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.
CHAPTER 8

SUMMARY AND GENERAL DISCUSSION
Chapter 8

SUMMARY

Pregnancy is an important life event, and a good start of the baby’s life is important for that baby and her family. Whenever a woman experiences an imbalance of her pregnancy (e.g. preterm birth, foetal distress during labour, high blood pressure), medical doctors are faced with the challenge to find a solution with the most favourable outcomes. This thesis focuses on improving evaluation research in obstetric interventions by:

- improving the outcomes used in evaluation research
- measuring long term outcomes of evaluation research
- integrating outcomes of obstetrical evaluation research in other study designs to give guidance for clinical decision making.

The thesis is divided in three parts. Part I of this thesis provides a core outcome set for obstetrical evaluation studies. In Chapter 2 the core outcome set for studies on prevention of preterm birth is presented, developed with an international e-Delphi consensus group. This core outcome set reflects the outcomes that are critically important to all relevant stakeholders (patients, obstetricians, midwives, neonatologists and researchers). One hundred and seventy-four individuals, representing five stakeholder groups, including obstetricians, midwives, neonatologists, researchers, and patients, from twenty-five countries participated in a modified e-Delphi procedure. We were able to reduce 227 outcomes identified by a systematic review of the literature and 33 outcomes suggested by participants to 13 consensus ‘core’ outcomes. This set contains four outcomes related to pregnant women: [1] maternal mortality; [2] maternal infection or inflammation; [3] preterm rupture of membranes; and [4] harm to mother from intervention. Nine of the core outcomes are related to the offspring: [1] gestational age at delivery; [2] offspring mortality; [3] birthweight; [4] early neurodevelopmental morbidity; [5] late neurodevelopmental morbidity; [6] gastrointestinal morbidity; [7] infectious morbidity; [8] respiratory morbidity; and [9] harm to offspring from intervention. Implementation of the core outcome set in future evaluation studies on preterm birth prevention will ensure that data from these trials can be compared and combined.
Part II of this thesis presents data on long-term outcomes of obstetrical intervention studies.

Chapter 3 evaluates the long-term outcomes in children born to mothers with a short cervical length that were given a pessary during twin pregnancy in a randomized controlled trial. The initial trial showed no benefit of pessary use in the overall group of asymptomatic women with a twin pregnancy, when compared to controls. However, in the women with a short cervical length (CL) at screening (<38mm), short-term benefit (measured as a composite outcome of neonatal mortality and morbidity, including prolongation of pregnancy) was found in the pessary group. In this follow-up study, long-term survival and neurodevelopmental outcomes were measured at three years corrected age in the children born to mothers with a short cervix who were offered a pessary (n=157) compared to controls (n=111). After three years 27 children had died, 6 (5%) in pessary vs 21 (26%) in control group, adjusted OR [95% CI] 0.14 [0.04 to 0.50]. To assess neurodevelopmental outcome Bayley-III scores were collected for 173 (72%) out of 241 surviving children (114 (75%) pessary vs 59 (66%) control group). The cumulative incidence of death or survival with a neurodevelopmental disability was 12 (10%) vs 23 (29%) in the pessary and control group, respectively, aOR [95% CI] 0.26 [0.09 to 0.75]. Neither statistical nor clinically relevant differences in Bayley-III scores between both groups were found. We concluded that in women with a twin pregnancy and a CL < 38 mm, the use of cervical pessary strongly improved survival of children without affecting neurodevelopmental disability at three years corrected age.

Chapter 4 evaluates the long-term neurodevelopmental and physical outcomes of the children born to mothers with a short cervical length (CL <30mm) in a singleton pregnancy. These women were included in a randomized controlled trial (Triple P) comparing the use of vaginal progesterone with placebo in the second and third trimester to prevent preterm birth. The initial trial showed no benefit of progesterone in the short-term but was underpowered. Due to the lack of long-term outcome data after the use of progesterone in pregnancy, and its safety being recently questioned, follow-up of this trial gives valuable information. This chapter reports on survival, as well as neurodevelopmental, neurological and physical outcomes of the trial at two years corrected age of the children. Of the 77 surviving children in the Triple P trial, 59 (77%) children were reached for follow-up of whom 57 (n=28
progesterone vs n=29 placebo) completed a Bayley III test. Neither statistical nor clinically relevant differences in Bayley-III scores between both groups were found. Congenital malformations were seen in 8 (30%) and 2 (11%) children in the progesterone and placebo group respectively RR [95% CI] 4.0 [0.93 to 17.1]. No differences in genital abnormalities and neurological examination were seen between both groups. We concluded that in low risk women with a short cervix the prescription of progesterone in second and third trimester is not associated with moderate neurodevelopmental delay at 2 years corrected age but that the difference in (minor) congenital anomalies should be further explored. Although the sample size of this follow-up study is too small to pick up small differences between study groups. These data can contribute to future meta-analyses that must answer whether progesterone is a safe drug to use during pregnancy.

Part III of this thesis focusses on the integration of outcomes derived from obstetrical evaluations into a systematic review/meta-analysis, cost-effectiveness analysis and budget impact analysis in order to give guidance for clinical decision making. Chapter 5 provides an overview of the existing literature to evaluate the predictive value of brain MRI results for long-term developmental outcomes in children born preterm or with a low birth weight. The study descriptively aggregates the results of 20 papers. The prognostic accuracy of MRI, performed at term equivalent age and evaluating the presence of moderate to severe white matter lesions, was found to be highest for prediction of cerebral palsy with a sensitivity of 51% and specificity of 93%. Its ability to predict other long-term outcomes such as neurocognitive and behavioural impairments is limited. We concluded that routine use of MRI in clinical practice is not recommended due to its moderate predictive value. However, routine use of MRI in a research setting with adequate and uniform recording of the data can help to generate future evidence on its prognostic capacity.

As most clinical studies only collect outcomes at short-term, long-term data is lacking in >80% of large evaluation studies in obstetrics.1 Especially for interventions that target long-term outcome improvement, the lack of long-term outcome data constitutes a blind spot for clinical decision-making. This is the case in studies that evaluate interventions to prevent perinatal asphyxia.
Asphyxia is a clinical condition of impaired oxygen supply or blood flow to the foetus and can occur before the onset of labour and during labour. This outcome is known to have immediate consequences, but can also lead to long-term neurological impairment. In Chapter 6 we used the data of an individual patient participant data meta-analysis that assessed the use of ST-analysis in electronic foetal monitoring (STAN) which aims to reduce neonatal asphyxia during labour. In a cost-effectiveness analysis two models were created: one model from a maternal perspective and the other from a neonatal perspective. Costs and effects at short and long-term were evaluated for women and children who were monitored with STAN compared to electronic foetal monitoring (EFM) only. Results from the neonatal model showed that the STAN strategy reduced metabolic acidosis (asphyxia) from 1100 to 900 per 100 000 newborns at an additional cost of € 14 509 to prevent one case of metabolic acidosis. In the maternal model we found a reduction of instrumental deliveries of 1.5% in favour of STAN. The cost to prevent one instrumental delivery was estimated at € 2602. The results of the long term benefit of STAN is very much depended on the association between short term metabolic acidosis and long term cerebral palsy (CP). Evidence on this association is very heterogeneous, providing a variation of probabilities between 0.5 to 9.3%. Explorative analysis showed that STAN becomes a cost-saving strategy if ≥1.3% of children exposed to metabolic acidosis acquire CP. This study therefore suggests that STAN, when compared to EFM alone, can be a cost-effective strategy from both maternal and neonatal perspective on the short term, and is potentially cost-saving on the long term.

Finally, by performing evaluation studies of obstetrical interventions our ultimate goal is to improve health outcomes of mothers and their children at acceptable costs. Therefore, implementation of trial results is a crucial step. Chapter 7 presents the predicted health and financial impact of the implementation of eight nationwide evaluation studies in obstetrics. The potential budget impact of the individual studies in terms of costs and effects was extrapolated to the situation in the Netherlands. When the results of these eight studies are implemented, a beneficial effect on health outcomes can be expected in: (1) women suffering from pregnancy induced hypertension and mild pre-eclampsia at term; (2) women in whom labour is induced and; (3) women with foetal monitoring by STAN analysis. De-implementation of non-
effective practices such as (4) prolonged tocolysis; (5) intra uterine pressure catheters; (6) progestagens to prevent preterm delivery in twins; (7) immediate induction of labour in preterm prelabour ruptured membranes; and (8) induction of labour at 35 completed weeks of gestation in intrauterine growth restriction, reduces costs. The potential cost reduction was estimated to be € 9.6 million per year on the basis of a one-time investment cost of € 3.1 million for the eight evaluation projects. We concluded that the financial and health benefits of useful clinical research more than offset the costs of performing it.

GENERAL DISCUSSION

In perinatal medicine, RCTs can help us to identify the best policies and interventions. A well-structured research question following the PICO(T) criteria is an important ingredient of a RCT. Systematic reviews and meta-analysis provide an overview of clinical trials on a topic and in doing so provide the highest level of evidence that is meant to give guidance for clinical decision making (Figure 1, introduction section).

There are, however, some problems in the design of randomised clinical trials and other clinical evaluation studies that hamper the usefulness of clinical research. First, there is a lack of standardization in the selection and operationalization of outcomes. This may lead to inefficiency in research and waste of resources. Second, in more than 90% of large clinical trials in obstetrics there is lack of follow-up outcomes. This can imply a blind spot in clinical practice, as not all effects (benefit and harm) become apparent on the short term. The Dutch famine study has provided examples for this phenomena. Finally, we are facing a gap between clinical research and its impact on clinical decision making. This thesis focused on each of these three problems, all being barriers to optimally use clinical research in order to improve the health of a pregnant woman who faces a problem in her pregnancy. Using different methodologies, we have explored several strategies that aim to improve evaluation research in obstetric interventions.
I Towards standardisation of outcome measurement in obstetrical evaluation research

At present the RCT is considered the optimal design to answer questions about the effectiveness of clinical interventions. There are standards for what should be addressed in the protocol (i.e. SPIRIT guideline7), standards for the conduct of the trial (i.e. ICH-Good Clinical Practice guidelines8) and on what should be addressed in the final report (i.e. CONSORT guideline9). Also, registration in a public trial registry at or before the onset of patient enrolment is a mandatory practice (e.g. ClinicalTrials.gov, WHO registry). One of the goals of these standard practices is the reduction of ‘waste’ across medical research, a phenomenon that has been estimated as consuming 85% of the billions spent on medical research each year.10 One of the items frequently mentioned in the literature exploring the different sources of research ‘waste’ is the use of inadequate outcomes in research. The outcomes that researchers have measured have not always been those that patients regard as most relevant11 and the variety of outcomes used hampers the comparison and meta-analysis of results. In preterm birth research for example, 72 different primary outcomes were reported in 103 clinical trials and 29 different outcomes in 33 Cochrane reviews.12

After international consensus has been achieved on how to perform and report on good quality trials, a next step is to establish international consensus on what outcome measures make trial results more suitable for clinical decision making. This could be achieved by defining so called ‘core outcome sets’; a minimal set of outcomes that relevant stakeholders consider as critical to evaluate interventions for specific health conditions.13

The idea of core outcome sets already came to practice in the 90ties, when a group of rheumatology researchers started to develop core outcome sets for rheumatology related health problems (Outcome Measures in Rheumatoid Arthritis Clinical Trials- OMERACT).14 This work has led to an increase in research effectiveness in this field due to the availability of comparable data and the possibility to pool data from different studies.15 The OMERACT team also shared their experience and stressed the importance of patient involvement in this process.11

There are now several international initiatives that support the idea of core outcome sets and deliberate how to improve its methodology and dissemination. The ‘Core outcome sets in effectiveness trials’ (COMET) initiative
launched in 2010 aims to foster methodological research in this area by publications on core outcome set methodology and organizing yearly scientific meetings (www.comet-initiative.org). It also developed a publicly available searchable database of completed and ongoing projects in core outcome set development. This will prevent duplication of core outcome set projects and inspire new relevant core outcome set projects. The ‘Core Outcomes in Women’s and Newborn Health’ (CROWN) initiative is led by journal editors, to harmonise outcome measurement and reporting in women's health research.\textsuperscript{16}

This consortium of 76 women’s health journals aims to encourage researchers to develop core outcome sets using robust consensus methodology and to organize peer-review and effective dissemination of manuscripts describing core outcome sets. By facilitating the dissemination of core outcome sets, the final goal is to improve synthesis of evidence to generate recommendations for clinical practice.

To date, the first core outcome set for preterm birth studies endorsed by CROWN has been published (chapter 2) together with summary publications in several journals and languages.\textsuperscript{17,18} At present, 23 ongoing Core outcome set projects on pregnancy and childbirth are registered in the COMET database, amongst them core outcome sets for hypertensive disorders in pregnancy, pre-eclampsia, gestational diabetes, intrauterine growth restriction, postpartum haemorrhage, pain management in labour, and hyperemesis gravidarum (www.comet-initiative.org).

Future implications

This work on Core outcome sets will only pay off if these sets are implemented in future study protocols and trials. We would like to suggest a roadmap on how to move forward:

1) Researchers, reviewers and guideline-developers experiencing the lack of Core outcomes in a specific research area targeting a specific population, need to address this. Either by the initiation of a Core outcome set project together with multinational key stakeholders, or by recommending this in their research- or review manuscript and guideline protocol. Guidelines on how to develop a proper Core outcome set project are needed.\textsuperscript{19}

2) Researchers and clinicians involved in national data registries can implement the Core outcome set in this registry.
3) Researchers and reviewers should incorporate the set of Core outcomes in their research protocol and motivate when they are not able to collect all Core outcomes of the set.

4) Journal reviewers should encourage authors to report all core outcomes included in the particular Core outcome set of a specific health area. Eventually, this could even be mandatory, just like reporting of trial registration number and ethical approval. The CROWN initiative is preparing a simple guideline for reviewers about Core outcome sets.\textsuperscript{16}

Improving awareness and dissemination of a Core outcome set will hopefully increase the amount of comparable data incorporating relevant outcomes. This will contribute to the reduction of unnecessary duplication of randomized trials and meta-analyses with the same research question and provide a valuable source of information for clinical guidelines.

II Evaluation of long-term outcomes in perinatal trials

Many interventions applied in pregnancy are evaluated for their efficacy and safety by measuring short-term maternal and neonatal outcomes only. There are numerous examples in the literature on evaluation studies showing either short-term benefit or no benefit, but remarkable long-term harm\textsuperscript{20,21} or warning signs for long-term harm.\textsuperscript{22,23} Performance of follow-up measurements of obstetrical evaluation studies is therefore pivotal. In chapters 3 and 4 we evaluated the long term follow-up of two RCTs. In chapter 3 we performed a follow-up study of children born to mothers that used a cervical pessary to prevent preterm birth in twin pregnancy. As positive effects of pessary on pregnancy prolongation and improvement of neonatal outcome only had been seen in women with a short cervix, we limited follow-up to that group. This choice was partly financially driven. With the available resources we acknowledged that follow-up in the group of women with the potential short term effect was most needed as potential long-term effects of the use of pessary in pregnancy (including harm) were unknown. We also calculated upfront that we had enough power to detect possible clinically important differences. This was different for the follow-up study performed in chapter 4. Here we were confronted with an underpowered RCT evaluating the use of progesterone to prevent preterm birth in women with a short cervical length at screening. After
reviewing the available literature, we were surprised about the small amount of long-term data on progesterone use in second and third trimester of pregnancy. Before the start of this follow-up study only two studies were published, both using parental questionnaires.\textsuperscript{24, 25} This confirmed that this follow-up study was highly needed due to this blind spot of long-term information of this widely used drug. New data on potential long term effects of progesterone published in the last year demonstrated that indeed there are some concerns related to long term health. The OPPTIMUM trial described an increase risk (although still with 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]).\textsuperscript{23} The follow-up of the PREDICT trial also reported an 8-fold increased risk of a cardiac problem in children exposed to progesterone at 8 years of age.\textsuperscript{26}

**Challenges in design**

To date there are no international standards on whether and how to perform follow-up of (obstetrical) trials. This can be explained by the fact that the design of the follow-up is related to the questions addressed in the initial trial and the potential harms expected according to pathophysiological reasoning.\textsuperscript{27} However, as explained in the above section about standardising outcomes, there is a huge potential gain in consistency in outcomes used; this implies short term outcomes, but also long term outcomes. Consistency in timing of follow-up and assessment methods will help in achieving follow-up data that can also be compared or pooled across studies. In the Netherlands a national working group is addressing the need for a national structured follow-up of all children that were discharged from a neonatal intensive care unit. In 2000, this group published recommendations on how to perform standardized follow-up in these children (Table 1).\textsuperscript{28} Just recently, this has been implemented in clinical practice. With this collaborative effort, more compatible data will be generated.
Summary and general discussion

Table 1. Recommendations of the national working group of neonatal follow-up in children discharged from Neonatal intensive care born at gestational age <30 weeks and/or birthweight <1000gr and/or birthweight <1500gr if below tenth percentile28 29

<table>
<thead>
<tr>
<th>Corrected age at follow-up</th>
<th>Type of assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>Neurological assessment and general physical examination by paediatrician</td>
</tr>
<tr>
<td>12 months</td>
<td>Neurological assessment and general physical examination by paediatrician</td>
</tr>
<tr>
<td>24 months</td>
<td>Neurological assessment and general physical examination by paediatrician. Bayley scales of infant and toddler development (Bayley) test30 31 Child Behaviour Checklist32</td>
</tr>
<tr>
<td>5 years</td>
<td>Neurological assessment and general physical examination by paediatrician Neurocognitive assessment (including IQ, language, executive function, visuomotor assessment) Movement ABC33 Child Behaviour Checklist</td>
</tr>
<tr>
<td>8 years</td>
<td>Neurological assessment and general physical examination by paediatrician Wechsler preschool and primary scale of intelligence (WPPSI) test34 Movement ABC Child Behaviour Checklist</td>
</tr>
</tbody>
</table>

A second problem of follow-up studies concerns that available follow-up tools/tests are sometimes not capable to detect subtle, but clinically relevant differences. Screening tools (i.e. questionnaires like ASQ) instead of diagnostic tools (i.e. Bayley test) are used with poor predictive value for long-term (neuro) development.35 Furthermore, due to logistic and financial reasons, it is hardly feasible to cover the whole spectrum of long-term developmental or healthcare related outcomes. A follow-up interpretation is therefore always restricted to the type of test and timing of test chosen.36

Third, a follow-up study faces the problem of loss to follow-up. This can be due to mortality, but there is a high risk of attrition that causes potential selection bias in the results and complicates data analysis to adjust for such bias.37

Challenges in interpretation of long-term follow-up data

The need for follow-up should not be misinterpreted with the need to prove effectiveness on the long-term.36 The reason for this is that some short-term outcomes indeed do not have long-term consequences, but can still be relevant for the first period of life. An example is the difference in interpretation of the follow-up results of the ORACLE I by two medical specialist organisations. In the original study, women with preterm rupture of membranes without signs...
of clinical infection were randomized to erythromycin or placebo. The study concluded that erythromycin decreased the risk of the primary outcome, a composite defined as death or major cerebral abnormality or chronic neonatal lung disease (11.2% vs 14.4%, p=0.02).38 The 7-years follow-up study, however, found no difference in the proportion of children with any functional impairment after prescription of erythromycin compared to placebo (38.3% vs 40.4%, OR [95% CI] 0.91 [0.79 to 1.05] and concluded that the prescription of antibiotics for women with preterm rupture of membranes did not show an persisting effect on the health of children.39 The Dutch society of obstetrics and gynaecology (NVOG) probably interpreted these findings as ‘antibiotics do not show an effect on the long-term in this group of women, and are therefore not useful’ by not recommending the use of antibiotics in its guideline using the ORACLE I follow-up in its references.40 The British society of obstetrics and gynaecology (RCOG) probably interpreted these findings as ‘antibiotics do show a relevant short-term effect, and do not show harm on the long-term, and therefore are useful in this group of women’ by recommending the use of antibiotics in its guideline.41 Follow-up outcomes should therefore not always be regarded as a proof of efficacy, rather as a proof of safety.36

**Future implications**

So long term follow-up is important to remove potential blind spots in clinical research by evaluating the full scope of potential effect and harm. However, realising this is challenging, because of financial, logistical and time-restraints, as well as to appropriately design and interpret the results of follow-up studies. When long-term follow-up cannot be achieved in clinical trials, we also can rely on long-term outcomes of cohort studies. A combination of short-term outcomes obtained from trials and long-term outcomes obtained from cohort studies might be the best achievable combination to assess the effectiveness and safety of a treatment.

Furthermore, working towards a standardization of follow-up will help to generate more consistently documented long-term follow-up data. Initiatives as the recommendations made by the national working group of neonatal follow-up (Table 1) will enhance this. Also the Dutch consortium for health evaluation in obstetrics and gynaecology is currently working on standardization and feasibility of follow-up assessment (e.g. using electronic questionnaires and
a mobile-bus as test location) for the randomized controlled trials performed within this consortium (www.studies-obsgyn.nl). This will help to generate long-term follow-up data that allows comparison and pooling across studies. Subsequently this framework might be a first step towards the development of a core outcome set for follow-up studies of obstetrical interventions.

III  Towards integration of outcomes of evaluation studies to guide clinical decision making

There is no such thing as a perfect study. However, a well thought-out, well-designed, appropriately conducted and analyzed clinical trial is an effective tool to generate valid and clinically relevant evidence. On the other hand, poorly designed and conducted trials can be misleading. Also, without supporting evidence no single study ought to be definitive. Therefore, integration of outcomes of well-designed evaluation studies can be powerful in providing crucial information for clinical decision making. In chapter 5 the method of meta-analysis is used to aggregate data from cohort studies to assess the prognostic value of term MRI in premature born infants on long term developmental outcomes. The work in this chapter started from a clinically based research question and therefore has a higher likelihood of being implemented in clinical practice. Furthermore, the Cochrane collaboration approached our research team to repeat this work within the Cochrane framework.

The method of cost-effectiveness analysis allows us to model the expected benefits and costs of two or more interventions in order to determine whether the expected results of an intervention are ‘worth’ the added costs. In chapter 6 we used data from an individual patient data (IPD) meta-analysis to model the potential benefit and costs of an intervention of ST-analysis in foetal monitoring during labour (STAN) compared to foetal monitoring only. We concluded that STAN can be a cost-effective strategy for both mother and child.

Finally, In this era when health-care budgets are constrained, the underlying goal of public health care allocation decisions is to attain maximal health benefit for a given budget. The ultimate goal of evaluation research is to maximize health benefit. When an intervention is found to be effective, this presumably results in a population health benefit when implemented in clinical practice. An example is the implementation of policies to discourage tobacco use during pregnancy on perinatal health in various populations (e.g. rates
of stillbirth, neonatal mortality, preterm birth an low birth weight).\textsuperscript{43,44} When an existing intervention is found to be ineffective, the de-implementation in clinical practice might reduce costs to a comparable health benefit.

The method of budget impact analysis allows us to estimate the potential impact of implementation and de-implementation of interventions on budget and health. We used this method to evaluate the potential impact of implementation of the results of obstetrical evaluation studies performed within the Dutch obstetrical consortium on national health and budget (chapter 7). Strikingly, the potential cost reduction was estimated to be € 9.6 million per year on the basis of a one-time investment cost of € 3.1 million for the eight evaluation projects. So, potential cost reduction for the healthcare budget driven by evidence from clinical research more than offsets the costs of performing this research, a conclusion in line with earlier findings.\textsuperscript{45}

This paper also contributed to a debate that healthcare insurance companies should consider whether they also have a role to invest in evaluation research,\textsuperscript{46} as healthcare insurance companies are the first to profit from cost reduction by implementation of effective interventions and de-implementation of ineffective interventions.

**Future implications**

Studies should be designed to provide an optimal contribution to the body of evidence to enable patients, clinicians, and decision makers to be confident about the magnitude and uncertainties of benefits and harms, and these studies should be judged based on clinical impact and their ability to change practice. Ideally, studies that are launched should be clinically useful regardless of their eventual results.\textsuperscript{47} Future development should therefore focus on proper study design, standardised assessment of short and long term outcomes and international collaboration with the ultimate goal of solid aggregation of data to guide clinical decision making. An Individual Participant Data (IPD) meta-analyses is a design containing all these key elements. IPD meta-analysis can provide additional relevant results by using patient level information, thus allowing evaluation of sub-groups and individualized approaches to health care delivery. An IPD can be performed retrospectively (when all studies are already completed) or prospectively (at the starting or recruitment phase of several studies). Ideally, we should all prospectively collect data anticipating future use within an international IPD collaboration. This will allow optimal
use of research data to generate and improve the available evidence to enable patients, clinicians, and decision makers to be confident about the effects and possible harm of clinical interventions.
REFERENCES


32. Achenbach T, Rescorla L. ASEBA Child Behavior Checklists for Ages 1.5-5 Years (CBCL/1.5-5) ASEBA.