Economic evaluations and clinical decision making in obstetrical care
van Baaren, G.J.

Citation for published version (APA):
van Baaren, G. J. (2014). Economic evaluations and clinical decision making in obstetrical care

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.
CHAPTER 10

Summary and general discussion
SUMMARY

The key question in clinical evaluation research used to be whether a specific intervention leads to better health outcomes than the next best alternative, usually standard care. Rising health care costs and scarce resources have increased the relevance of economic evaluations for supporting decision making. Nowadays, the ultimate purpose of evaluating interventions used in clinical practice is to provide the most possible care, in other words most value, for money.

The aim of the research reported in this thesis was to assess the effectiveness, costs and cost-effectiveness of a number of diagnostic and treatment strategies in obstetrical care. The thesis encompasses two main parts. Part one relates to clinical decision making in threatened preterm labour. Part two presents four economic evaluations based on clinical trials conducted within the Dutch Obstetric Consortium. Various methods of economic analysis were used.

PART ONE: CLINICAL DECISION MAKING IN THREATENED PRETERM LABOUR

In the decision analytic analysis study described in Chapter 2 we evaluated whether implementation of the fetal fibronectin test in the Netherlands could reduce unnecessary treatments. We used a decision tree analysis, using model parameters based on currently available international literature. A cost-effectiveness analysis comparing several testing strategies, including fibronectin testing and cervical length measurement, showed that a combination of both tests can lead to major cost reductions without significantly compromising health outcomes. Performing a fibronectin test in women with moderate shortening of the cervix could lead to national cost savings in the order of 2.8 to 14.4 million Euros annually.

In Chapter 3 we present the APOSTEL-I study. APOSTEL-I is a prospective nationwide cohort study that was performed in all 10 perinatal centres in the Netherlands. Women with symptoms of preterm labour between 24 and 34 weeks of gestation with intact membranes were eligible. In all women, qualitative fibronectin testing (0.050-microgram/mL cut-off) and cervical length measurement were performed. Logistic regression was used to predict spontaneous preterm delivery within 7 days after testing. Women with predicted risk of less than 5%, corresponding to the risk for women with a cervical length of at least 25 mm, was considered at low risk. Fibronectin results and cervical length were available for 665 women, of whom 80 (12%) delivered within 7 days. Women with a cervical length of at least 30 mm or with a cervical length between 15 and 30 mm with a negative fibronectin result were at low risk (less than 5%) of spontaneous delivery within 7 days. In comparison with cervical length measurement only, an additional fibronectin test in case of a cervical length between 15 and 30 mm identifies 67 women (10%) who will not deliver within 7 days. In these women admissions, treatment and/or referral could be prevented.
In the study reported in Chapter 4 we evaluated whether withholding tocolytic treatment would increase the rate of subsequent preterm delivery in fibronectin negative women with symptoms of preterm labour and moderate shortening of the cervix (10-30 mm). We performed this discordance trial within the APOSTEL-1 cohort study. We randomly allocated 73 consenting fibronectin negative women to the tocolytic nifedipine or to placebo. We did not observe an effect of nifedipine, mainly because the outcome in the placebo group was already favourable. This was confirmed in women that were not randomised. We concluded that placebo is not inferior to nifedipine, which further strengthens the conclusions of Chapter 3.

In Chapter 5 two prediction models are presented for the outcome preterm delivery within 7 days. We used the APOSTEL-I cohort, in which we included women with symptoms of preterm labour and intact membranes between 24 and 34 weeks gestation. Because a substantial part of the data was missing we used imputed datasets. In the first prediction model only cervical length and fibronectin were included as predictors. This model is similar to the model presented in Chapter 3. In the second prediction model we also used patient characteristics in addition to cervical length and fibronectin. We used backward elimination to arrive at a parsimonious model. Patient characteristics included in the model were maternal age, previous preterm birth, parity, multiple pregnancy, vaginal bleeding, C-reactive protein, and digestion. Both prediction models had a high discriminative capability, with an AUC of 0.90 (95% CI 0.87 to 0.93) for the first and 0.94 (95% CI 0.91 to 0.96) for the second model. Women with a predicted risk of less than 5%, corresponding to the risk for women with a cervical length of at least 25 mm, were considered at low risk. Using our final multivariable model, an additional 35 (5%) women could be identified as low risk women, compared to the model with only fetal fibronectin and cervical length.

PART TWO - ECONOMIC EVALUATIONS ALONGSIDE CLINICAL TRIALS

Chapter 6 gives an overview of the neonatal costs due to premature delivery, in both singleton pregnancies and multiple pregnancies. We merged data from the APOSTEL-I study, APOSTEL-II trial, the ProTWIN trial and the Amphia trial; this represents data from 2802 women in total, with 1090 singleton pregnancies, and 1712 multiple pregnancies. Analyses were performed on a maternal level. The costs were calculated per gestational age, starting at 24 weeks, and include the neonatal costs until final discharge of the newborn in the first year postpartum or neonatal death. Adverse neonatal outcomes were also analysed per gestational age. For singleton pregnancies costs are around the €75,000 per woman delivering between 24 and 27 weeks gestation. From 27 weeks gestation costs decrease with the increasing gestational age. For multiple pregnancies costs increase between 24 and 26 weeks gestation and reach their maximum of €169,571 per woman at 27 weeks. This is because many newborns die immediately postpartum in that period, and did not generate excessive additional costs. At a gestational
age of 26-27 weeks, relatively many neonates survive with initially poor health. Preventing and delaying preterm birth after a gestational age of 27 weeks, even a minor delay of a few days/weeks, is associated with significant reduction in costs and improvement of neonatal health.

Chapter 7 presents the results of the economic evaluation of using a cervical pessary to prevent preterm delivery in women with a multiple pregnancy. The ProTWIN trial showed that insertion of a pessary in women with a cervical length below 38 mm at 16-22 weeks significantly reduced both the risk of poor neonatal outcome and of very preterm birth. Costs were estimated for health care utilisation between randomisation and 6 weeks postpartum. We separately analysed outcomes in the prespecified subgroup of women with a cervical length (CL) below the 25th percentile (<38 mm). The mean costs in the pessary group (401 women) were €21,783 versus €21,877 in the group without a pessary (407 women), for an average difference of -€94 (95% CI, -€5975 to €5609). In the prespecified subgroup of women with a CL below 38 mm the mean costs in the pessary group (78 women) were €25,142 versus €30,577 in the no pessary group (55 women); a difference of -€5436 (95% CI -€11,001 to €1456). Our analysis showed a probability of 94% that pessary treatment is the dominant strategy (more effective and less costly) in women with CL<38 mm. We concluded that treatment with a cervical pessary in unselected women with a multiple pregnancy generates comparable costs, compared to no treatment. However, a screen-treat program using a pessary in women with a CL<38mm results in better outcomes and lower costs.

Chapter 8 presents an economic analysis conducted alongside the PROBAAT trial. In this trial 411 women were induced using Foley catheter; in 408 women PGE2 gel was used. The number of caesarean sections in both trial arms was similar, with less hyperstimulation, fewer neonatal admissions to the intensive care unit and fewer cases of post-partum haemorrhage in the Foley catheter group. The cost-effectiveness analysis studied scenarios in which women are admitted to the antenatal ward or monitored as out-patient during ripening. Foley catheter and PGE2 gel inductions were found to generate comparable costs, but Foley catheter induction resulted in fewer neonatal admissions and asphyxia/post-partum haemorrhage compared to prostaglandin induction. In a scenario where a Foley catheter is used in an outpatient setting, costs could be substantially reduced in favour of the Foley catheter, by reducing the time spent in the labour ward.

Chapter 9 described a cost-effectiveness analysis alongside the HyRAS study, demonstrating that postpartum screening for cardiovascular risk factors and subsequent treatment in women with a history of term gestational hypertension or pre-eclampsia is very likely to be cost-effective. The cost-effectiveness analysis was performed using two explorative Markov models, based on hypertension screening and screening for metabolic syndrome. Data from the HyRAS study and
medical literature were used. It was shown that compared to current practice, both screening for hypertension and for metabolic syndrome in women with a history of term gestational hypertension or pre-eclampsia resulted in an increase in life expectancy (hypertension screening 0.19 years (95% CI -0.28 to 0.66); metabolic syndrome screening 0.05 years (95% CI -0.26 to 0.35)) and event free survival (hypertension screening: 0.42 years (95% CI -0.39 to 1.23); metabolic syndrome screening 0.09 years (95% CI -0.25 to 0.44)). Health gain in QALYs was limited (hypertension screening 0.04 QALYs (95% CI -0.12 to 0.20); metabolic syndrome screening 0.03 QALYs (95% CI -0.14 tot 0.19)). All incremental cost-effectiveness ratios were below the commonly used threshold of €60,000 euros per QALY gained, and the corresponding interventions can be considered cost-effective.

**IMPLICATIONS FOR CLINICAL PRACTICE AND FUTURE RESEARCH**

Based on the work presented in this thesis, we can provide some recommendations for clinical practice and future research. We performed a number of economic analyses in order to systematically compare commonly used and/or innovative interventions to improve obstetrical care. The results can contribute to clinical decision making for individual women, as well as to clinical guidelines and evidence based policy making regarding threatened preterm labour, prevention of preterm delivery, induction of labour, and cardiovascular risk assessment in women. Yet the growing need for economic evaluations in obstetrics – and probably other medical areas - is still not fully acknowledged. There is a split between the number of economic evaluations during the past decades, the way economic evaluations are performed, and the way results of economic evaluations are implemented, if implemented at all.1-4

**Timing of economic evaluations**

Economic evaluations are not performed by default alongside clinical trials. In addition economic evaluations are not seldom secondary analyses, presented after the initial results on effectiveness. Results of the original trial may suggest that a new intervention is worth introducing into clinical practice, while an economic evaluation may lead to a different conclusion.

One should consider whether an economic question is embedded within a research project before starting the clinical trial. In this process one should consider and decide upon the type of economic evaluation, the perspective, and the time horizon, all before the study starts.5 Considering methodological issues of the economic evaluation upfront allows for a proper inventory of which data, both costs and health related outcomes have to be collected during the study. This is a crucial component of a high quality economic evaluation study, especially in trial-based economic evaluations in which prospective patient level data are used. Results could be part of clinical decision making while simultaneously presented with the effectiveness results.
Long-term time horizon

A limitation of our economic evaluations is the restricted time horizon, which was used for reasons of practical and financial feasibility. In most cases, only clinical outcomes concerning the period between randomization and childbirth or hospital discharge were collected. In the field of obstetrics there is growing awareness of the importance of long-term follow-up of the child after obstetric interventions. For evaluating treatment options during pregnancy or labour, long-term consequences (and costs) for both mother and child might be relevant, and considering these is recommended by the National Institute for Health and Care Excellence in the United Kingdom. Long-term consequences can be integrated in economic evaluations in two ways: (1) by empirically evaluating long term effects and costs of a perinatal intervention through follow-up of the neonates after the trial or, (2) by extrapolating short-term trial data using modelling techniques. At present, it is not standard policy to measure long-term effects of a perinatal intervention by following-up the neonate in an obstetric trial. Long-term follow-up is time-consuming, expensive and does not fall within the funding-period of most obstetric trials. Teune et al. concluded in their review that only a small proportion of large perinatal trials report the long-term follow-up of the child. They therefore emphasize that future obstetric trials should consider performing long-term follow-up at the start of the trial. It is also mentioned that empirical long-term follow-up of infants may not always be feasible.

One could argue that short-term outcomes are often sufficient as input for decision making, especially when downstream consequences of short-term outcomes are more or less known. In some studies, such short term results may suffice, e.g. when costs and effects are in favour of the same strategy, and long-term costs will just be an extrapolation of this dominance. However, when short-term costs and effects are in pointing in different directions contradicting (or opposite findings in major health outcomes), the evaluation could benefit of a formal synthesis of long-term consequences. The question is whether actual, empirical measurements are always required or modelling will suffice in case associations between short-term and long-term outcomes are considered known. Rational decision making regarding follow-up measurements is required for an efficient allocation of the limited resources for evaluation research.

Consortia

The studies presented in this thesis were conducted within the Dutch Consortium for Research in women’s health, which is a national research network in the field of obstetrics and gynaecology with more than 50 participating centres, including teaching, non-teaching and general hospitals. Both the number and variety of recruiting hospitals, make study samples highly representative for the population of pregnant women in the Netherlands. Another major advantage of this network is the multicenter infrastructure that facilitates a standardized way of performing randomized clinical trials. This standardization includes general trial management and coordination, patient recruitment, design of research instruments, data collection and
data entry, analyses and reporting of results. This standardization also stretches to (piggyback) economic evaluation studies, as they may use (adaptations of) previous questionnaires, cost calculations and analytical procedures. Expertise in performing economic evaluations is maintained in such networks, and improves the quality of the results.

**Benefits of economic evaluations**

Economic analyses have more benefits than providing support for decisions. A recent study performed by the Dutch Obstetric Consortium showed that clinical trials performed in the field of obstetrics in the Netherlands were a good investment, as in the end they are cost-saving to society and beneficial for patients.9 Multiple trial based economic evaluations were included, and their potential health gains and budget impact were estimated, based on realistic implementation percentages. This study estimated that a potential cost saving in the order of €9.6 million per year can be achieved, when the results of eight trials are effectively implemented, with an investment of €300,000 per year over 7 years for conducting the studies.9

This budget impact study mainly focused on innovative interventions, but could be extended with diagnostic and prediction studies, like the one evaluated in the APOSTEL-I study. From this point of view economic evaluation research is not only necessary to critically appraise new as well as existing interventions to improve cost-effective care, but also a worthwhile investment.

**Implementation of results**

The Dutch Association of Obstetrics and Gynaecology (NVOG) aims to introduce guidelines to support physicians and patients to make better choices in daily practice. These evidence-based guidelines are developed based on the available evidence on effectiveness and costs of interventions. The actual content of these guidelines is defined after a consensus procedure within a professional group.10 Because of the lack of clear regulations many new technologies and interventions are introduced without adequate evaluation. Paradoxically, once the evidence from cost-effectiveness studies becomes available, it is often found hard to change practice based on these results. A good example is the introduction of the fetal fibronectin test.

The fetal fibronectin test was widely implemented internationally after a systematic review was published, concluding that the negative predictive value was high. Since a majority of women presenting with signs of preterm birth do not deliver within 7 days, and many of these will eventually deliver at term, a test could be clinically useful, if a negative result indicates that delivery within 7 days is very unlikely, allowing care providers to postpone or avoid unnecessary treatments.11

By itself, the characteristics of the fibronectin test do not guarantee that its adoption into a clinical setting will generate the theorized clinical benefits. The sometimes intricate relations between testing and subsequent treatment can obscure evaluations of the potential effect of
the test itself. For example, if clinicians ignore the test results in their subsequent management, patient outcomes will not be different as a result of testing, regardless of test accuracy or the test's potential to affect management. On the other hand, if clinicians use the test results to guide therapy decisions but subsequent therapy is not effective, or only effective for a small subgroup of patients, differences in health outcomes that could be achieved as a result of using the test may not be appreciated. A new diagnostic test should therefore be evaluated taking in account the diagnostic tests that are used in current practice (Figure 1).

In the Netherlands the fibronectin test was not implemented prior to the start of the APOSTEL-I study. This study was conducted following a strict treatment protocol, in which therapeutic choices were mandatory based on the test results. We also assessed the accuracy of the fibronectin test in addition to the diagnostics that are used in symptomatic women in current practice, including cervical length measurement and vaginal examination.

It seemed likely that the APOSTEL-I study findings and recommendations based on them could be implemented within a short time horizon. We accounted for potential biases, all 10 perinatal centres in the Netherlands participated in the study and were therefore familiar with the test, and the study results were presented at (inter)national congresses and published within a year after completion of the study. Furthermore, results were consistent with the available literature, which concluded that women with a cervical length below the 15 mm have a high risk for preterm delivery on a short term, while women with a cervical length above the 30 mm have a low risk. Women with a cervical length between the 15 and 30 mm have an intermediate risk on preterm delivery, which might be influenced by other diagnostic results. And finally the proposed decision rule is easy to use.
Unfortunately, the fibronectin test has so far not been implemented in all perinatal centres, not to mention the general hospitals. Many factors may have played a role. These include: (1) the introduction of a quantitative version of the fibronectin test was introduced during the course of the study, giving a value between 0 and 500 microgram/mL instead of a positive or negative result, (2) the fact that general hospitals did not participate in the study, and (3) the substantial variability of results across different studies and settings, a phenomenon also seen in intervention studies. This variability is not yet fully understood, and most studies are for that reason repeated in order to confirm the beneficial effects or the absence of adverse effects, before they are introduced into clinical practice.

A delay in implementing potentially (cost-)effective interventions may result in missed opportunities to reduce costs and improve health outcomes. A solution is to implement recommendations or interventions based on original studies in observational studies, for example the decision rule including the fibronectin test in case of the APOSTEL-I study, in order to establish the original findings. It is desirable that as much as possible centres participate, in order to get results fast. It has been shown that implementation of an intervention is easier in participating centres than in non-participating centres after completion of the study. In this respect, one could even think of building more international collaborative networks. An example is the one responsible for the EuFiS study, which investigates the role of the quantitative fibronectin test in combination with cervical length measurement in women with threatened preterm labour before 34 weeks of gestation.

**Concluding remarks**

In this thesis we presented data on the effectiveness, costs and cost-effectiveness of a number of diagnostic and treatment strategies in obstetrical care, which could contribute to the development of clinical guidelines and evidence based policy making. In view of the rising health care costs and clear budget constraints, economic analyses to guide decision-making are gaining importance. Professionals need to take the responsibility for mounting these evaluations of new interventions and for covering the associated costs prior to implementation. With respect to prediction of preterm birth in women with threatened preterm labour, other diagnostic tests, like the quantitative fibronectin test and the actim-partus® test, need further investigation. The developed prediction models need further validation and evaluation for their clinical and economic benefits.
REFERENCES


