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 2

Equipoise and the RCT

M Braakhekke
F Mol
S Mastenbroek
BW Mol
F van der Veen

Clinical decisions in reproductive medicine are often made in uncertainty. To reduce uncertainty and to improve clinical decision-making, randomized controlled trials are increasingly called upon. A key concept underpinning the ethics of randomized controlled trials is equipoise. Here we aim to dissect the basic reasoning behind the concept of equipoise and we propose a line of thinking delineating under which conditions it is ethical to design and execute a randomized controlled trial. This might prevent a priori negative trials, reduce research waste, and aid in the design of meaningful ones. It is these trials that will provide insight on how to safely and effectively assist subfertile couples.

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 subfertile couples, but the truth is that many clinical decisions are made in uncertainty because of a lack of comparative data on effectiveness on both group and individual level (Farquhar et al., 2015). In absence of these data, personal preferences then start to play a role and this, in turn, may not always lead to the best decisions.

To improve clinical decision-making, the concept of critical appraisal was introduced by David Sackett in the mid 1980-ies, later referred to as 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. The highest level of evidence in clinical research would be generated by aggregated evidence of randomized controlled trials (Guyatt et al., 2000). Standards detailing how to execute and perform randomized clinical trials have well been set (Schulz et al., 2010). So, at present, the why and the how of performing randomized
controlled trials are no longer a matter of debate and as a consequence, an increasing amount of randomized controlled trials is performed each year.

This in itself, is to be applauded, although it is estimated that 85% of research funding across the entire biomedical research range is wasted (Chalmers and Glasziou, 2009). In 2014 the Lancet published a series of five reviews assessing research waste at its several stages; the relevance and priorities of the questions being asked, the design, conduction and reporting of research (Chalmers et al., 2014, Ioannidis et al., 2014, Salman et al., 2014, Chan et al., 2014, Glasziou et al., 2014). Although the series has provoked discussion and is on the agendas of several key players, many steps still have to be made.

To help researchers, clinicians and policymakers in establishing research priorities, we here address the question under what conditions a randomized controlled trial should be performed. The key concept in this is equipoise (Fried, 1974). Equipoise refers to a state of regarding two treatments as an equal bet in prospect (Edwards et al., 1998). Since 1974 there is an ongoing debate on whether equipoise can be limited to an individual doctor being genuinely uncertain or whether equipoise refers to uncertainty in the medical community at large, a state also known as clinical equipoise (Freedman, 1987; Lilford and Jackson, 1995). Also, if clinical equipoise is accepted as the state in which we are allowed to randomise patients, it is debated how to deal with personal preferences and beliefs of the clinician and the patient that are relevant in the choice between treatments (Miller and Brody, 2007; Hey and Truog, 2015; Lilford, 2005).

In our opinion, the basic reasoning behind the concept of equipoise is missing in this debate and we will make the argument that the interaction between pathophysiological processes and mechanisms and observational data is of paramount importance in assessing whether one is in equipoise. To do so, we will delineate three scenarios framing the balance between the understanding of the pathophysiological processes underlying a disease and observational data gathered in the population on the basis of the biological plausibility behind the effect of an intervention.

**Scenario 1: Biological plausibility behind the effect of an intervention is known**

If certain pathophysiological processes or mechanisms have beyond doubt identified the cause of a disease, it has traditionally been assumed that correction of these pro-
cesses will cure the patient. This may well be the case, but obviously any positive effect of a treatment overcoming the pathophysiologic hurdle should be observed before this treatment can be routinely implemented.

If a treatment works exactly as it was expected from the pathophysiologic reasoning and overcomes a large gap between treated and non-treated patients, a single observation can be sufficient to provide evidence of the efficacy of a treatment (Glasziou, 2007). If this is followed by multiple observations confirming the initial observation, a randomized controlled trial is not needed and is even unethical to perform. An example of such an intervention is IVF leading to the birth of Louise Brown in 1978, the first test tube baby. Her mother, Lesley Brown was sterile since her – diseased – fallopian tubes had been surgically removed (Steptoe and Edwards, 1978). In this case it is clear that Louise Brown could only have been conceived by IVF and never by natural conception. When the first case report on the success of IVF was followed by multiple reports of similar events, it was apparent that IVF was a causal treatment for women with surgically removed or blocked tubes that otherwise would have no chance to conceive. Thus, if it is evident that a treatment ‘works’ based on insight into the underlying pathophysiologic processes and a few clinical observations, it is unnecessary and even unethical to perform a randomized controlled trial.

Scenario 2: Biological plausibility behind the effect of an intervention is absent

If biological plausibility behind the effect of an intervention is absent, the a-priori chances that this treatment is effective are low. Bayesian logic dictates that in these situations it is unlikely to detect any meaningful difference in a randomized controlled trial. Such trials should not be performed. Examples are trials in which homeopathic interventions are randomized. Effectiveness of these interventions is not underpinned by any pathophysiological data and the probability of the treatment being effective is zero. Another explicit example is a trial in which intercessory praying for IVF was evaluated. In 2001 a randomized controlled trial was published that assessed the potential effect of intercessory prayer, i.e. the act of praying on behalf of others, on pregnancy rates in women being treated with in vitro fertilization and embryo transfer. In this trial 218 Korean women were randomized between receiving intercessory praying and conventional IVF (Cha and Wirth, 2001). It resulted in a higher pregnancy rate in the women that received intercessory praying, 50% vs. 26%. This randomized controlled trial should not have been performed since there is no pathophysiologic reasoning
behind the effect of intercessory praying on pregnancy rates. Performing this kind of trials is a waste of time, resources and effort, effective treatments are withheld to the patients and they are exposed to possible side effects. Even more importantly, raising false hope to subfertile couples is morally abject.

Scenario 3: Biological plausibility behind the effect of an intervention is uncertain

If biological plausibility behind the effect of an intervention is uncertain and one is thus in doubt whether the proposed intervention might have a positive effect, observational data might be helpful to decide if a randomized controlled trial is necessary. Treatment effects may be observed in cohorts, cases or pilot studies.

If multiple clinical observations show a negative effect of a treatment, like death or in our field irreversible infertility, randomized controlled trials should not be performed. If multiple observations show a positive or no effect of a treatment, we are in equipoise. The same condition is met if some cohorts show a positive effect, while others show a negative effect. These interventions should all be subjected to randomized controlled trials to provide evidence of their effectiveness before introduction to routine clinical practice.

An example of a trial in this category is the use of ICSI versus IVF in non-male-factor-infertility. ICSI leads to higher fertilization rates compared to IVF – physiological plausible and epidemiological data show a positive effect – which lead to the idea that this technique may also generate higher pregnancy rates than conventional IVF (van Rumste et al., 2000). A multicentre randomized controlled trial assessing this topic indeed showed a higher fertilisation rate per oocyte inseminated after ICSI than after IVF, but pregnancy rates per cycle were not higher, thus arguing against the expanded role of ICSI in non-male-factor-infertility (Bhattacharya et al., 2001). This trial countered the hypothesis that the higher fertilization rates observed in ICSI treatments would translate to higher pregnancy rates and provided the evidence to not introduce ICSI in couples with non-male-factor-infertility. Thus, if there is uncertainty on the biological plausibility behind the effect of an intervention and observational data suggest that a treatment might work, we are in equipoise and it is ethical to perform a randomized controlled trial to test the effectiveness of a new intervention.
Many interventions of which the biologic plausibility is uncertain and for which evidence of effectiveness is lacking are already applied on a wide scale. These treatments should not have been introduced in clinical medicine in the first place, but have become part of medicine. In these cases, randomized controlled trials can serve to debunk bad medicine. An example is the administration of progesterone in women with a history of unexplained recurrent miscarriage. Despite the lack of clear pathophysiologic or observational data, progesterone is prescribed by many clinicians. This widespread application without any evidence of its effectiveness provided the underpinning of the PROMISE-trial. The PROMISE-trial randomized these women between progesterone or placebo (Coomarasamy et al., 2015). The trial showed that progesterone therapy did not result in a higher rate of live births among women with a history of unexplained recurrent miscarriage and therefore this treatment should be abolished. Trials to debunk bad medicine might be regarded as fulfilling a policing function and are therefore ethical to perform.

The paradox however is that these trials, which are the most urgent ones to do, are also the ones for which recruitment of patients is the most difficult, especially if the interventions we want to debunk have been applied by the majority of clinicians and for a long period of time. It is then difficult for many to realize and to accept that what is regarded as common practice is based upon hardly anything. As a consequence, patients experience randomization as decreasing their chances of success. Decent counseling of patients by dedicated doctors is of the utmost importance in these trials.

**EPILOGUE**

Since our basic reasoning rests heavily on the biological plausibility behind the effect of an intervention, we feel the need to add a few remarks on the role of biologic plausibility.

First, we have framed our description of what typifies biologic plausibility in the setting of scientific medicine, in which biological hypotheses are generated and after thorough fundamental research rejected or not. The Oxford Dictionaries Online defines the scientific method as ‘a method or procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment,
and the formulation, testing, and modification of hypotheses’ (Oxford University Press, 2010). It follows that types of medicine that do not fulfill these criteria are not part of scientific medicine. An example of this is acupuncture. The mechanisms behind the effect of acupuncture are based on the so-called chi and body meridians, of which existence there is no scientific evidence (Kaptchuk, 2002).

Second, researchers may be so convinced of their own beliefs and reasoning that they shift from uncertain biologic plausibility to known biologic plausibility, even when scientific data underpinning this shift are undeniably absent. An example of this is the biologic plausibility of dehydroepiandrosteron (DHEA) suppletion for women with low ovarian reserve. Proponents of administrating DHEA argue that a randomized controlled trial is not needed, since the androgen-dependency of follicle growth and development is well established (Gleicher, 2016). Thereby, they map DHEA suppletion to scenario 1, known biologic plausibility. Although androgen-dependency of follicle growth is indeed well established, there is limited understanding of how DHEA metabolizes to androgens and subsequently regulates follicular development and whether administrating this precursor of testosterone would stimulate follicle recruitment (Prizant et al., 2014). The biological plausibility therefore remains uncertain and should be mapped to scenario 3. Thus, a well powered, well designed randomized controlled trial is needed.

Third, data on the pathophysiological processes or mechanisms of a disease may be well established, but not taken into account. An example is preimplantation genetic screening (PGS), not to be confused with preimplantation genetic diagnosis (PGD). PGS has been offered in routine clinical practice for more than two decades now. With this procedure, aneuploid embryos are discarded after single-cell analysis. It is thought that the transfer of embryos with such aneuploidies does not result in live birth (Wilton, 2002). Yet, the phenomenon of mosaicism, i.e. the fact that not all cells in a preimplantation embryo have the same chromosomal content, was already reported in 1993, two years before the first birth after PGS, and confirmed in subsequent years undermining the biological plausibility of PGS (Delhanty et al., 1993, Verlinsky et al., 1995, van Echten et al., 2011). Once randomized controlled trials were performed, PGS was shown to be ineffective (Mastenbroek et al., 2011, Twisk et al., 2006). If the existing data on biological plausibility had been seriously taken into account from the beginning, then randomized controlled trials would have preceded instead of followed.
routine clinical application, and women would not have been offered this costly and harmful addition to their IVF treatment (Mastenbroek and Repping, 2014). Now the first generation of PGS techniques has been abandoned and replaced by new methods that use a different timing and different methods for the analysis. And again, these PGS methods are introduced on a vast scale in routine clinical practice without proper randomized controlled trials, while the biological plausibility is certainly not beyond doubt (Gleicher et al., 2015, Geraedts and Sermon, 2016, Greco et al., 2015, Mastenbroek and Repping, 2014, Sermon et al., 2016).

CONCLUSION

As a consequence of performing randomized controlled trials in the absence of equipoise, subfertile couples are exposed to non-effective and often costly interventions and time, effort and money are wasted. We all should be aware that couples with infertility belong to a vulnerable group, that will do anything for a pregnancy (Nap and Evers, 2007). We should only perform carefully designed randomized controlled trials when equipoise is reached or to debunk non-tested interventions that have crept into regular practice.
REFERENCE LIST


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

Chalmers I, Glasziou P. Avoidable waste in de the production and reporting of research evidence. Lancet 2009;374:86-89


Fried C. Medical experimentation: personal integrity and social policy. Amsterdam, North Holland Publishing. 1974


Gleicher N. Randomised controlled trials on dehydroepiandosterone supplementation in female infertility still not conclusive. BJOG 2016;123:1106


Guyatt GH. Evidence-based medicine. ACP J Club 1991;114:A–16


Lilford RJ. Ethics of clinical trials from a bayesian and decision analytic perspective: whose equipoise is it anyway? BMJ 2005;326:980-981


Nap AW, Evers JLH. Couples with infertility belong to a very vulnerable group, they should not be exploited. Hum Reprod 2007;22:3262–3263


