Improving evaluation of obstetric interventions
van 't Hooft, Janneke

Citation for published version (APA):
van 't Hooft, J. (2016). Improving evaluation of obstetric interventions

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
‘drawings from children that participated in the ProTwinkids follow-up study (chapter 3)’
CHAPTER 1

GENERAL INTRODUCTION
Chapter 1

GENERAL INTRODUCTION

In most pregnancies the synergy between mother and her unborn child is adequately balanced, resulting in the birth of the baby at the end of an uncomplicated pregnancy. Unfortunately, not all pregnancies and deliveries remain in such optimal balance. In fact, the day of birth is a high risk event for both mother and child, with a more than 5 times greater risk of dying for the mother and 400 times greater risk of dying for the baby than travelling 370 km by car.\(^1\) Moreover, pregnancies can be complicated by high blood pressure of the mother, suboptimal growth of the foetus, foetal distress before or during labour, or preterm birth.

Many new and existing interventions can be offered to pregnant women who face a problem in pregnancy or during labour. In order to guide clinical as well as policy decision making, evaluation research is needed to establish evidence on effectiveness and potential harm of these interventions. A randomized controlled trial (RCT) is worldwide considered as the best instrument to evaluate the effectiveness of medical interventions. It is defined as a *prospective study comparing the effect and value of intervention(s) against a control in human beings.*\(^2\) By randomly allocating subjects, an RCT incorporates a control group which does not differ from the intervention group except for the intervention being studied (Figure 1). But although RCTs represent primary research with the highest level of evidence, a single RCT is still prone to chance for false positive or false negative results, limited generalisability or various forms of bias. In addition, research that is relevant to evaluate whether an intervention is effective can be scattered all over the literature and published in different languages. Systematic reviews and meta-analyses identify these relevant studies, appraise their quality and summarize their results using scientific methodology.\(^3\) The aggregated evidence gives a more balanced answer to a research question and therefore systematic reviews and meta-analysis are considered to have a higher level (hierarchy) of evidence compared to separate RCTs (Figure 1).\(^4\) An individual patient data (IPD) meta-analysis is a specific type of systematic review. Rather than extracting summary (aggregate) outcomes from study publications, the original research data are sought directly from the researchers responsible for each study. These data can then be re-analyzed centrally and combined, if appropriate, in a meta-analysis. IPD-meta-analysis
can provide additional relevant results by analyzing associations at the individual patient level. They include the ability to allow in depth exploration of patient factors and subgroup analyses and have been described as the gold standard of systematic reviews.\textsuperscript{5}

**Variation in outcomes used in obstetrical evaluation research**

When an RCT or systematic review (SR) addresses a relevant question regarding a specific population, in which an intervention group is compared to a comparison group, it will measure and report on key outcomes that provide a better understanding on the effectiveness and safety of that intervention at a specific time point (PICO structure, Figure 1). Outcomes used in RCTs and SRs are ideally of real importance to the population. However, if researchers have a more biological/mechanistic oriented question the outcomes chosen might be different compared to more clinically related research questions. Within the context of clinical evaluation research, we will limit our exploration of the problem of variation in outcomes used in RCTs and SRs to clinical outcomes of obstetric interventions. In the design phase of a clinical trial about prevention of preterm birth, for example, the chosen ‘outcome’ can be ‘gestational age at delivery’ or ‘admission to neonatal intensive care’ or ‘respiratory problems of the neonate’. Besides collection of outcomes with the greatest (therapeutic) importance for the patients,\textsuperscript{6} outcomes are selected because of their available (internal and external) validated measurement tools. However, researchers may need to make pragmatic decisions when designing a trial. Funding and time limitations may mean that outcomes with higher event rates that are easy to measure are more attractive, increasing the statistical power of the trial at the expense of relevance for patients. Also historical perspective (outcomes that are already used by other researchers in the same field) and special interest of the researchers team can influence the list of outcomes used.

In preterm birth clinical research the lack of consistency in choice of outcomes has led to over 72 different primary outcomes being reported in 103 clinical trials.\textsuperscript{7} The same lack of consistency in the choice of outcomes exists in SR en meta-analyses: in 33 Cochrane reviews on preterm birth, 29 different primary outcomes were reported.\textsuperscript{7}
Chapter 1

Figure 1. An overview of evaluation research in obstetrics and some of the problems we face
Theoretically, total freedom of research teams to choose the outcomes used for their RCT or SR gives rise to several problems:

1) The selected outcomes may not be the most relevant ones (especially relevant for patients and clinicians).

2) When relevant outcomes cannot be easily acquired, there may be a tendency to report on intermediate, surrogate or proxy outcomes. This can give misleading (and even harmful) results, further explained in the next section.

3) Similar trials may use a wide variety of outcomes, outcome measurement tools and definitions which hampers the comparison and meta-analysis of results of various trials with similar goals, and thus leads to inefficiency and waste of research.

These problems of lack of consistency in outcomes reported in RCT and SR, and the lack of reporting of relevant outcomes can be addressed by introducing the use of a core outcome set (COS) in research, i.e. a set of critical and important outcomes that should be measured and reported, as a minimum, in a standardised manner. A core outcome set captures the key outcomes to be used in trials on a specific topic, defined through an international consensus involving all relevant stakeholders (including patients) using proper methodology. The introduction of core outcome sets, enhances the translation and integration of research in decision making into clinical practice (i.e. Evidence Based Medicine).

Long-term outcomes

Many interventions applied in pregnancy are evaluated for their efficacy and safety by measuring short-term maternal and neonatal outcomes (Figure 1). Neonatal follow-up often ends at the moment of the child’s discharge from the hospital or within 6 to 10 weeks after the expected term date. In obstetrical research to evaluate the effect of a specific perinatal intervention only a small minority (approximately 16%) of large RCTs reports on long-term follow-up of the child. These short-term outcomes can be surrogate outcomes or short-term clinical outcomes. An example of a surrogate outcome is the Apgar score at 1, 5 and 10 minutes after birth as a surrogate for short-term and long-term mortality and morbidity. Although there may be an association between a surrogate outcome and long-term outcome (e.g. there is an association of
Apgar score <7 at five minutes with increased risk of neurologic disability) the vast majority of children born with a low Apgar score grow up without disability. Moreover, use of surrogate outcomes in clinical research can also have serious harmful effects. There are numerous examples of drugs used in the past for heart diseases that had been approved on the basis of surrogate outcomes, but were ultimately proven to be harmful by increasing mortality rates. So restricting conclusions to short-term surrogate outcomes can lead to seriously erroneous conclusions due to the fact that these outcomes may not reflect any possible clinical effect.

Subsequently, more reliable short-term outcomes (e.g. admittance to neonatal intensive care, or problems related to early neonatal life such as respiratory distress syndrome) still have their drawbacks because they do not show the full scope of information necessary to assess clinical impact. Thus, restricting conclusions to short-term outcomes can also have serious drawbacks due to the fact that the risk-benefit ratio of any perinatal interventions may change considerably both for the pregnant woman and her infant, between the period immediately after birth and later on in childhood. This was shown, for example with the ORACLE II study, on use of antibiotics for women in spontaneous imminent preterm labour. In this study no short-term benefit in the use of antibiotics compared to placebo was seen in the initial trial. At follow-up after seven years a potential harmful effect of the use of erythromycin in the children was found, indicated by an increased risk of cerebral palsy RR [95%CI] 1.69 [1.07 to 2.67]. Another trial, evaluating the use of vitamin K and phenobarbital to prevent intracranial haemorrhage in newborns less than 34 weeks gestation, also showed no effect on the short-term, but significantly lower Bayley scores in the treatment group compared to the placebo group (mean scores (SD) of 104 (21) vs 113 (22), p=0.023). Warning signs of long-term harm were seen in trials evaluating the use of progesterone, and the use of repeated doses of corticosteroids in women with a high risk of imminent preterm labour. The OPPTIMUM trial showed an increased risk (although still of low frequency) for problems related to renal, gastrointestinal, and respiratory systems in the progesterone group (e.g. gastrointestinal disability in 4 (1%) in placebo vs 9 (2%) in progesterone group, OR [95%CI] 2.67 [1.37 to 5.20]), while repeated doses of corticosteroids evaluated in another trial showed an increased risk (though not significant) for cerebral palsy RR [95%CI] 5.7 [0.7 to 46.7] compared to single dose corticosteroids in imminent preterm labour. Another famous example
of prenatal effects that only came to expression later in adulthood is the Dutch famine study, a historical cohort that provided information on the effects of famine exposure during specific periods of gestation on outcomes measured at birth and outcomes in adulthood. Data from 821 children exposed in utero to famine (divided in subgroups of early-, mid- and late gestation of exposure) were compared to data of 1593 children that were conceived before and after the period of famine. Babies exposed to maternal famine in late- or mid gestation were lighter, shorter, thinner and had smaller head circumference than babies that had not been exposed to famine. The long term consequences found (metabolic syndrome –including high blood pressure, obesity, misbalanced lipid profiles and glucose intolerance- breast cancer, depression, airways disease and renal function) were however to a large extent independent of size at birth, underlining the fact that programming may take place even without effects that are not visible immediately after birth. Long-term follow-up of mothers and children participating in obstetrical trials is therefore pivotal.

Integrating outcomes of obstetrical evaluation studies to guide clinical decision making
Now that we have introduced the importance of consistency in (relevant) outcomes reported in RCTs and SRs, and the added value of long-term outcomes of obstetrical interventions, it will be clear that most ideally, the measured outcomes in clinical research will have an impact on clinical practice. An efficient system of research addresses health problems of importance to populations and interventions and outcomes considered important by patients and clinicians. However, much has been written about research waste due to low priority questions, inappropriate study design and problems in access to study data and obtaining unbiased reports. This subsequently leads to difficulties of implementation of research into clinical practice. A quote of dr. Ioannidis in a published essay entitled ‘why most clinical research is not useful’ demonstrates this: ‘Practicing doctors and other health care professionals will be familiar with how little of what they find in medical journals is useful. The term “clinical research” is meant to cover all types of investigation that address questions on the treatment, prevention, diagnosis/screening, or prognosis of disease or enhancement and maintenance of health. Experimental intervention studies (clinical trials) are the major design intended to answer such questions, but observational studies may also offer relevant evidence. “Useful clinical research”
means that it can lead to a favorable change in decision making (when changes in benefits, harms, cost, and any other impact are considered) either by itself or when integrated with other studies and evidence in systematic reviews, meta-analyses, decision analyses, and guidelines.

In this thesis we will address some clinically based research questions, and we will discuss the integration of outcomes from obstetrical evaluation studies (as suggested by Ioannidis in the above quote) in systematic review/meta-analysis, cost-effectiveness analysis and budget impact analysis in order to give guidance for clinical decision making. We will start with an example of clinically based research and introduce some of the methodologies.

Clinically based research questions
An example of a clinically based research question addressed in this thesis originated from doctors working in the neonatal intensive care in Amsterdam. It is know that preterm birth is associated with an increased risk of neurodevelopmental problems. However, not all children born preterm will develop developmental problems, and if there are problems, there is a broad range in type of problems (cognitive, motor, visual, etc) and severity (mild to severe). Predicting the long term impact of a preterm birth can be of great value as it may help parents to better prepare for the future and improve selection of children that may benefit from early intervention programs (i.e. physiotherapy or speech therapy) to improve outcomes. However, if the predictive value is poor, it may invoke unwanted effects, as parents may worry unnecessarily about the possible abnormal development of their child. Neonatologists were in doubt whether to perform brain MRI in all very preterm born neonates at term equivalent age. Several studies reported high predictive value of term equivalent MRI on long term development of these children. But no systematic review and meta-analysis was available on this topic. However, after discussing this topic with international colleagues at conferences apparently many of them were already convinced by this technique and were using this as standard care in their clinical practice. Instead of blindly implementing this imaging technique in standard care, the department of neonatology conducted a systematic review and meta-analysis (a chapter incorporated in this thesis). This example shows that research aimed at answering a research question arising from clinical practice may have a higher chance of influencing clinical practice than research which does not have such a close connection with daily clinical practice.
Meta-analysis using bivariate model to assess predictive value of a prognostic tool

To determine the predictive value of a prognostic tool is challenging. First, the time-frame between the performed prognostic test (e.g. Apgar score, cord blood pH, brain MRI) and the outcome of interest (e.g. neurodevelopment) is broad, resulting in the lack of studies evaluating this topic due to feasibility reasons. Second, the outcome of interest (unfavourable neurodevelopment) can vary (e.g. cerebral palsy, visual and/or hearing problems, motor-, neurocognitive- and behavioural problems). However, information of each cohort studies (e.g. prospectively following a consecutive sample of patients presenting with the prognostic dilemma) can be useful (acknowledging the increased risk of bias due to confounding factors in this type of studies). These studies report estimates of sensitivity (correctly detecting those with the target condition) and specificity (correctly identifying those without the target condition) of the prognostic test, which can be pooled in meta-analytic approaches. Because sensitivity and specificity-values of a test are related and very much depending on the used cut-offs, the bivariate model has the advantage of preserving the two-dimensional nature of the underlying data and gives insight in the optimal use of the predictive test in clinical practice.26

Cost-effectiveness analysis of an obstetrical intervention using data from clinical research

As the rising health care expenditures and affordability questions become increasingly relevant for health care decision making, interventions also need to be evaluated in terms of economic outcomes. Many RCTs are therefore complemented with an economic evaluation. Health care use and costs are estimated as economic outcomes, and related to clinical outcomes through cost-effectiveness analyses, cost-utility analyses or cost-benefit analyses. Cost-effectiveness analyses most often involves the comparison of two or more interventions/alternatives where the health gains of one intervention is related to the additional costs (or cost savings) associated with that intervention, relative to the comparator. In case health gains are achieved at increased cost, the question is whether the health gains are ‘worth’ the extra costs (willingness to pay).27 In case an intervention leads to health gains and cost savings at the same time, implementing this intervention would benefit patients while reducing health care expenses.
Chapter 1

Innovative (and costly) interventions and drugs often result in cost increases, and need consideration whether society is willing to pay for the anticipated benefits. In many other cases, where interventions optimize existing care arrangements, the improvement in health and reduction of costs often go hand in hand, and cost savings can be expected.

**Budget impact for exploring potential health and budget impact in a population before and after implementation of clinical trial results.**

Finally, if we succeed to perform clinical research that has impact on clinical practice we can estimate its potential (health and economical) impact by performing a budget impact analysis (BIA). The purpose of a BIA is to estimate the financial and healthcare consequences of adoption and diffusion of a health-care intervention within a specific health-care setting or system. For this purpose a BIA can be used for prediction of a shift in prevalence of disease (or health conditions) and cost after implementation of interventions that are found to be effective and de-implementation of interventions that are found to be ineffective in evaluation research. A BIA can provide additional information that helps to motivate clinicians to implement evaluation research into clinical practice.

**PROBLEMS**

In summary we face the following problems in evaluation research in obstetric interventions:

1) a lack of standardization in the selection and operationalization of outcomes. This may lead to inefficiency in research and waste of resources.

2) a lack of systematic follow-up data of randomized controlled trials, leaving a blind spot in clinical research.

3) a gap between clinical research and its impact in clinical decisions. Patients therefore do not fully benefit from the available evidence.
AIMS OF THE THESIS

This thesis focuses on improving evaluation research on obstetric interventions. The aims are to:

• develop a core outcome set that can be used in obstetrical evaluation studies
• measure long-term outcomes of obstetrical evaluation studies
• integrate outcomes of obstetrical evaluation studies in order to guide clinical decision making

OUTLINE OF THE THESIS

The thesis is divided in three parts. Part I describes the development of a core outcome set that can be used in obstetrical evaluation studies. Chapter 2 presents the development of a core outcome set (COS) for studies on prevention of preterm birth developed with an international e-Delphi consensus group. This COS reflects the outcomes that are critically important to all relevant stakeholders (patients, obstetricians, midwives, neonatologist and researchers).

Part II explores ways to measure long-term outcomes of obstetrical intervention studies. Chapter 3 evaluates the long-term effects in children born to mothers with a short cervical length that were given a pessary during twin pregnancy in a randomized controlled trial. Chapter 4 evaluates the long-term effects in children born to mothers with a short cervical length in a singleton pregnancy. These women were included in a randomized controlled trial comparing the use of vaginal progesterone in the second and third trimester with placebo to prevent preterm birth.

Part III deals with outcomes of obstetrical evaluation studies in systematic reviews, cost effectiveness analysis or budget impact analysis to give guidance for clinical decision making. Chapter 5 evaluates the predictive value of brain MRI results for long-term developmental outcomes in children born preterm or with a low birth weight. Chapter 6 models the short and long-term costs and effects of using an advanced form of foetal monitoring (ST-analysis) during
labour when compared to conventional foetal monitoring from a maternal and neonatal perspective.

Finally, by performing evaluation studies of obstetrical interventions our ultimate goal is to improve health outcomes of mothers and their children at an acceptable cost. Therefore, implementation of trial results is a crucial step. In Chapter 7 explores the potential impact of implementation of nationwide evaluation studies in obstetrics on health and costs at a national level. It assesses whether evaluation research leads to cost-savings and if this covers the cost of performing them.
REFERENCES

Chapter 1


