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 6

Effectiveness and safety as outcome measures in reproductive medicine

M Braakhekke
El Kamphuis
F Mol
RJ Norman
S Bhattacharya
F van der Veen
BW Mol

Human Reproduction 2015;30(10):2249-2251
Chapter 6

ABSTRACT

The aim of reproductive medicine is to help couples with an unfulfilled child wish to have a child by offering them the best treatment option. The choice of treatment reflects effectiveness and safety. While 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.

In an attempt to integrate effectiveness and safety, healthy singleton live birth (at term) has been suggested as the ideal outcome measure for evaluative research in reproductive medicine. Although intuitively desirable, this proposal overlooks the fact that assessment of effectiveness and safety in this context cannot be measured as a single outcome. In this paper, we explain why effectiveness and safety outcomes in reproductive medicine should be assessed independently, and later synthesized to inform clinical decision-making.
INTRODUCTION

A consensus paper from the European Society of Human Reproduction and Embryology states that the aim of reproductive medicine is to help couples with an unfulfilled child wish to have a healthy child (Land and Evers, 2003).

To achieve this goal fertility treatments need to be both effective and safe. Assessment of effectiveness and safety of interventions is therefore – not surprisingly – one of the most important elements of clinical research. In our field a number of potential outcome measures have been proposed which combine effectiveness and safety. Examples are singleton live birth per initiated treatment and birth emphasizing a successful singleton at term (BESST) (Min et al., 2004; Wennerholm and Bergh, 2004).

EFFECTIVENESS

Effectiveness relates to how well a treatment works in clinical practice, i.e. to the extent to which a treatment is helping the patient achieving its intended purpose; have a baby. Consequently, for a fertility treatment to be effective, it has to increase the chances of a couple to have a baby, and the birth of a baby is therefore a relevant measure of effect in studies evaluating the effectiveness of a fertility treatment.

SAFETY

On the other hand, the safety of a particular fertility treatment could be reflected in how healthy that baby is. Since most fertility treatments are directed at women, safety outcomes should also include any adverse events that may involve the mother. In reproductive medicine, safety therefore refers to the possible negative consequences of a treatment to the mother or her offspring, occurring either directly as a consequence of the treatment itself, as pregnancy complications or as the impact of the treatment on the long-term health of the mother or child.

In most areas in clinical research, the unit of randomisation is the patient and adverse events are reported in those receiving the intervention, e.g. bleeding disorders in the
treatment of deep venous thrombosis, and arrhythmias in the treatment of asthma with sympathomimetics. In these situations, adverse events can occur irrespective of whether the treatment is effective. Effectiveness and adverse-events are thus measured and reported as different entities, and subsequent clinical decision-making is based on the balance between expected benefits and harms.

**THE DISSECTION OF EFFECTIVENESS AND SAFETY**

In reproductive medicine, effectiveness and adverse events are less independent, since the baby is the outcome of the intervention rather than a participant. In addition, effectiveness and adverse events in our field relate to two subjects who are of interest to us; mother and child. The birth of a healthy child after an uncomplicated pregnancy requires a treatment to be both effective and safe for the baby and the mother. The use of composite outcome measures, such as healthy singleton live birth, might therefore be attractive in terms of capturing effectiveness and safety simultaneously, but makes it impossible to balance effectiveness and safety against each other.

The strategy of reporting combined outcome measures in reproductive medicine also ignores the fact that effectiveness and safety are difficult to assess in a single trial. The best method to evaluate effectiveness of interventions is the RCT in which patients are randomly allocated to a treatment or a control group (Guyatt *et al.*, 2000). Reporting the results of a RCT in a composite outcome measure incorporating safety and effectiveness will have little meaning for safety when there are large differences between the prevalence of effectiveness and safety outcomes. Most randomized infertility trials lack power to show meaningful differences in infrequently occurring adverse events.

The need for separate assessments of effectiveness and safety is well illustrated in a recently published RCT comparing clomiphene citrate and letrozole for treating anovulatory women with polycystic ovary syndrome (Legro *et al.*, 2014). Women who received letrozole had more cumulative live births compared to women who received clomiphene citrate, 27% versus 19% (p 0.007), but there were more congenital anomalies in the letrozole group; 4/102 versus 1/66.
Effectiveness and safety as outcome measures in reproductive medicine

Since these numbers are too small to support a definitive statement on the safety of letrozole, the authors conclude that letrozole is more effective, but also point out that a larger (cohort) study should be performed to clarify the safety and teratogenic risks possibly associated with letrozole. In doing so, the authors make the case for separating effectiveness from safety, and highlight the fact that safety still needs to be evaluated in a larger study.

If the authors had chosen healthy singleton live birth as outcome measure they would have concluded that letrozole is a more effective and safe treatment. Readers might have missed that there might be a safety issue with letrozole as ovulatory agent.

Similarly, the need for separate reporting of effectiveness and safety is demonstrated in a trial of IVF single embryo transfer (SET) versus IVF double embryo transfer (DET) (Thurin et al., 2004; Thurin et al., 2006). When two good quality embryos were available, a fresh SET and, if necessary, a subsequent transfer with a single frozen-and-thawed embryo resulted in a 39% pregnancy rate for a multiple rate below 1%, while performing DET resulted in a pregnancy rate of 43%, but for a 33% multiple pregnancy rate. The corresponding preterm birth rates were 12% after SET and 29% after DET (Thurin et al., 2006).

Separate calculation of relative risks of DET versus SET for measures of effectiveness and safety resulted in a relative risk of 1.1 (95% confidence interval (CI) 1.0 to 1.4) for ongoing pregnancy, 33.4 (95% CI 5.2 to 652) for multiple pregnancies and 2.5 (95% CI 1.5-4.5) for preterm birth (Thurin et al., 2004; Thurin et al., 2006). These measures indicate no significant difference in effectiveness for a much better safety profile of SET. If the authors had integrated effectiveness and safety into a single outcome ‘singleton live birth’, the relative risk would have been 1.3 (95% CI 1.1 to 1.7) indicating only a minor advantage with regards to SET. Again, this example demonstrates that integration of effectiveness and safety in one endpoint clouds our interpretation of results.

**CLINICAL IMPLICATIONS**

We therefore argue for independent measurement of the desired effects (the birth of a child) and adverse events (i.e. preterm birth, congenital anomalies, long-term health
risks, ovarian hyperstimulation syndrome, pre-eclampsia) associated with fertility treatments, as is done in other fields of medicine. In our opinion this offers subfertile couples and their doctors the best opportunity to compare risks and benefits in the context of shared decision-making.

As patients and doctors might value effectiveness and safety differently, both elements should be defined in explicit terms (Dancet et al., 2014). Women with infertility are prepared to accept adverse outcomes in their offspring, short of perinatal death (Scotland et al., 2007). In that respect, they behave in a similar manner as pregnant women, who value the survival of their child as far more important than any residual disability (Bijlenga et al., 2011). Here, effectiveness, i.e. having a child, may be of more value to subfertile couples than safety, i.e. the health of the child.

CONCLUSION

Combining effectiveness and safety into a uni-dimensional outcome impedes accurate evaluation of treatments in reproductive medicine. These outcomes should be assessed independently and interpreted in the context of a couples’ values and preferences as well as their clinical, social and economic circumstances to optimise the process of shared clinical decision-making.
REFERENCE LIST

Bijlenga D, Birnie E, Mol BW, Bonsel GJ. Obstetrical outcome valuations by patients, professionals, and laypersons: differences within and between groups using three valuation methods. *BMC Pregnancy Childbirth* 2011;11:93


Min JK, Breheny SA, MacLachlan V, Healy DL. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. *Hum Reprod* 2004;19:3–7


Wennerholm UB, Bergh C. What is the most relevant standard of success in assisted reproduction? Singleton live births should also include preterm births. *Hum Reprod* 2004;19:1943–1945