system with autonomous midwives providing primary care for pregnant women and obstetricians providing secondary, used to be an example of well-organized maternity care with low rates of medical intervantions [2]. However, this conservative approach underlying the Dutch system became subject of debate due to high perinatal mortality rates in the Netherlands as reported by two successive European perinatal health reports [3-5].

A system strictly divided into two separate levels of care, such as Dutch obstetric care, may suffer from disadvantages such as insufficient risk awareness and selection, discontinuity of care, and an increased risk of inaccurate communication [6]. Due to the European perinatal health reports, there was an increasing call for a reform of obstetric care into a system of integrated client-centered care with a more proactive approach [3,6]. In Limburg, the obstetric consortium, consisting of obstetric healthcare professionals representing the region, chose to achieve this by designing and implementing a risk-based care (RBC) approach: an obstetric healthcare system relying on an individual risk assessment with basic care pathways for low-risk women and additional recommendations for women identified with an increased risk for pregnancy related complications. Furthermore, RBC pathways might stimulate integrated care by intensifying the collaboration between autonomous midwives and gynecologists.

The majority of perinatal deaths in the Netherlands are related to either asphyxia, preterm birth (PTB), small-for-gestational-age infancy (SGA), or congenital anomalies [7]. Hypertensive disorders in pregnancy, such as pre-eclampsia (PE), are strongly associated with SGA and PTB [8]. Gestational diabetes mellitus (GDM) increases the risk of with birth injuries and asphyxia [9,10]. Therefore, preventing these adverse outcomes could eventually lead to a reduction of perinatal mortality.

A number of interventions have shown to be effective in the prevention of adverse pregnancy outcomes, such as low-dose aspirin treatment in case of PE [11-13], adequate management of GDM [14,15], and progesterone administration in women at risk of spontaneous PTB [16]. Most of the interventions, however, are not suitable for all pregnant women, because of either possible adverse effects, patient burden, or costs. Consequently, healthcare professionals need a risk assessment in order to decide which women may, on average, benefit most from such preventive measurements.

In care-as-usual (CAU), the Dutch obstetric indication list is used to check whether there is a predefined risk factor present, or a complication during pregnancy that warrants transfer from primary to secondary care [17]. However, this list does not assess an individual woman's absolute risk and is unable to take a combination of factors into account simultaneously. Furthermore, it does not describe the contents of obstetric care that should be offered.

The Expect Study was designed to improve the risk selection of pregnant women and consists of two parts [11,18]. Expect Study I was aimed at the external validation of in total 39 non-invasive prediction models predicting important pregnancy related complications. Expect Study II, which is reported on in this thesis, focused on the implementation and evaluation

of RBC using prediction models, combined with obstetric healthcare paths tailored to the individual risk assessments [18].

Expect Calculator

To implement RBC we designed an online prediction tool, the Expect Calculator. This tool combines the selected prediction models, risk-thresholds, and care paths to enable RBC. Risk assessment of pre-eclampsia and gestational diabetes mellitus is performed with the aid of externally validated prediction models. Risks of spontaneous preterm birth and fetal growth deviations are assessed with regional guidelines which were provided by the Limburg obstetric consortium.
To facilitate the shared decisional approach, the results of the risk assessment are visualized at a linear scale. Moreover, the tool automatically provides patient information brochures tailored to the results of the individual risk assessment. As shown in Figure Add.1, displaying the number of risk assessments made per month, the Expect Calculator was increasingly and intensively used. Although the Expect Calculator was specifically developed for healthcare professionals of Limburg, it can be easily used by any obstetric healthcare professional.

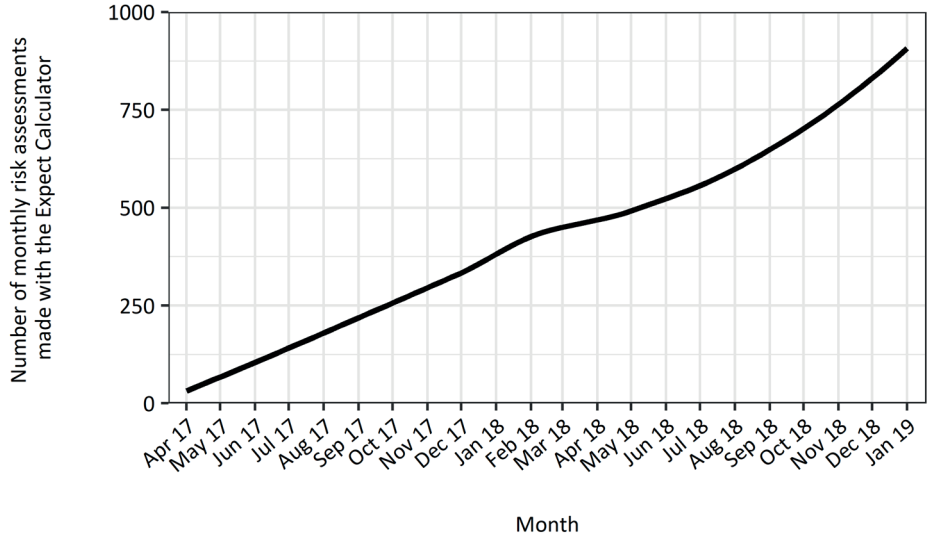


Figure Add.1 Number of risk assessments performed by the Expect Calculator per month

Societal impact

Merely explaining the societal relevance of improving obstetric healthcare would quickly result in stating the obvious. However, when accompanied with some statistics it may be easier to realize the potential societal impact of it. RBC focuses on early detection and prevention of pregnancy related complications with the aid of prediction models. The studies in this thesis indicate that implementation of RBC resulted in an increased usage of preventive measurements and a reduction of neonatal adverse outcomes. The following paragraphs discuss the potential societal impact of these improvements in Dutch obstetric care.

Roughly 8,500 women give birth in Limburg annually [19]. Applying the incidence rate of PE, approximately 3% [8], to this number, means that every year 255 women in Limburg suffer from PE. Fortunately, the majority of the women PE will be manifest term or near-term. In these cases labor will be induced and often further adverse events are either prevented or remain manageable. However, for a minority PE truly becomes a life threatening disease, either for the mother or for the neonate [20]. Preterm PE, especially extremely preterm PE, frequently results in preterm birth and is often combined with low birthweight and prolonged hospitalization of the mother and the neonate. Although the management of PE has improved, a cure that would preserve the pregnancy and thereby diminishing the sequelea accompanied with preterm birth remains unavailable. Therefore, preventive measures play a pivotal role in decreasing the burden of PE [8]. The absolute reduction of PE depends upon the combined effectiveness of low-dose aspirin prophylaxis and adequate calcium intake. A recent meta-analysis examining the effectivity of aspirin solely, indicated a relative risk of 0.56 (95% confidence interval, 0.43-0.75) if aspirin was initiated at ≤ 16 weeks of gestation [21]. This would mean that, if all women at risk were identified and used low-dose-aspirin as recommended, 112 of the 255 annual cases could be prevented.
In this thesis, a composite outcome was used for the evaluation of the neonatal outcome. Interpreting a composite outcome may be somewhat complicated. Still, the relevance of the reduction of this outcome is clear since all components of the composite outcome are important determinants of child mortality and morbidity. The results of chapter 8 in this thesis indicate that RBC was associated with a 44% reduction of the adverse neonatal composite outcome in nulliparous women. In the Netherlands, excluding multiple pregnancies, roughly 161,000 children are born annually. Of these children, 71,000 (44%) are born to nulliparous women [19]. Assuming a 6% prevalence rate of the adverse neonatal composite outcome, the observed prevalence rate observed in nulliparous women receiving CAU, means 4.260 neonates in the Netherlands suffer from such an adverse outcome. Applying the 44% reduction rate as indicated by the analyses of chapter 8, would mean nationwide implementation of RBC in the Netherlands could prevent 1,874 new-borns having an adverse outcome. This number is equal to the number of children of roughly seven averagely sized elementary schools.

Economic impact

Ideally, decisions regarding recommendations and preferred follow-up in general are primarily based on clinical arguments. However, the potential costs associated with provided healthcare services cannot be neglected as resources, be it healthcare costs or trained staff, are not infinite. The Dutch government, as most governments of developed countries, struggles with increasing healthcare expenditures that threaten the sustainability of the healthcare system. When healthcare expenditure remains to increase at the same speed as it did during 2006-2016, a household would spent half of its income on healthcare by 2040 [22]. As a result, reformation of a healthcare system should be accompanied with an economic evaluation. This evaluation should firstly answer whether the reform results into increased healthcare costs. If so, the next question is whether the reform is cost-effective, or in other words, whether the degree of improved outcomes justify the increased costs. Chapter 8 of this thesis describes the economic evaluation of RBC in detail. The results indicate that RBC is cost-effective and result in a substantial direct cost reduction of approximately €2,700 per pregnant woman. This would mean that nationwide implementation, taking into account

163.826 pregnant women [19], may result in a cost saving of 442 million euro per year.

Future implications

The studies in this thesis provide useful insights regarding the potential impact of RBC relying on a prediction tool that enables an individual risk assessment. Before a prediction model can be put to practice thresholds should be selected that indicate which risks are considered as increased [23]. This thesis covers how the obstetric consortium of Limburg handled this process and tried to incorporate all stakeholders. Although there are many different strategies imaginable to accomplish the implementation of a prediction model into daily practice, our study design may serve as an example for others.

Dissemination of guidelines or stating recommendations does not automatically result in adherence by healthcare professionals. Implementation of effective preventive interventions often suffers from low adherence rates [24-26]. The research in this thesis gives a first glance of the uptake of recommendations that emerged from an individual risk assessment provided by a prediction tool. Furthermore, a first insight of potential barriers that may hamper the uptake is provided.

By using a qualitative study design, the potential barriers and opportunities involved in this process could be evaluated in depth. Such a study, with the aid of focus groups, is currently performed regarding the recommendations of adequate calcium intake during pregnancy. Further research like this, will be necessary to improve the utilization of recommended preventive measures. This would increase our insight how to optimize the implementation of RBC, increase the uptake of preventive interventions, and how RBC could be implemented best in other regions.

References

1. Drooge, L.v., Vandeberg, R., et al., Waardevol: Indicatoren voor Valorisatie, 2011, Rathenau Instituut: Den Haag.
2. De Vries, R., Nieuwenhuijze, M., et al., What does it take to have a strong and independent profession of midwifery? Lessons from the Netherlands. Midwifery, 2013. 29(10): p. 1122-1128.
3. Visser, G.H.A., Obstetric care in the Netherlands: relic or example? J Obstet Gynaecol Can, 2012. 34(10): p. 971-975.
4. Euro-Peristat Project with SCPE and EUROCAT, European Perinatal Health Report. The Health and care of pregnant women and babies in Europe in 2010. 2013: p. 113.
5. Euro-Peristat project with SCPE EUROCAT EURONEOSTAT. European Perinatal Health Report 2004. 2008.
6. Perdok, H., Jans, S., et al., Opinions of maternity care professionals and other stakeholders about integration of maternity care: a qualitative study in the Netherlands. BMC Pregnancy and Childbirth, 2016. 16(1): p. 188.
7. Bonsel, G., Birnie, E., et al., Lijnen in de perinatale sterfte, Signalementstudie Zwangerschap en Geboorte 2010. 2010, Rotterdam: Erasmus MC.
8. Koullali, B., Oudijk, M.A., et al., Risk assessment and management to prevent preterm birth. Semin Fetal Neonatal Med, 2016. 21(2): p. 80-8.
9. Meertens, L., Smits, L., et al., External validation and clinical usefulness of first-trimester prediction models for small- and large-for-gestational-age infants: a prospective cohort study. Bjog, 2019. 126(4): p. 472-484.
10. Henriksen, T., The macrosomic fetus: a challenge in current obstetrics. Acta Obstet Gynecol Scand, 2008. 87(2): p. 134-45.
11. Meertens, L.J.E., Scheepers, H.C.J., et al., Should women be advised to use calcium supplements during pregnancy? A decision analysis. Matern Child Nutr, 2017.
12. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. American journal of obstetrics and gynecology, 2017. 216(2): p. 110-120. e6.
13. Rolnik, D.L., Wright, D., et al., Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med, 2017.
14. HAPO Study Cooperative Research Group, Hyperglycemia and adverse pregnancy outcomes. N Engl j Med, 2008. 2008(358): p. 1991-2002.
15. Poolsup, N., Suksomboon, N., et al., Effect of treatment of gestational diabetes mellitus: a systematic review and meta-analysis. PLoS One, 2014. 9(3): p. e92485.
16. Dodd, J.M., Jones, L., et al., Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. Cochrane Database Syst Rev, 2013(7): p. Cd004947.
17. Commissie Verloskunde van het College voor zorgverzekeringen, verloskundig vademecum - Verloskundige indicatielijst (VIL), 2003.
18. van Montfort, P., Willemse, J.P., et al., Implementation and Effects of Risk-Dependent Obstetric Care in the Netherlands (Expect Study II): Protocol for an Impact Study. JMIR Res Protoc, 2018. 7(5): p. e10066.
19. Perined, Perinatale Zorg in Nederland 2017, 2019: Utrecht.
20. Mol, B.W.J., Roberts, C.T., et al., Pre-eclampsia. Lancet, 2016. 387(10022): p. 999-1011.
21. Roberge, S., Nicolaides, K., et al., The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. Am J Obstet Gynecol, 2017. 216(2): p. 110-120.e6.
22. Rapport 'De zorg: hoeveel extra is het ons waard?' (Healthcare: how much extra it is worth to us?). 2016, The Dutch Ministry of Health, Welfare, and Sport.
23. Reilly, B.M. and Evans, A.T., Translating clinical research into clinical practice: impact of using

prediction rules to make decisions. Annals of internal medicine, 2006. 144(3): p. 201-209.
24. Segaar, D., Bolman, C., et al., Identifying determinants of protocol adoption by midwives: a comprehensive approach. Health Educ Res, 2007. 22(1): p. 14-26.
25. Offerhaus, P., Fleuren, M., et al., Guidelines on anaemia: effect on primary-care midwives in The Netherlands. Midwifery, 2005. 21(3): p. 204-11.
26. Zeitlin, J., Manktelow, B.N., et al., Use of evidence based practices to improve survival without severe morbidity for very preterm infants: results from the EPICE population based cohort. Bmj, 2016. 354: p. i2976.

Dankwoord

Allereerst wil ik alle zwangere vrouwen die mee hebben gedaan aan de Expect studie hartelijk danken voor hun moeite om de vragenlijsten in te vullen en voor het delen van hun medische dossier. Zonder jullie bereidheid om deel te nemen had dit proefschrift niet kunnen bestaan.

Dan nu een officieel woord van dank aan mijn voltallige promotieteam, prof. dr. L.J.M. Smits, dr. H.C.J. Scheepers en prof. dr. M.E.A. Spaanderman.

Luc, of beter gezegd, professor Smits; gedurende mijn promotie behaalde ook jij een mijlpaal in je carrière met jouw benoeming tot hoogleraar. Erg bijzonder en leuk voor mij om dat proces van dichtbij te mogen meemaken. Allereerst denk ik dat ik je vrouw mijn excuses aan moet bieden, aangezien ik waarschijnlijk medeverantwoordelijk ben voor een paar zilveren wijsheids-haren. Weet je nog die keer dat ik even was vergeten om het copyright te checken? …oeps…
Waarschijnlijk heb ik je ook menig keer het vuur aan de schenen gelegd tijdens onze frequente bijeenkomsten. Gelukkig waren we het nog veel vaker volledig met elkaar eens. Onze discussies hebben daarbij niet alleen inhoudelijk bij gedragen aan de kwaliteit van het proefschrift, maar ook mij als persoon en wetenschapper verder gevormd. Voor mij ben je een hele fijne begeleider geweest, waarvan ik enorm veel heb kunnen leren. We hebben samen een leuke tijd gehad op Deb., maar ook daarbuiten met onze Expect-etentjes en congressen. Zoals dat buitenkansje in Sint-Petersburg, waar wij niet alleen de laatste ontwikkelingen van de perinatologische wetenschap hebben mogen ontdekken, maar ook de Hermitage en lekkere restaurants. Luc, bedankt voor al je inspanningen!

Liesbeth, het is inmiddels zes jaar geleden dat wij elkaar leerde kennen. Ik had eigenlijk al een stageplek geregeld en de deadline was al verstreken toen ik mij realiseerde dat ik toch veel liever mijn combistage bij de gyn wilde lopen. Gelukkig had jij meteen alle vertrouwen in mij. Inmiddels heb ik geleerd dat deze situatie voor jou geen uitzondering was, jouw enthousiasme laat het nooit afweten! Het enthousiasme spat daarbij soms letterlijk van het scherm met zinnen die geen einde...
Je passie voor het vak en de wetenschap werken voor mij aanstekelijk. Juist op momenten van tegenslag, wist jij er vaak weer een positieve impuls aan te geven. Verder kon ik rekenen op je scherpe klinische blik bij de interpretatie van de resultaten en de translatie daarvan naar de dagelijkse praktijk. Daarmee heb je mij enorm geholpen bij het vormgeven van de discussies in dit proefschrift. Je doorgaans nuchtere en relativerende houding hebben mij tevens geholpen menig knoop door te hakken. Ik hoop dat ik ook in de toekomst daarvoor bij je mag blijven binnen lopen. Enorm bedankt voor je inzet en betrokkenheid!

Marc, altijd een bomvolle agenda, maar ook altijd een begripvol oor. Regelmatig was een herinneringsmailtje, of twee, en misschien nog een keertje langslopen, noodzakelijk. Bewonderenswaardig vond ik het dat, wanneer we dan om tafel zaten, die drukte als vanzelf naar de achtergrond verdween en jij alle tijd nam voor mij. Jouw rustige, wel overdenkende houding hebben zeker hun positieve invloed gehad. Vaak was jij een van de eerste coauteurs die mijn manuscripten las, zonder er vanaf de eerste letter bij betrokken te zijn. Voor mij

vormde jouw kritische blik dan ook de vuurdoop en met je constructieve feedback kon ik mijn manuscripten verder verbeteren. Op deze manier kon ik mijn manuscripten met vertrouwen indienen voor publicatie.

De leden van de beoordelingscommissie, prof. dr. I.C.W. Arts, prof. dr. R.F.P.M. Kruitwagen, dr. R. Hermens, dr. M. Nieuwenhuijze en prof. dr. M.A. Joore, wil ik bedanken voor de beoordeling van dit proefschrift.

De leden van de Expect projectgroep wil ik eveneens graag bedanken: prof. dr. M.E.A. Spaanderman, prof. dr. R.G. de Vries, prof. dr. C.D. Dirksen, dr. I. Korstjens, prof. dr. A.L.M. Mulder, dr. M.J. Nieuwenhuijze, prof. dr. J.G. Nijhuis en prof dr. O.C.P. van Schayk.

Zonder de medewerking van alle betrokken verloskundigenpraktijken en ziekenhuizen was de Expect Studie niet tot stand gekomen. In het bijzonder wil ik een woord van dank richten aan de LOQS-projectgroep en werkgroepleden:
De gynaecologen Annemieke van Wijck, Ella Wijnen-Duvekot, Iris Zwaan, Ivo van Dooren, Josje Langenveld, Martine Wassen, Robert Aardenburg, Maartje Zelis, Liesbeth Scheepers, Marc Spaanderman, maar ook niet-werkgroepleden Salwan Al-Nasiry en Stijn van Teeffelen. Jullie hebben elk een sleutelrol gespeeld bij de implementatie en het overwinnen van diverse uitdagingen. Bedankt voor jullie inzet!
Uiteraard wil ik ook benadrukken dat de verloskundigen Lilian Custers, Badia El Haddad, Cindy Bastings, Claudia Brouwers, Jose ten Thije, Nicky Maassen, Wendy Gijsen, Desiree Greven en Hilde Coolen met hun inzet om de 'massa' in beweging te krijgen een belangrijke rol hebben gespeeld.
Verder ben ik er vrijwel zeker van dat ik nu een aantal namen niet heb genoemd van mensen, die buiten mijn zicht, zich hard hebben gemaakt voor de Expect en ervoor zorgden dat er genoeg vrouwen deelnamen aan de studie. Zo ook de kraamcentra, met in het bijzonder, Babette Peeters en Marie-Louise Verstappen-Wouters, hebben hierin ook een belangrijke rol gespeeld. Ik ben jullie allen zeer dankbaar.
Ten slotte verdiend ook Jolanda Willems-Roberts, research nurse van het Zuyderland, het om hier genoemd te worden. Jolanda enorm bedankt jouw inzet en betrokkenheid bij het werven van deelneemsters voor de Expect Studie.

Yvonne, toen we eenmaal op elkaar ingespeeld waren wisten wij perfect wat we aan elkaar hadden. Jouw inzicht, ervaringen en advies zijn zeer waardevol geweest bij de totstandkoming van de implementatie.
Marianne en Lonneke, ik heb maar korte tijd met jullie samengewerkt, maar jullie zijn een echte aanwinst voor de LOQS.

Bij deze wil ik alle coauteurs hartelijk danken voor alle feedback en jullie kritische blik op mijn manuscripten. Stéphanie, we hebben intensief samengewerkt aan een gedegen manuscript. Het is mooi om te zien dat al onze inspanningen inmiddels beloond worden en hun vruchten beginnen af te werpen. Veel succes bij de verdere afronding van je proefschrift!
Sander, in het begin was jij voor mij de rots in de branding toen ik nog pas startte met R. Later in het promotietraject bleef je bereikbaar voor goed advies en nuttige tips met betrekking tot de analyses. Daarmee was je voor mij een zeer gewaardeerde coauteur. Bedankt voor

je tijd en inzet.
Carmen, bedankt voor alle hulp, wijsheid en bevestiging omtrent het kosteneffectiviteitsonderzoek. Dit hoofdstuk was niet tot stand gekomen zonder jouw bijdrage en inzet, waarvoor dank!
Laure, nadat Luc eerst een paar maanden naast zijn schoenen liep vanwege de goede kandidate die bij hem had gesolliciteerd, was ik toch erg nieuwsgierig geworden. Al gauw kwam ik erachter dat hij niet overdreven had. In de korte tijd die we samen hebben gewerkt heb ik veel geleerd van jouw statistisch inzicht, wat nog op het laatste moment een stempel heeft gedrukt op mijn proefschrift, bedankt.

Stagiaires Pia, Mandy, Babette, en Clémence, ik ben jullie dankbaar voor jullie inzet voor de Expect Studie en prettige samenwerking.

Ik wil MEMIC centrum voor data en informatie management bedanken voor de fijne samenwerking. Dirk en Luc, jullie hebben een belangrijke bijdrage geleverd aan de uitvoering van onze studie opzet, de Expect calculator en de benodigde datakoppelingen. Luc, vaak had jij meteen tijd voor mij wanneer dat nodig was. Dirks enthousiasme is zelfs overgeslagen op zijn vrouw die vervolgens heeft deelgenomen aan de Expect ;-)

De heren van Sience Vision, Ger en Jule, wil ik bedanken voor hun bijdrage en inspanningen om de Expect Studie alsook de nieuwe geboortezorg met beeldmateriaal te promoten.

Pieter Kubben, we hebben elkaar leren kennen toen ik nog als coassistent deelnam aan de eerste Hackathon Dutch Hacking Health te Maastricht, luchtig borrelend met een lekker wijntje bij de vooraankondiging. Als neurochirurg had je inhoudelijk geen enkele relatie met mijn project, maar we deelde samen een duidelijk enthousiasme en passie voor (digitale) innovaties in de zorg. Je advies en nuchtere vrolijkheid hebben mij geholpen de Expect Calculator vorm te geven, en hebben mij gestimuleerd mijn R-skills verder te ontwikkelen wat zeer van pas is gekomen gedurende het promotietraject.

Cyriel Heuts, bedankt dat je voor ons een netelige kwestie goed hebt weten op te lossen.

De dames van het secretariaat op Deb., Yvonne, Mariëlle, Petra en Irma, wil ik bedanken voor alle hulp en ondersteuning op de afdeling.
Monique Janssen, Manon Gordijn en Conny de Zwart wil ik in het bijzonder bedanken voor hun inspanningen. Het plannen van onmogelijke afspraken is toch echt jullie specialiteit! Gelukkig mocht ik altijd een beroep op jullie doen om een overleg in te plannen, uiteraard meestal op korte termijn.
Ella wil ik bedanken voor het feit dat de logistiek van mijn onderwijstaken tiptop georganiseerd was.

Harry en Jos bedankt voor jullie technische ondersteuning. Jos, als behulpzame aanpakker sta je voor anderen klaar. Of ik nu problemen had met mijn bureau, beeldscherm, of (meermaals) met een compleet gecrashte pc, jij dacht met mij mee en wist het snel op te lossen. Harry, bedankt voor alle vernuftige oplossingen waarmee menig tijdrovend ICT-karweitje ‘on the fly’ werd uitgevoerd. Daarmee was jij, een niet te onderschatten stille

kracht van ons project, wat kan een aio zich nog meer wensen?

Ilse Mesters, jouw expertise op het gebied van gezondheidsvoorlichting heeft duidelijk bijgedragen aan onze patiëntfolders afgestemd op onze doelgroep. Ik kan het wellicht niet evidence-based staven met mijn data, maar ik ben ervan overtuigd dat dit een positieve invloed heeft gehad op de mate waarin vrouwen de nieuwe adviezen opgevolgd hebben.

Mijn kantoormaatjes Jessica, Linda, Louise, Karlijn en Vivianne wil ik bedanken voor de vele gezellige uurtjes op A2.049.
Karlijn, wij hebben samen het langste de kamer gedeeld tijdens mijn promotietraject. Ik hoop dat jij ook een nieuwe werkplek vindt waar je je als een vis in het water voelt. Waar je ook terecht komt, jouw doorzettingsvermogen en harde inzet zal overal gewaardeerd worden!

Mijn Expect-buddies Jessica en Linda! Wat hebben we het naar onze zin (gehad) met zijn drietjes.
Linda, als stagiaire startte ik op 'jouw' project. Een beter voorbeeld als promovendus had ik mij niet kunnen wensen. De lat ligt bij jou zo hoog, dat ik soms twijfel of er überhaupt wel ergens een lat ligt. Jouw precisie en grondigheid hebben voor mij de toon gezet als wetenschapper. Verder vulden wij elkaar, juist door onze verschillen, denk ik perfect aan en hebben we elkaar regelmatig gesteund bij de tegenslagen die iedere onderzoeker op zijn bordje krijgt. Het was dan ook erg wennen toen ik ineens zonder jou verder moest met de Expect studie. Gelukkig zien we elkaar nog steeds regelmatig. Ik wens je heel veel succes bij de huisartsgeneeskunde!
Jessica, onze enthousiaste turbo van het team. Ik kan mij niet herinneren dat ik jou eens *niet* vrolijk heb meegemaakt. De positieve energie straalt als een aura om je heen en heeft zelfs via de telefoon een effect. Je bent dan ook de perfecte kandidaat voor het kwalitatieve onderzoek van de Expect. Het is noemenswaardig hoe je dit weet te combineren met je klinische werkzaamheden. Heel veel succes met de afronding van het onderzoek en de huisartsgeneeskunde!

Ik wil alle overige collega's en aiossen van de afdeling bedanken voor de fijne werksfeer en gezelligheid tijdens de koffiemomentjes en teamuitjes. De verrassingslunch die jullie voor mij hadden georganiseerd was ook echt heel leuk, bedankt daarvoor.
In het bijzonder wil ik Sophie en Jacqueline nog bedanken, we liepen regelmatig bij elkaar de kamer binnen en hebben menig lunch samen genuttigd. Vooral in de tijd dat ik alleen zat op mijn kantoortje en de muren soms op mij afkwamen was ik daar zeer dankbaar voor. Mijn buurman, Adrie, altijd in voor een praatje en geïnteresseerd in mijn onderzoek wil ik eveneens in het bijzonder benoemen.

Vrienden en vriendinnen, ik verheug mij enorm op de momenten die wij samen doorbrengen en heb altijd wel iets om naar uit te kijken.
Allon, Joost, Laurens, Luuk en Rob, de mannenavond, inmiddels mannenweekend, is toch echt een begrip en vast ritueel geworden! Inmiddels is de groep bijna verdubbeld met Jill, Judith, Lieke, Stephanie en Zoë. Het is hartstikke leuk dat we samen zoveel plezier hebben – ultimate kaos! ;-) – en lief en leed met elkaar kunnen delen.

Laurens, wij hebben elkaar ‘echt’ leren kennen op weg naar de open dag van Universiteit Utrecht. Het is zeker niet bij die ene dag in trein gebleven en ben je naast een goede vriend ook schoonbroer geworden. Inmiddels hebben we niet meer de luxe om op doordeweekse middagen urenlang op het water door te brengen. Gelukkig zien we elkaar nog regelmatig. Minimaal één keer per jaar omslaan met de boot lukt ons trouwens nog steeds. Bedankt voor alle steun en gezellige momenten.
Lennart, ons promotietraject verliep de laatste jaren grotendeels parallel aan elkaar. De vele cappuccino’s, het hardlopen, of het fietsen door het heuvellandschap waarbij we ons hoofd konden leegmaken zijn erg gezellig. Dat we nog flink wat kilometers samen mogen maken!
Etiënne, we kennen elkaar al sinds de A-KO en ook na het behalen van ons artsendiploma zien we elkaar geregeld, zo niet wekelijks. We hebben al menig serie samen versleten en ’s zomers steken we de BBQ graag aan, op naar een nieuw seizoen!

Mijn familie en schoonfamilie wil ik bedanken voor de warme momenten samen en hun interesse in mijn onderzoek. De prijs van vrolijkste Expectbaby gaat zonder meer naar Nore. Pap en mam, toen ik eenmaal vertrok naar Utrecht om diergeneeskunde te studeren was volgens mij inmiddels alle hoop dat ik nog enigszins in de buurt zou blijven wonen inmiddels vervlogen. Hoe anders is die situatie nu, wonend in heuvelland, bij de vierde (soort van) diploma-uitreiking. Jullie hebben mij door dik en dun gesteund en staan altijd voor mij klaar. Heel veel dank voor de support die jullie mij hebben gegeven en nog steeds geven!
Dorien, wie had dat gedacht; dat we nog een tijdje zouden samenwonen nadat we het ouderlijk huis hadden verlaten. Inmiddels zijn we allebei weer diverse malen verhuisd, ik heb nog steeds rugpijn van die @#$* wasmachine op zolder..., en zien we elkaar door de afstand minder vaak. Desondanks kunnen we nog steeds op elkaar rekenen.
Lieke, mijn zusje, collega, vriendin, onze wegen kruisen elkaar overal en daar ben ik dankbaar voor. Ik schiet nog steeds in de lach als ik die verwarde blik van de verpleging herinner toen we als semi-arts voor patiëntinformatie ons naar elkaar lieten doorverbinden. Wie weet staan we in de toekomst nog eens samen op dezelfde verloskamer, team dokter van Montfort, dat is pas continuïteit van zorg!

Lieve Judith, mijn onderzoek was vaak voor jou, net zoals voor de rest van mijn familie en vrienden, een ver-van-mijn-bed show. Desondanks steun(de) jij mij door dik en dun. Gelukkig komen mijn opgedane vaardigheden nog goed van pas bij de planning van de achtertuin. Wat hebben we het toch goed samen!

Curriculum Vitae

Pim van Montfort was born on June 14th, 1991 in Roermond, the Netherlands. In 2009 he completed secondary school at Connect College in Echt with the distinction cum laude. Pim started studying Veterinary Medicine at Utrecht University (Utrecht, the Netherlands) and obtained a Bachelor of Science degree with honor in 2012. After his graduation, Pim enrolled in the four-year master of Medicine and Clinical Research at Maastricht University (Maastricht, the Netherlands). During the master he performed a research internship at the department of Epidemiology, joining the Expect Study group. Pim obtained his degrees Medical Doctor and Master of Science in 2016. Consecutively, he started his PhD research at the Department of Epidemiology at Maastricht University (Care and Public Health Research Institute, CAPHRI) under supervision of prof. dr. Luc J.M. Smits, dr. Liesbeth (H). C.J. Scheepers, and prof. dr. Marc E.A. Spaanderman. The research, Expect Study II, was performed in collaboration with the Limburg Obstetric Consortium. The scientific work presented in this thesis, was published in peer-reviewed journals and presented on several national and international conferences. Pim currently works at the department of Obstetrics and Gynaecology at Zuyderland Medical Centre.

Add.

List of publications

van Montfort P, Smits LJM, van Dooren IMA, Lemmens SMP, Zelis M, Zwaan IM, Spaanderman MEA, Scheepers HCJ. Implementing a pre-eclampsia prediction model in obstetrics: cut-off determination and healthcare professionals' adherence. Medical Decision Making. 2019.

Smits LJM, **van Montfort P**, Meertens LJE, Lemmens SMP, Röselaers YCM, Spaanderman MEA, Scheepers HCJ. De Limburgse aanpak: Populatiebrede preventie van pre-eclampsie. Nederlands Tijdschrift voor Obstetrie en Gynaecologie. 2019; 132(3):130-132.

Meertens LJE*, **van Montfort P***, Scheepers HCJ, van Kuijk SMJ, Aardenburg R, van Dooren IMA, Langenveld J, Zwaan IM, Spaanderman MEA, Smits LJM. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta Obstetricia et Gynecologica Scandinavica, 2018; 97(8):907-20

van Montfort P, Willemse JPPM, Dirksen CD, van Dooren IMA, Meertens LJE, Spaanderman MEA, Zelis M, Zwaan IM, Scheepers HCJ, Smits LJM. Implementation and effects of risk-dependent obstetric care in the Netherlands (Expect Study II): Protocol for an impact study. JMIR Research Protocols. 2018.

van Montfort P, van der Steen E, Hukkelhoven M, de Jong A, Rennenberg R, van Mook W. Wat verwacht de maatschappij van een huisarts buiten diensttijd? Huisarts en Wetenschap. 2018; 61(7).

van Montfort P, Stijntjes M, van de Pol M, Lipman L. Wat moet de praktiserende dierenarts weten van het European Bat Lyssavirus? Tijdschrift voor Diergeneeskunde. 2013 138(4).

Submitted

Lemmens SMP*, **van Montfort P***, Meertens LJE, Spaanderman MEA, Smits LJM, de Vries R, Scheepers HCJ. Perinatal factors related to pregnancy and childbirth satisfaction: a prospective cohort study.

van Montfort P, Scheepers HCJ, van Dooren IMA, Meertens LJE, Zelis M, Zwaan IM, Spaanderman MEA, Smits LJM. Low-dose-aspirin usage among women with an increased pre-eclampsia risk: a prospective cohort study.

van Montfort P, Scheepers HCJ, van Dooren IMA, Meertens LJE, Wynants L, Zelis M, Zwaan IM, Spaanderman MEA, Smits LJM, Adherence rates to a prediction tool identifying women with an increased gestational diabetes risk: an implementation study.

van Montfort P, Scheepers HCJ, Dirksen CD, van Dooren IMA, Meertens LJE, van Kuijk SMJ, Wijnen EJ, Zelis M, Zwaan IM, Spaanderman MEA, Smits LJM. Impact on perinatal health and cost-effectiveness of risk-based care in obstetrics: a before after study.

*Contributed Equally