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 9

Summary and implications for the future
SUMMARY

Reproductive Medicine aims to assist subfertile couples in a period of overwhelming stress. Unfortunately, many clinical decisions in reproductive medicine are made in uncertainty because of a lack of data on effectiveness of existing interventions. Without knowledge on effectiveness it is impossible to provide guidance.

Data on effectiveness can be obtained by various research methods, like case reports, cohort studies or randomized controlled trials (RCTs). The choice for a specific method depends on the research question and the a priori chance of effect. Modern Reproductive Medicine started with a convincing, observational n=1 study. 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. When this success was followed by multiple successful attempts in similar patients, IVF was quickly adopted in clinical practice and became an established treatment.

Groundbreaking developments like these are rare. More often, the differences between the course of the disease after treatment and the natural course of the disease are small. These differences may be expressed by signal to noise ratios, where the signal represents the course of the disease after treatment and the noise the natural course of the disease. 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 and pregnancy rates after IVF do not reach 100%. In these cases randomized controlled trials are needed to study effectiveness. The Cochrane Collaboration, in our field the Cochrane Gynaecology and Fertility Group, aggregates comparative information provided by RCTs in a systematic and standardized way. The problem however is that most Cochrane reviews conclude that there is insufficient or limited evidence of effectiveness and further research is needed. This is no small matter, since this means that the clinical research we perform up to today is not reaching one of its goals, ie a medicine that is supported by good data.

In this thesis, ‘Randomized Controlled Trials in Reproductive Medicine – Disclosing the Caveats’, we addressed some major caveats in the reporting of randomized controlled trials in reproductive medicine related to various factors contributing to waste in medical
Chapter 2 tries to make explicit a line of thinking that we can use to decide under what conditions it is ethical to execute a RCT. We dissect the basic reasoning behind the concept of equipoise, a key concept for underpinning the ethics of RCTs. We make the argument that the interaction of pathophysiological mechanisms and observational data is of paramount importance in assessing whether one is in equipoise or not.

We delineate three scenarios based on the biological plausibility behind the effect of an intervention. If biologic plausibility is known, a few clinical observations of effect can be sufficient to provide evidence of efficacy of a treatment. In this scenario it is unnecessary and unethical to perform a RCT. If biologic plausibility is absent, the a-priori chances that this treatment is effective are low, and it is unlikely that a meaningful difference is detected by a RCT. Such trials should not be performed. If biological plausibility is uncertain it depends on observational data whether a RCT is necessary. If observational data show no effect or a positive effect we have reached equipoise. In this scenario, interventions should be subjected to RCTs to provide evidence of their effectiveness.

This line of thinking may help in setting research priorities, preventing a priori negative trials, reducing research waste, and aiding in the design of meaningful ones.

In chapter 3 we assess whether randomized clinical trials on In Vitro Fertilization (IVF), intrauterine insemination (IUI) and Ovulation induction (OI) are designed to evaluate cumulative pregnancy rates by reporting on multiple cycles of treatment. These data are relevant for clinical decision making, because, as well as for natural conception, in medically assisted reproduction cumulative pregnancy rates rise with additional cycles. Getting pregnant takes time.

Of all 253 RCTs, 48 (19%) reported on multiple cycles. Reporting of multiple cycles was significantly more common in trials on IUI and OI compared to trials on IVF. Our analysis shows that the majority of randomized clinical trials, especially those on IVF, do not report cumulative pregnancy rates. The clinical significance of these trials is thus limited, and they add little value to a medicine that is supported by good data.

In chapter 4 we present a review of the RCTs on interventions in subfertile couples included in Cochrane reviews until December 2012. To improve and harmonize international registration of the outcomes of fertility treatments, the European Society of Human Reproduction and Embryology (ESHRE) in 2003 recommended the reporting of ‘singleton live birth’ as outcome measure of effect and to report neonatal and maternal outcomes as well.

We assessed whether or not this recommendation was followed by the RCTs included in the Cochrane reviews. Since the recommendation was done in 2003, we compared trials published before- and after 1-1-2004. We identified 910 RCTs that reported on fertility treatments, of which 182 RCTs (20%) reported on ‘singleton live birth’ (before 1-1-2004 96/518 (19%); after 1-1-2004 86/392 RCTs (22%). ‘Singleton live birth’ was the primary outcome in 68 RCTs (7.4%). Only 44 RCTs (4.8%) reported on neonatal outcome, and 52 RCTs (5.7%) reported on maternal outcome.

In conclusion, the use of the outcome measure ‘singleton live birth’, as recommended by ESRHE, is not implemented in clinical research. Neonatal and maternal outcome are reported in a minority of the trials.

**Chapter 5** ponders on the most appropriate primary outcome measure in RCTs in reproductive medicine. Although we agree with the recommendation of ESHRE that ‘live birth’ is the aim of clinical practice, we feel that there are several arguments why ‘ongoing pregnancy’ best serves the many purposes of a primary outcome measure and best reflects the effectiveness of a treatment.

First, the incidence of ‘ongoing pregnancy’ is higher and thereby the use of ongoing pregnancy reduces the required sample size. Second, interim analysis can be performed six months earlier. These two consequences of the use of ongoing pregnancy as outcome measure cause the results of studies to become available sooner and provide
the opportunity to test futility before recruitment of patients has ended. Third, ‘ongoing pregnancy’, compared to ‘live birth’ is less subject to random error not related to the treatment under study. Fourth, ‘ongoing pregnancy’ reflects effectiveness and is not masked by safety issues like ‘singleton live birth’ is. These two consequences of the use of ongoing pregnancy as outcome measure result in estimates of the treatment effect that are more precise. Finally, ‘ongoing pregnancy’ as primary outcome instead of ‘singleton live birth’ does not ignore the wish of many patients who consider a twin pregnancy desirable, since this gives them a family with more than one child.

We conclude that ‘ongoing pregnancy’ is the primary outcome measure of choice. It will enhance uniformity in future RCTs and allow better interpretation of comparative effectiveness research.

In chapter 6 we expand our thinking on outcome measures of RCTs in reproductive medicine. There is a trend to try and combine safety with effectiveness in one outcome measure. ‘Healthy singleton live birth (at term)’ has been suggested as the ideal outcome measure for evaluative research in reproductive medicine. Effectiveness refers to the extent to which a treatment increases the chance of a couple in having a baby; safety relates to adverse effects associated with such a treatment.

There are several arguments why safety and effectiveness cannot be combined in one outcome measure. First, effectiveness and adverse events in reproductive medicine relate to two subjects who are of interest to us: mother and child. ‘Healthy singleton live birth’ seems to cover safety, but ignores the safety of the mother. Second, a strategy of reporting combined outcome measures ignores the fact that effectiveness and safety are difficult to assess in a single trial. Effectiveness is best evaluated in RCTs. There are large differences between the prevalence of effectiveness and safety outcomes and most RCTs lack power to show meaningful differences in adverse events that occur much less frequently than effectiveness. Third, patients and doctors might value effectiveness and safety differently, and a separate assessment enhances patients to make the decision of their choice.

Combining effectiveness and safety into a uni-dimensional outcome does not help accurate evaluation of treatments in reproductive medicine. Separate assessment of effectiveness and safety will help to improve the process of decision-making.
In chapter 7 we assess whether the issues around the choice of outcome measures leads to selective outcome reporting in RCTs on IVF and ICSI and if so, whether this is related with sponsorship. Selective outcome reporting is a form of bias that can be caused by the choice of the primary outcome measure and by the reporting of only a selection of the outcomes of a trial.

We systematically searched RCTs on IVF and ICSI published between January 2009 and March 2016 in MEDLINE, EMBASE, the Cochrane Central Register of Controlled Trials and the publisher subset of PubMed. We analysed 415 randomized controlled trials. For our primary analysis 235 RCTs were excluded, because the funding source was not reported. Of the 180 RCTs included in our analysis, 7 trials did not report on any primary outcome measure and 107 of the remaining 173 trials (62%) reported on surrogate primary outcome measures like the amount of oocytes. Of the 114 registered trials 21 trials (18%) provided primary outcomes in their manuscript that were different from those in the trial registry. This indicates selective outcome reporting. We found no association between selective outcome reporting and sponsorship.

To reduce waste caused by selective outcome reporting, primary outcome measures should be pre-defined and to answer clinical research questions, clinical primary outcome measures should be chosen.

Chapter 8 provides insight in the consistency and quality of information provision on websites of fertility clinics. Next to researchers and clinicians, patients also have the right to be well informed. The Internet is the most important information source for subfertile couples.

We included 221 websites of fertility clinics in Europe, reporting to the European IVF Monitoring of ESHRE. We analysed the overall quality of information on the websites, including the treatment specific information. To do this, we used the Health on the Net (HON) principles. We also analysed the consistency in the reporting of success rates.

The HON code assessment shows that the overall quality of information, including the treatment specific information on the websites is very diverse and not uniform. 55% of the websites reported on success rates in very diverse ways, using 8 different numerators and 5 different denominators. The majority of the websites reporting on
success rates added an explanatory statement on the interpretation of the success rates.

In view of the important role of the Internet for patients to obtain information, we believe that an ESHRE guideline might be the solution to create better online information sources for subfertile couples in Europe.

IMPLICATIONS FOR THE FUTURE

In reproductive medicine, RCTs can help us to identify the best interventions and strategies to assist subfertile couples in getting a baby. However, the RCTs we currently perform do not bring us reproductive medicine that is supported by good data. In this thesis we disclosed the caveats in RCTs in reproductive medicine that can be influenced by researchers themselves and discussed various ways to create useful clinical research resulting in trusted evidence and informed decisions.

Based on these discussions, we propose the following recommendations to clinical researchers: First, only conduct randomized controlled trials when equipoise, according to the concept based on biological plausibility as described in chapter 2, exists. This helps setting research priorities, prevents a priori negative trials, reduces research waste, and aids in the design of meaningful ones. Second, getting pregnant takes time; therefore analyze cumulative pregnancy rates and multiple cycles within a clearly defined time horizon when conducting randomized controlled trials on the effectiveness of medically assisted reproduction. Third, create uniformity in the choice for the primary outcome measure to enhance the possibilities of data pooling and consider ongoing pregnancy as the preferred primary outcome measure of choice for effectiveness trials in reproductive medicine. Fourth, separate the assessment of effectiveness from the assessment of safety in clinical research to improve clinical decision-making and weigh these according to the values of clinicians as well as patients. Fifth, pre-register primary and secondary outcome measures of clinical interest to prevent selective outcome reporting and thus, prevent difficulties interpreting the data. Finally, create transparent and consistent online patient information.
In line with the issues addressed in this thesis, there are some developments that warrant discussion here.

One of the international initiatives to make clinical research more useful is the CONSORT initiative (Begg et al., 1996). CONSORT stands for Consolidated Standards of Reporting Trials and was developed to alleviate the problems arising from inadequate reporting of RCTs. Recently, various initiatives have been undertaken to adapt the CONSORT guideline for specific medical disciplines. For Obstetrics, CONSORT-OB has been proposed, with more than 30 modifications to the current statement (Chauhan et al., 2013). In the field of reproductive medicine IMPRINT has been proposed, specifically aiming at increasing the reporting of benefits and risks of infertility treatments (Harbin Consensus Working Group, 2014). The consensus group recommended that the preferred primary outcome of all infertility trials is live birth or cumulative live birth and proposed to track the change in quality of RCTs that these guidelines may produce. Until today it has not yet been evaluated whether these recommendations are now being followed.

Another, more recent, international initiative is the Core Outcomes in Women’s health (CROWN) initiative. It is led by journal editors, and aims to harmonize outcome reporting in women’s health research by developing core outcome sets (Kahn, 2014). The development of core outcome sets is a time-taking process, in which health care professionals but also patients are involved. Several core outcome sets have been developed, but in reproductive medicine there is no core outcome set yet.

Involving patients in the development of core outcome sets relates to patient centered health care. Listening to patients’ preferences and needs would be especially helpful for fertility patients, since the inability to conceive creates overwhelming stress and carries a high psychological burden (Verhaak et al., 2007, Wirtberg et al., 2007). This means the use of patient centered, personalized outcome measures (Huppelschoten et al., 2015). Whatever future research reveals on these outcomes, the basis of patient centered care should always be unbiased and useful data acquired through comparative effectiveness research.

When all is said and done, there is another, more structural, problem that is more difficult to be influenced by well-willing and motivated researchers. The recruiting of
patients for randomization is getting more difficult every day. Often, patients have already decided which treatment suits them best before visiting a doctor, since online information sources are widely available.

A randomized controlled trial recently started in the Netherlands comparing IVF with expectant management in women over 38 years old with unexplained subfertility (registered NTR5484). Despite this, in hospitals not participating in this trial IVF is still offered as a first line treatment. Women that decided to go for IVF will consult a clinic that offers IVF, and thus, not participate in the trial. The same applies to a Dutch trial randomizing between IUI and expectant management for unexplained subfertility (Registered NTR5599).

In contrast, in New Zealand the TUI trial, also randomizing IUI and expectant management, successfully includes patients (registered ACTRN 12612001025820). In New Zealand the policy differs from the policy in our country. Public funded fertility clinics accept couples after five years of subfertility. Participating in a trial randomizing IUI and expectant management before ‘normal’ acceptance to the clinic is an attractive option. Also, due to the geographical situation visiting fertility clinics abroad for other or earlier treatment is no option for most couples in New Zealand.

The other side is that doctors are not willing to randomize. In 1951, Austin Bradford Hill was the first to describe the difficulties in getting doctors to randomize their patients, for ethical reasons but too often because the relevance was not appreciated (Bradford Hill, 1951). This attitude towards randomization has barely changed. In the case of randomizing between IUI and expectant management, clinicians at forehand think they know that their patients are not willing to participate and are worried that their patients will go to another clinic to be treated (unpublished data NVOG questionnaire). The money driven phenomenon that new reproductive techniques are implemented without evidence for effectiveness, by claiming that it is no longer ethical to keep these promising techniques away from patients also does not help (Evers, 2015). We have all sworn ‘first do no harm’ but how can we be sure to do no harm if there is no trusted evidence for effectiveness and safety?

Randomised controlled trials are essential for evidence-based medicine, but recruitment of eligible patients remains a structural problem. Our studies and those of others
have shown that doctors, for whatever reasons, are not able to generate trusted evidence in the field of reproductive medicine in an efficient manner.

In the Netherlands, several institutes, like the Patientenfederatie, ZonMW, Zorg Instituut Nederland (ZIN), Nederlandse Zorgautoriteit (NZa), Santeon and the Nederlandse Federatie van Universitair Medische Centra (NFU) are key players in public healthcare. So far, the issues we have addressed have largely escaped their notice. Clinical research and health innovations will not reach its goals unless these institutes combine their forces and restructure clinical healthcare into a system in which randomization is the norm.
REFERENCES


Farquhar C. The intra uterine insemination with stimulation in women with unexplained infertility study. Registered Australian New Zealand Trial Registry. ACTRN 12612001025820


Kahn K. The CROWN initiative: journal editors invite researchers to develop core outcomes in women’s health. *BJOG* 2014;121:1181-1182

Mochtar M. Cost-effectiveness of IVF for unexplained or mild male subfertility in women from 38 years up. A multicenter randomized controlled trial. Registered Dutch trial register NTR5484

Mol F. Intrauterine insemination for unexplained or mild male subfertility. Registered Dutch trial register. NTR5599
