The design, publication and interpretation of research in Subfertility Medicine: uncomfortable issues and challenges to be faced

David H.Barlow
Editor-in-Chief

In the past decade we have seen almost universal acceptance of the concepts encompassed by the words ‘evidence-based’. Whilst some interpret this as meaning that only randomized controlled trial (RCT) evidence is of value many would take the view expressed by David Sackett that evidence-based medicine is...“about integrating individual experience and the best external evidence” (Sackett et al., 1996). In the field of subfertility medicine it is especially important that this latter interpretation is used since large portions of clinical practice are devoid of RCTs. Subfertility medicine is not unique in facing this difficulty, it being recognized that it is more difficult to set up RCTs. This is a difficulty shared with other fields, particularly surgical specialties.

In this issue of Human Reproduction we publish two papers and an Associate Editor Commentary that illustrate well some of the challenges to be faced by our field. Whilst addressing them may not be straightforward the potential gains for effective clinical practice in subfertility are substantial.

When the UK Evidence-Based Clinical Guidelines on the Management of Subfertility were published by the RCOG (Primary, Secondary and Tertiary Guidelines, see www.rcog.org.uk) it was noticeable how many of the recommendations for practice could not be based on evidence classifiable as level A (RCT based). The Cochrane Menstrual Disorders and Subfertility Group (MDSG) compiles the literature of RCTs in the subfertility field (Farquar et al., 1999). As one of the Editors of MDSG I have been aware of the many gaps in the evidence and of the progressive effort to provide a spectrum of Cochrane reviews covering the topics in subfertility practice. One of the papers published in this issue of Human Reproduction is from the Cochrane MDSG editorial base in Auckland. It analyses the quality of the evidence revealed by the spectrum of the Group’s subfertility reviews published in the Cochrane Library (Johnson et al., 2003). The paper highlights the many gaps in current knowledge and, importantly, demonstrates that of 38 subfertility reviews in the Cochrane Library there was insufficient evidence of effectiveness in 26 of them. The reviewers in 23 called for further research to address their respective questions despite the large volume of papers published on subfertility topics over many years. A clue to this perceived continuing shortage of high quality evidence may be provided by another paper published in this issue of the journal (Vail and Gardner, 2003).

Vail and Gardner indicate that even where subfertility RCTs are being reported in what we consider to be leading journals of the field (Fertility and Sterility and Human Reproduction) there are significant shortcomings in many papers and key problems are highlighted. The issues raised are further illustrated and discussed by an excellent Associate Editor Commentary by Daya (2003) who expands on the individual points. This commentary should be basic reading for any researchers setting out to perform RCTs in subfertility. He stresses the importance of the CONSORT Guidelines for the reporting of trials and some key points in relation to subfertility trials. It is important that researchers are thinking about the CONSORT criteria at the point of study design rather than when they are actually writing their paper when it might be too late to correct design problems (Moher et al., 2001). The criteria are best understood by reference to (www.consort-statement.org).

We know that the RCT model is not perfect but it is a design aimed at reducing biases that might lead to wrong conclusions being drawn from a trial. The Cochrane plea is for adequately powered and conducted RCTs but, as highlighted by Vail and Gardner and expanded by Daya, there are important issues of study design and study reporting that need particular attention in the subfertility field. If these are not addressed the study may contain inherent biases which could distort the results obtained.

It is not difficult to understand that if an IVF study permits some patients to participate more than once then the results may be biased if it is analysed as ‘per cycle’ rather than as ‘per patient’ so I find it difficult to understand why those designing such studies permit the double entry in the first place.

The issue of the importance of analysing by intention-to-treat (ITT) can seem counter-intuitive when this means that, for example, couples randomized in an IVF study are to be included in the analysis even if they did not get as far as receiving the intervention being tested. Daya’s commentary provides a clear explanation of the importance of ITT analysis and useful advice on how to minimize the problem of non-treatment by appropriate study design.

A matter which provokes discussion is the requirement that subfertility trials which have pregnancy outcomes should be reporting success as ‘live births’ rather than as ‘pregnancies’. We appreciate that live birth is the most meaningful goal of treatment but Vail and Gardner report in their analysis 34 of 39 trials with a pregnancy outcome failing to report live births as the trial end point. This is consistent with much of the
subfertility literature. How are we to resolve the conflicting pressures of wishing to see reporting of trials with this optimal end-point information and the inherent delays incurred in awaiting live birth outcomes which can be a problem for the length of studies (both for young researchers and for sponsors) and in some settings the difficulties associated with obtaining this late outcome data? I suspect that the subfertility literature will be dominated by evidence based on pregnancy end points rather than live birth end points for some time to come but this should not deter us from seeking to establish acceptance that the live birth outcome is optimal and ask authors to justify failing to provide that information.

How are journals to move forward on the quality of study design, analysis and reporting? It is very clear that well-powered RCTs which comply with all of the points made by Daya and Vail and Gardner should be of high quality and should be more likely to provide a reliable answer to the questions they study. It is to be expected that referee and editorial processes will detect shortcomings and assign lower weight to the publishing of trials where some of these shortcomings exist. In my experience of the editorial process at *Human Reproduction* many submitted trial papers are rejected on methodological grounds. If the original design is fatally flawed we would hope that our processes would not encourage resubmission but papers originally rejected because of inadequate reporting of the trial processes and/or the analysis may be accepted on resubmission after revisions which address the points in question. The fact that there may still be shortcomings, as shown in the Vail and Gardner paper, raises the question of whether trials should not be published if there are shortcomings as described. In deciding on publication there is emphasis on whether the research question is important, on the originality of the work and on the quality of the study design and analysis. A study scoring highly on all these criteria will be very likely to be published but studies falling short are more likely to be refused. I very much look forward to a time when we shall be able to address the key questions in subfertility with flawless trials. The Commentary by Daya serves to aid this by his exposition of the important principles. Prospective authors should note that our Associate Editors will be reminded of these important issues for their assessment of papers and we shall be re-emphasising the journal’s compliance with the CONSORT Criteria. When the journal receives a paper involving an RCT the refereeing process cannot commence until we have seen a completed CONSORT statement of compliance from the authors including the trial flow chart. Where this is not presented with the paper on submission there is inevitable delay in seeking the statement from the authors before the refereeing can start. The compliance statement document is available from the CONSORT website (http://www.consort-statement.org/) and we would advise authors to send this in at the original submission of the paper in order to avoid creating delays.

The CONSORT statement relates to the publishing of RCTs and seeks to ensure clarity in the reporting. The QUOROM statement is a parallel development that seeks to provide clarity in the reporting of meta-analyses (Moher et al., 1999). *Human Reproduction* now publishes Cochrane reviews, which by definition have already undergone rigorous peer-review. Authors submitting meta-analyses that have been performed independently of Cochrane will be encouraged to provide a statement of QUOROM compliance before refereeing can take place (http://www.consort-statement.org/QUOROM.pdf).

A third and quite recent development is the STARD initiative which focuses on the complete and accurate reporting of studies of diagnostic accuracy and this once again relies on a checklist and flow chart to aid authors in reporting their work (Bossuyt et al., 2003). As with RCTs and meta-analyses authors of studies on diagnostic accuracy will be encouraged to provide a completed STARD checklist and flow chart before reviewing is initiated (www.consort-statement.org/statement_test.pdf).

In the reviewing process a journal seeks to ensure that the interpretation and claims presented in a paper are appropriate and proportionate to the data. One place where this is important is that where a trial has been sponsored by a pharmaceutical company, the temptation might be to maximize the positive interpretation. A less obvious but equally important matter of interpretation is in the opposite direction where there can be a risk of disproportionately negative interpretation being placed on a study reporting some form of risk with a danger that inappropriate alarm might result. This alarm might originate in the paper or in the interpretation of the paper as presented in the media. Examples of this were seen recently in the publicity resulting from publications on the detection of imprinting disorders and retinoblastoma in babies conceived through assisted reproduction (DeBaun et al., 2003; Moll et al., 2003). Many couples who had conceived through IVF or were contemplating IVF contacted clinics expressing anxiety following the publicity. There appeared to be little understanding that what was being reported was an increase in the risk of rare conditions. In other words the reports were suggesting children conceived by assisted reproduction have an increased ‘relative risk’ of imprinting disorders which themselves have a very low ‘absolute risk’. Thus despite the increase in relative risk the absolute risk remains very low. Examples are the report on the imprinting disorder, Beckwith-Wiedermann Syndrome (BWS) (DeBaun et al., 2003) and the report on the childhood cancer, retinoblastoma (Moll et al., 2003).

DeBaun et al. carried out a prospective observational study from the Washington University BWS Registry and reported that the prevalence of ART in the BWS series was 4.6% against a background rate of ART of 0.8% in the United States population. Thus a six-fold increase in the prevalence of the condition is indicated in association with ART but although BWS is referred to as a ‘rare congenital malformation’ in the discussion of the paper no estimate of the absolute risk or the prevalence of this condition is provided. A second report of an increased relative risk of BWS in association with ART has just been published (Maher et al., 2003) and it does quote an estimate of the prevalence of BWS to be at least 1.3 per 100 000 newborn infants from a Spanish study (Arroyo Carrera et al., 1999).

The other report used data from the Dutch retinoblastoma registry and The Netherlands cancer registry of an increased risk of retinoblastoma in children in association with IVF.
treatment. The increased relative risk is reported to be in the range 4.9 to 7.2. The report does provide incidence information indicating that retinoblastoma occurs in 2.6 babies per 100,000 in the first year of life and as 0.9 per 100,000 children between ages 1–4. This provides the necessary low absolute risk context in which to view the importance of the finding in terms of advice to patients.

In both examples the suggested increased risks in association with ART are scientifically noteworthy and we must continue to seek to know about risks of problems in children conceived through ART but I would suggest that public discussion of these risks must be in the context of the rarity of the conditions. I believe that it is important that some estimate of prevalence or absolute risk is given to provide that context. Otherwise the reader cannot judge what importance they should place on the report of an increased risk.

It would be useful if it were more common to put reports of relative risk in the context of absolute risk and that this important contextual information be placed in the abstract of such papers. With the ready availability of the abstracts of scientific papers on electronic media it is important that those who browse abstracts have immediate access to this aspect of the clinical perspective of the report. Good examples of such helpful information in another field of reproductive medicine can be seen in some major reports on the risks associated with HRT use where the reader of each abstract is given a clear indication of the absolute effect of the reported increased risks (Collaborative Group on Hormonal Factors in Breast Cancer, 1997; Beral et al., 2002; Writing Group for the Women’s Health Initiative Investigators, 2002). The editorial policy of Human Reproduction will be to encourage authors to include such information in the abstracts of the papers we publish.

References


DOI: 10.1093/humrep/deg270