EDITORIAL

The sun is not hurried by early risers

J.L.H. (Hans) Evers
Editor-in-Chief

‘If the clinician knows, or has good reason to believe, that a new therapy (A) is better than another therapy (B), he cannot participate in a comparative trial of therapy A versus therapy B. Ethically, the clinician is obligated to give therapy A to each new patient with a need for one of these therapies’ (Shaw and Chalmers, 1970). This ethical dilemma would make participation in clinical trials for many of us impossible. It may also hinder, or even thwart, the development of new therapies. For this reason, Benjamin Freedman introduced the term equipoise (Freedman, 1987). Equipoise means that in the medical community at large—so, not in individual clinicians—there is genuine uncertainty about either the benefit of a new therapy or about the superiority of a new therapy over an existing one. Given that randomized clinical trials presume equipoise, very large treatment effects will therefore rarely be found. They usually are found in first, small, observational studies, especially in those that present surrogate outcomes. If a large treatment effect is found in such a first small study, it usually occurs by chance, at one of the extremes of a normal distribution, and regression-towards-the-mean will, by definition, have subsequent studies show smaller effects (Ioannidis, 2008).

Be aware that this may concern perfectly valid studies with rigorous, robust methods. Pereira and co-workers (2012) studied over 80000 forest plots from 3000 Cochrane reviews (the highest quality evidence). They found that 90% of very large treatment effects (i.e. odds ratio [OR] >5.0) in first trials became smaller in meta-analyses that included subsequent trials. The mean OR decreased from 12 to 4. Many trials lost their nominal significance. They conclude: ‘Most large treatment effects emerge from small studies, and when additional trials are performed, the effect sizes become typically much smaller’, and—referring to the effect of regression-towards-the-mean—‘(…) most large treatment effects should be considered with caution: many are spurious findings, while the vast majority may represent substantial overestimations’.

Ours is the clinical specialty with the smallest, utterly convincing, observational, $n=1$ study in the medical literature: the birth of a healthy baby to a woman without Fallopian tubes (Steptoe and Edwards, 1978). Not all our studies are that clear-cut however, and especially many ‘firsts’ have to be taken with a pinch of salt. We want you to be aware of this. Human Reproduction wishes to publish high-quality reviews of high-quality original research, but we want you to be aware of the limited significance of systematic literature reviews and meta-analyses of the first few positive outcome studies. No one will publish a small study on a new debatable treatment if the outcome is negative, but small, poorly performed positive studies will make headlines. This may impinge on the conclusions of a meta-analysis performed (too) early, especially when careful and critical appraisal of the methods sections of the original papers is lacking. On the other hand, the 40+ RCT’s included in the most recent Cochrane review (Van Wely et al., 2011) of trials comparing urinary with recombinant gonadotrophins (and showing no difference) may indicate that we could have reached the same conclusion many years ago already, that we spent much time, effort and costs on an issue that had been solved already, that patients volunteered for a trial in vain, and that we perhaps had better addressed one of the countless other, urgent but unanswered questions that abound in clinical reproductive medicine.

In this issue of the journal, Armstrong and co-workers (2014) appraise the literature on another of the many new ‘toys for the boys’, time-lapse monitoring of embryo development. They critically evaluate the method sections of current studies, assess the (lack of) evidence of benefit of this new technology, and advocate rigorous study designs to assess the actual contribution of time-lapse imaging (Armstrong et al., 2014). These expert authors highlight the ethical aspects that are involved in hastily adopting a novel technology without robust and reliable evidence of improved live birth rates, safety and cost-effectiveness (Armstrong et al., 2014).

So when is the right time to undertake a systematic review? I agree with Ed Hughes and co-workers (2012), who, in this journal, answered this question with ‘In the case of Cochrane reviews, the simple answer is ‘whenever valid data are available’. Cochrane reviews should not be postponed, waiting for more evidence. On the contrary, they should be undertaken and published when an important clinical question has been addressed by clinical trials’. But reader, be aware (as Cochrane does) of publication bias and the ‘Texas sharp shooter phenomenon’, and carefully scrutinize the quality of the individual studies included.

Or, as we teach during our ESHRE Journals’ Academic Authorship Programme: first study the methods, and only then decide whether to read the paper.

References


