Randomized controlled trials in reproductive medicine
Braakhekke, M.W.M.

Citation for published version (APA):
Braakhekke, M. W. M. (2017). Randomized controlled trials in reproductive medicine: Disclosing the caveats
Chapter 1

Introduction and outline of the thesis
Subfertility, defined as failure to conceive after 12 months of unprotected intercourse, affects up to 1 in 10 couples that aim to get a child and manifests itself as an acute and unanticipated life crisis (Gnoth et al., 2005). Subfertility creates overwhelming stress and tests coping strategies as it is unforeseen and lasts for an undetermined length of time (Wirtberg et al., 2007). The inability to conceive carries a high psychological burden and couples may exceed their physical and mental boundaries to achieve the desired pregnancy.

Reproductive medicine aims to assist these couples. This would be a trivial pursuit, if there was certainty on how to effectively assist sub fertile couples. However, many clinical decisions are made in uncertainty because of a lack of comparative data on effectiveness (Farquhar et al., 2015).

Such data on comparative effectiveness of treatments can be obtained by various research methods. These methods range from case reports to systematic reviews with meta-analysis. Reproductive medicine is the clinical specialty with the smallest, convincing, observational, case report in the medical literature: On 25 July 1978 Louise Brown, the first IVF baby, was born. Her mother, Lesley Brown underwent a procedure, later known as IVF, because she was sterile since her – diseased – fallopian tubes had been surgically removed (Steptoe and Edwards, 1978).

Groundbreaking developments like these are rare; more often we have to deal with small signal to noise ratios. This ratio reflects the difference between the course of the disease after the treatment started -the signal- relative to the natural course of the disease that was to be expected without the intervention -the noise- that can be observed outside a clinical trial (Glasziou, 2007). An example of a small signal to noise ratio is pregnancy rates after IVF in couples with unexplained subfertility, since these couples have considerable chances to conceive spontaneously (Brandes et al., 2011). To be able to make clinical decisions in case of small signal to noise ratios we are dependent on evidence obtained via empirical research.

Austin Bradford Hill was one of the first who pioneered the concept of empirical research. In 1948, he was statistician on the Medical Research Council that performed the first randomized controlled trial in humans, evaluating the use of streptomycin in tuberculosis (Marshall et al., 1948). In 1951 he started the British Doctor Study, the
first prospective observational cohort study to investigate the relationship between smoking and lung cancer. This study was probably the first to herald a new type of scientific research on showing the relevance of epidemiology and medical statistics (Doll and Hill, 1956). In 1965 he presented the Bradford Hill criteria, nine criteria that can be used to determine causality between a specific factor and a disease (Hill, 1965). These criteria form the basis of evaluation research until today.

In the mid 80-ies David Sackett introduced the concept of Evidence-Based Medicine (Sackett et al., 1985, Guyatt, 1991). Evidence-Based Medicine aims to de-emphasize intuition and clinical experience as sufficient ground for clinical decision-making and stresses the appraisal of evidence from clinical research next to pathophysiologic reasoning. In case of small signal to noise ratios, the highest level of evidence in clinical research would be generated by aggregated evidence of randomized controlled trials in which participants are randomly allocated to either the treatment under study or the control treatment (Guyatt et al., 2000).

The Cochrane Collaboration aggregates comparative information provided by randomized controlled trials in a systematic and standardized way. In our field, the Cochrane Gynaecology and Fertility Group - formerly the Menstrual Disorders and Subfertility Group - was registered in June 1997. This group produces systematic reviews of all relevant randomized controlled trials on interventions for women with menstrual disorders and for subfertile couples. In doing this, it has transpired that the Cochrane Reviews frequently conclude that there is insufficient or limited evidence of effectiveness and that further research is needed (Johnson et al., 2003). In December 2016 there were 204 reviews in the database of the Cochrane Gynaecology and Fertility Group, of which 156 reviews (76%) made that conclusion. It follows that the clinical research we perform until today is not reaching one of its goals, a medicine that is supported by good data.

Clinical research that is not providing us good data can be regarded as waste. Waste across medical research can occur at several stages and it is estimated at 85% of the total investment in biomedical research (Chalmers and Glasziou, 2009)

Several factors contributing to this waste have been identified. In 2014, the Lancet published a series in which waste across medical research is summarized by identifying five factors contributing to this waste: first, research priorities which are not well set,
weaknesses in design, conduct and analysis of trials; second, increasing difficulties in obtaining regulatory and governance approval; third, incomplete reporting of outcome measures or reporting of unusable outcome measures; fourth, complete background information about studies and fifth, inaccessible datasets and non-published research (Chalmers et al., 2014, Chan et al., 2014, Glasziou et al., 2014, Ioannidis et al., 2014, Salman et al., 2014).

All in all, as John P. A. Ioannidis, a key player in pointing out why clinical research is not good enough to reach its goals, has already stated, it is more likely for a research claim to be false than true (Ioannidis, 2005).

**BACKGROUND OF THIS THESIS**

In this thesis, we aim to highlight some major caveats in the reporting of randomized controlled trials in reproductive medicine related to various factors contributing to waste in medical research; research priorities, study design, reporting of outcome measures and access to complete information on randomized controlled trials. In our opinion, these factors are the ones that can be influenced by researchers themselves and therefore deserve our primary attention. We will discuss various ways on how to reduce waste and create useful clinical research resulting in trusted evidence and informed decisions.

**Research Priorities**

In reproductive medicine research priorities are not well set. New promising techniques are not prioritized to go through randomized controlled trials to test effectiveness. Time lapse monitoring for embryo development is an explicit example. It was introduced in clinical practice after a few observational studies showed a positive effect, without evidence from a well powered randomized controlled trial that assessed safety and effectiveness (Armstrong et al., 2015). On the other hand, treatments already shown to be non-effective, like homeopathy and acupuncture, are repeatedly investigated in randomized controlled trials. In our opinion, analyzing the biologic plausibility behind a possible effect of a new treatment is the way to decide which level of evidence is needed. In the **second chapter** of this thesis we use the concept of biological plausibil-
ity to ponder under what conditions a randomized controlled trial should or should not be performed.

**Study design**

One factor topical in the design of randomized controlled trials in Reproductive medicine is the number of treatment cycles that is analyzed. Just as for natural conception, in medically assisted reproduction cumulative pregnancy rates rise with additional cycles (Smith *et al.*, 2015). One treatment cycle can therefore not be seen as independent and effectiveness can only be assessed when multiple cycles and - in some instances, depending on the research question - even multiple treatments are reported (Daya, 2003). We assessed in chapter 3 whether the design of multiple cycles is used in randomized controlled trials on assisted reproductive techniques (ART), and which trial characteristics are determinants in doing so.

**Reporting of outcome measures**

In 2003, during a European Society for Human Reproduction and Embryology (ESHRE) consensus meeting, it was agreed that the outcome measure for effectiveness in Reproductive Medicine should be singleton live birth, and that multiple pregnancies were thereby an outcome measure for safety (Land and Evers, 2003). This agreement fuelled a fundamental debate series on what would be the most relevant standard of success in assisted reproduction. The debate never led to new recommendations and as such it would seem that singleton live birth is currently the accepted outcome measurement. To find out whether singleton live birth is really reported as outcome measurement, we assessed the reporting of singleton live birth and neonatal and maternal morbidity and mortality in randomized controlled trials on in vitro fertilization (IVF) and intracytoplasmic semen injection (ICSI) in chapter 4.

In the following chapters, 5 and 6, we reflect on the question whether singleton live birth is indeed the most relevant primary outcome measure in Reproductive medicine. We reflect on the purposes of a primary outcome measure in order to enlighten importance the choice of the primary outcome. We also discuss the advantages of an independent assessment of safety and effectiveness instead of a combination of those two in a combined primary outcome measure like ‘healthy singleton live birth at term’ in order to be able to translate research into clinical practice.
Introduction and outline of the thesis

Accessibility of complete datasets

In reproductive medicine a wide variety in outcome measures is reported, making research susceptible for selective outcome reporting. Selective outcome reporting occurs if some outcomes are reported while others are not, which potentially renders published results misleading (Higgins and Green, 2011). Full information on all possible research outcomes could be used to correct for this, but is often not accessible. One of the drivers behind selective outcome reporting is sponsorship. Sponsorship by industry is known to influence valid interpretation of trial results by more often reporting on outcomes favourable to the sponsor’s products than publicly funded trials (Lundh et al., 2012). In chapter 7 we aimed to assess if randomized controlled trials on IVF and ICSI are subject to selective outcome reporting and if this is related with sponsorship.

Translation of research findings into patient information

Next to complete information and data access for researchers and clinicians, patients also have the moral right to receive complete information and data. Patient information is currently disseminated on a large scale by fertility clinics via the Internet. The question then arises whether this information is unbiased. In chapter 8 we assessed if the information provided on websites of fertility clinics in Europe is consistent with respect to overall quality and the reporting of success rates.

OUTLINE OF THIS THESIS

In chapter 2 we discuss the role of biological plausibility in the decision making to start a randomized controlled trial.

In chapter 3 we assess whether randomized controlled trials (RCTs) on In Vitro Fertilization (IVF), intrauterine insemination (IUI) and Ovulation induction (OI) report multiple cycles of treatment.

In chapter 4 we assess the reporting of singleton live birth and neonatal and maternal morbidity and mortality in RCTs on IVF and intracytoplasmic semen injection (ICSI).
In chapter 5 we reflect on the most relevant primary outcome measure in RCTs in reproductive medicine.

In chapter 6 we explain why safety and effectiveness best can be independently assessed in RCTs in reproductive medicine.

In chapter 7 we assess if randomized controlled trials on IVF and ICSI are subject to selective outcome reporting and if this is related with sponsorship.

In chapter 8 we evaluate if the information provided on websites of fertility clinics in Europe is consistent with respect to overall quality and the reporting of success rates.

In chapter 9 we provide a summary of this thesis and the implications for further research.
REFERENCES


Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Metin Gulmezoglu A, Howells DW, Ioannidis JPA, Oliver S. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156-165


Daya S. Pitfalls in the design and analysis of efficacy trials in subfertility. *Hum Reprod* 2003;18:1005-1009

Doll R and Hill AB. Lung cancer and other causes of death in relation to smoking. *BMJ* 1956;nov:1071-1081


Chapter 1


Ioannidis JPA. Why most published research findings are false. *PloS Med* 2005;*8*:e124

Ioannidis JPA. How to make more published research true. *PloS Med* 2014;*11*:e1001747


