Progress in evidence based reproductive surgery

J. Bosteels[1,2], S. Weyers[3], C. Siristatidis[4], S. Bhattacharya[5], T. D'Hooghe[6]

[1] *Department of Obstetrics and Gynaecology, Imeldahospitaal, Imeldalaan 9, 2820 Bonheiden, Belgium.*
[2] *CEBAM, Centre for evidence-based medicine, the Belgian branch of the Cochrane Collaboration, Capucijnenvoer 33, blok J, 3000 Leuven, Belgium.*
[3] *Universitaire Vrouwenkliniek, University Hospital Gent, De Pintelaan 185, 9000 Gent, Belgium.*
[4] *Assisted Reproduction Unit, 3rd Department of Obstetrics & Gynecology, "Attikon" University Hospital, Athens, Greece.*
[5] *Aberdeen Maternity Hospital, School of Medicine and Dentistry, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, United Kingdom.*
[6] *Department of Obstetrics and Gynaecology, University Hospital Gasthuisberg, KU Leuven, Herestraat 49, 3000 Leuven, Belgium.*

Correspondence at: Dr. Jan Bosteels, Department of Obstetrics and Gynaecology, Imeldahospitaal, Imeldalaan 9, 2820 Bonheiden, Belgium.
E-mail: jan.bosteels@med.kuleuven.be

Abstract

The Consolidated Standards of Reporting Trials (CONSORT) was introduced in 1996 to improve the methodological quality of published reports of randomised controlled trials. By doing a systematic review of randomised controlled trials on reproductive surgery, our group can demonstrate that the overall quality of the published reports of randomised studies on reproductive surgical interventions has improved after CONSORT. Nevertheless, some problems still remain. By discussing the benefits and pitfalls of randomised trials in reproductive surgery, our opinion paper aims to stimulate the reader's further interest in evidence-based practice in reproductive surgery.

Key words: Evidence-based medicine, randomised controlled trials, reproductive surgery.

Introduction

Traditionally, a wooden spoon was given every year to the student with the lowest score at the comprehensive mathematics examination at St Johns College in Cambridge University. It was awarded for the last time in 1909. Its possession implied that its owner was actually better equipped to be a cook than a scholar.

In 1979 Archie Cochrane awarded a wooden spoon to Obstetrics and Gynaecology because the uptake of designing randomised controlled trials (RCTs) in this discipline was almost non-existent. Some time before, he had criticized the medical profession by writing that" we have not organised a critical summary, by specialty or subspecialty, updated periodically, of all relevant randomised controlled trials" (Cochrane, 1979). Initially, Cochrane's challenge was taken up in perinatal medicine. In the field of reproductive medicine, the first systematic review of the effectiveness of subfertility treatments was published in 1993 (Vandekerckhove *et al.*, 1993).

In surgery a "non-evidence-based" approach to practice has been traditionally present (Johnson *et al.*, 2008). The latest surgical technique is often embraced by the clinical community either when it seems rational or revolutionary or whenever it demonstrates the technical skill of the surgeon.

In this opinion paper we will present data and some conclusions on the current methodological quality of published reports of randomised studies in reproductive surgery. By discussing the benefits and pitfalls of RCTs in reproductive surgery, we aim to stimulate the reader's interest in evidence-based practice in reproductive surgery.

Evidence-based medicine

"Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett *et al.*, 1996). The practice of evidence-based medicine (EBM) stands for the integration of individual clinical expertise with the best available external clinical evidence from systematic research. The essentials of EBM include five consecutive steps: first of all, to ask the right questions; secondly, to find the best level of evidence available; thirdly, to appraise critically the evidence for risk of bias, clinical relevance and applicability; fourthly, to implement the results of the appraisal in every day clinical practice and fifthly to evaluate the changes in practice (Farquhar and Vail, 2006). The highest level of evidence is derived from well written critically appraised systematic reviews of RCTs. The randomised controlled trial is generally accepted as being the least biased measure of the effectiveness of interventions. Although observational studies are considered vastly superior to RCTs in detecting adverse events e.g. surgical complications, they are often misleading when they are employed in searching for moderate treatment benefits. Systematic reviews comparing observational studies with randomised trials of the same interventions for the same conditions in the same study populations concluded that the former were clearly unreliable and consistenly overestimated the treatment effect (Britton *et al.*, 1998; Kunz *et al.*, 2004).

RCTs in surgery: the benefits

Gynaecology has evolved to becoming a specialty in which the interventions are increasingly exposed to the gold standard of RCTs (Johnson *et al.*, 2003). An overview of 23 systematic reviews including 94 gynaecological surgical trials in the Cochrane Database of Systematic Reviews (CDSR) (Selman *et al.*, 2008) has ended up to the final conclusion that the quality of the RCTs has significantly improved since the Consolidated Standards of Reporting Trials (CONSORT) was introduced in 1996 (Begg *et al.*, 1996). Using meta-regression analysis the authors have demonstrated that the proportion of studies reporting allocation concealment has significantly increased after the introduction of the CONSORT statement (60% versus 26%, p = 0.002). In parallel, a reduction in the magnitude of the effect estimate was observed over time (log of the ratio of odds ratios per year 0.96, 95% CI 0.93-0.99, p = 0.05) together with a trend towards higher precision of the estimation of the treatment effect (inverse of variance of the log odds ratio 0.12, 95% CI 0.02-0.23, p = 0.03) (Selman *et al.*, 2008). In a second overview of 30 reviews in the CDSR, the same authors' group found that only 7 out of 30 reviews reported evidence of a significant effect, 11 out of 30 reviews concluded that there was some evidence of significant effects for primary outcomes along with some evidence gaps while in the remaining 12, the authors found insufficient evidence of effectiveness (Johnson *et al.*, 2008).

In conclusion, apart from providing up to date unbiased evidence on health care interventions, systematic reviews of RCTs can identify 'gaps of knowledge' where there is insufficient or no evidence at all. Several knowledge gaps in the evidence for fertility treatment have already been identified in a review of RCTs from the Cochrane Menstrual Disorders and Subfertility Group (MDSG) database (Johnson *et al.*, 2003).

RCTs in surgery: the problems and pitfalls

There are two major categories of methodological challenges that need to be at least identified if not solved during the design phase of RCTs on surgical interventions (McCulloch *et al.*, 2002; McLeod, 1999).

The first category concerns issues on the design and conduct of surgical trials. The surgical learning curve raises an interesting dilemma for the timing of surgical trials: it is well known that the individual surgeons' complication rates fall significantly as the procedure is carried out on more and more patients. While drugs in trials work the same regardless of the competence of the prescribing physician, there are surgeon-to-surgeon differences in the preferences for and the expertise in performing different surgical procedures (Devereaux *et al.*, 2005). In a recent Cochrane review on the effectiveness of excisional versus ablative surgery for ovarian endometriomata an effect favoring the excision of the cyst wall compared with its drainage and ablation was demonstrated the odds for a spontaneous pregnancy at 12 months after excision of the endometriotic cysts was higher compared with the control group which was treated by drainage and ablation (OR 5.2, 95% CI 1.9-14) (Hart *et al.*, 2011). One of the included trials provided evidence for a treatment effect in favor of the excision technique for the spontaneous pregnancy rate at 12 months (OR 4.8, 95% CI 1.6-14, 62 patients) (Alborzi *et al.*, 2004) while another smaller trial demonstrated a higher point estimate favoring excision over ablation but failed to reach statistical significance (OR 8.0, 95% CI 0.69-93, 26 patients) (Beretta *et al.*, 1998). In the former trial, the intervention was performed by the same surgeon in two university centres, whereas in the latter no information on the number of surgeons involved was available. Additionally, in both of them no information on the expertise of the performing surgeons was given. The need for head-to-head comparisons between different surgical techniques inevitably necessitates that the same surgeon prefers both techniques and is an expert in performing both of them. This is difficult in practice and impossible to achieve through studies. Therefore a strong case can be made for "expertise-based" trials in which consenting patients are allocated to different expert surgeons, who carry out the procedure they prefer and are expert in performing. While improving the internal validity of the trial, this potentially diminishes the external validity of the trial as well, meaning that the results of the RCT cannot be generalised as such without caution.

In the same context, in the same trials mentioned above, the application of both techniques raises further considerations: how sure can we be that both surgical teams were using comparable techniques? Did they selectively coagulate visible endometriotic lesions or was the whole cyst wall evaporated? No such clarifications were presented in the published reports. In addition, in one trial (Beretta *et al.*, 1998) hydroflotation was used in contrast to the second trial (Alborzi *et al.*, 2004). The previous remarks illustrate the great difficulty to standardize a surgical intervention since each individual surgeon develops his own modification of a standard technique e.g. for dissection, hemostatis and/or management of complications. There have been some attempts, though, to comprehensively standardise the technical steps of surgical interventions (Kapiteijn *et al.*, 1999).

Another point which needs adressing is the difficulty of blinding or masking of a surgical procedure combined with the legal obligation of the treating physician to obtain informed consent. This can be a major problem if "soft" outcome measures, e.g.pain or quality of life are being assessed through selfreporting by unblinded patients or determined by unblinded assessors.The emotional consequences of knowing one's treatment may significantly affect the reporting of outcomes. "Sham" surgical procedures have been conceived in the past to try to overcome the issue. Ethical problems may potentially arise (Moseley *et al.*, 2002) and therefore "hard" outcome measures, e.g. live birth rate are mostly preferred. The latter are relatively independent to the knowledge of a patient's treatment, but there still is the need for the outcome assessors to be blinded to the allocated treatments.

A second category of problems concerns the interpretation of the results of the trials. The mixture of data from trials conducted by less experienced surgeons together with others done by more expert ones may negatively affect the magnitude of the effect estimate, since differences in treatment outcomes are expected. In surgery, it is logical that a surgical intervention has a more favorable outcome when the provider is more experienced. Finally, a common problem in RCTs concerns the statistical power of surgical trials: a large survey of 90 "negative" surgical trials found that only 24% had sufficient power to detect relative risk reductions of 50% and only 29% reported a formal sample size calculation (Dimick JB *et al.*, 2001). Power calculation is currently considered as an absolute must in the proper conduct of an RCT. It constitutes one of the main endpoints which a reviewer has to judge for a clinical trial and gives the adequate power to the results and therefore the interpretation of the trial's data.

When is it ethical to design an RCT in surgery?

It is essential to define the circumstances under which an RCT can be conducted to determine whether a surgical procedure is more effective compared to other surgical or non-surgical treatments. We usually undertake trials because we hypothesise that a new surgical procedure can be better than the current standard practice in terms of efficacy, safety or cost but we are uncertain whether this statement is true or false. The limits of uncertainty include the possibility that the new technique may not be better or even worse than the current standard practice. The true uncertainty on the part of the expert professional community about the benefit to harms balance of two or more treatments for a well-defined study population has been described as the "clinical equipoise" (Freedman, 1987). When clinicians, methodologists and ethics committees or institutional review boards are uncertain whether an intervention is beneficial, an RCT is judged to be appropriate.

In addition we need to consider the uncertainty in the patient-clinician relationship, through an active patient participation in the inclusion/ exclusion process. If the patient is certain that a specified treatment is better or safer, then the patient should not be included in the trial. Similarly, if the physician judges that a particular patient is clearly better off with a treatment, ethically, he is obliged to inform the patient and to assist in seeking the most appropriate treatment, excluding the patient from the trial. If both physician and patient are uncertain which treatment to choose, the patient should be offered inclusion in the trial.

The above principle of equipoise should always be considered as the gold standard in deciding whether or not to design an RCT. Some consider the evidence provided by non-randomised studies as an

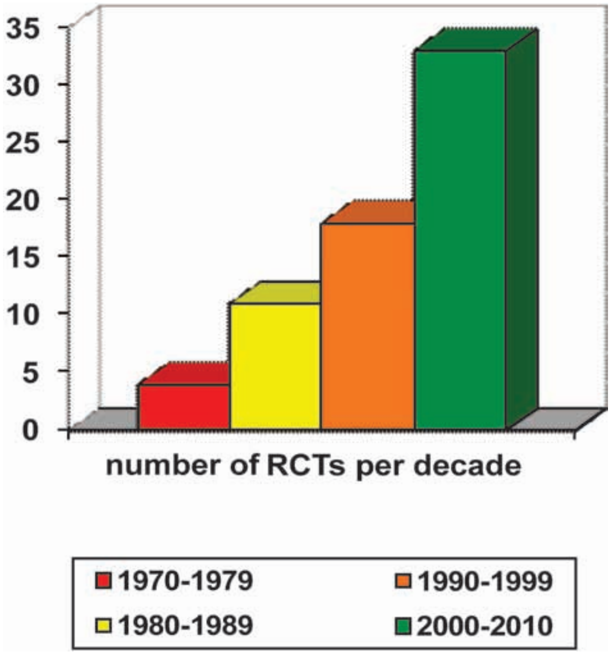

Fig. 1. — RCTs on the effectiveness of reproductive surgery

ethical basis to discard the need for further research. Their certainty based on the results of studies with a high risk of bias should nevertheless be put aside in deference to the reasoned uncertainty existing within the larger community of experts (Haynes *et al.*, 2006).

RCTs in reproductive surgery: the present state

Our group has published a systematic review on the effectiveness of reproductive surgery for treating female infertility (Bosteels *et al.*, 2010). We conducted a search in *the Cochrane Library*, MEDLINE and EMBASE for RCTs on reproductive surgery in subfertile women. Our findings demonstrated a steady increase from 1970-2010 in the number of RCTs on the effectiveness of reproductive surgery per decade (Figure 1).

Nearly 75% of the included 63 RCTs had an adequate random sequence generation and nearly 50% had adequate allocation concealment (Figure 2). The percentage of RCTs on reproductive surgical interventions with adequate allocation concealment (26 out of 63 studies or 41%) was similar (p = 0.67) to the findings of the review of gynaecological surgical trials available in the Cochrane Library (42 out of 94 studies or 45%) (Selman *et al.*, 2008).

The number of trials with adequate random sequence generation has nearly doubled from the pre- compared to the post-CONSORT era (RR 1.7; 95% CI 0.98-3.1) (Figure 3): the difference was marginally insignificant (p = 0.06). Although the number of RCTs in the field of reproductive surgery with adequate allocation concealment has nearly doubled from the pre- (4 out of 16 studies or 25%) compared to the post-CONSORT era (22 out of 47 studies or 47%), the current sample size in our review is too small to draw definitive conclusions (RR 1.9, 95% CI 0.76-4.6) (Figure 3). Despite the non-significant p-value (p = 0.17) our data are nevertheless consistent with the findings of the review of gynaecological surgical RCTs in the Cochrane Library which did demonstrate both an important and statistically significant increase (p = 0.002) (Selman *et al.*, 2008). The absence of evidence of a better methodological quality (RR 1.0, 95% CI 0.23-4.6) concerning blinding pre-versus post-CONSORT illustrates the great difficulty of adequate blinding in surgical trials (Figure 3). The methodological quality of the trials on reproductive surgery as determined by random sequence generation, allocation concealment and blinding has improved after the CONSORT statement (RR 1.7, 95% CI 1.1-2.7); the p-value was compatible with a statistically significant difference (p = 0.03) (Figure 3).

Live birth rate was reported as the primary outcome measure in 16 out of 63 studies or 25% of the included RCTs.

In 7 out of 15 topics there was evidence of a significant effect for primary outcomes; in 5 out of 15 topics there was some evidence of effect for primary outcomes along with some evidence gaps; in 3 out of 15 topics there was insufficient or no evidence. A summary of the grading of the evidence for different topics in reproductive surgery is presented in Table 1.

Discussion and future perspectives

The limited and poor quality evidence provided by 63 RCTs indicated a positive role for some surgical reproductive interventions. Overall the methodological quality of the RCTs published after the CONSORT statement in 1996 has improved but this conclusion should be made with caution given the limited numbers of the included trials in our systematic review. In addition it is evident that not every methodological problem has been solved. Since reproductive medicine was one of the first domains where the need for evidence-based practice was stressed (Vandekerckhove *et al.*, 1993), it seems logical that research in reproductive surgery should also be further exposed against the gold standard of RCTs. We agree with others that evidence-based reproductive surgery "is no passing fad" (Johnson *et al.*, 2008).

In many publications on the methodological aspects of studies, the concealment of allocation

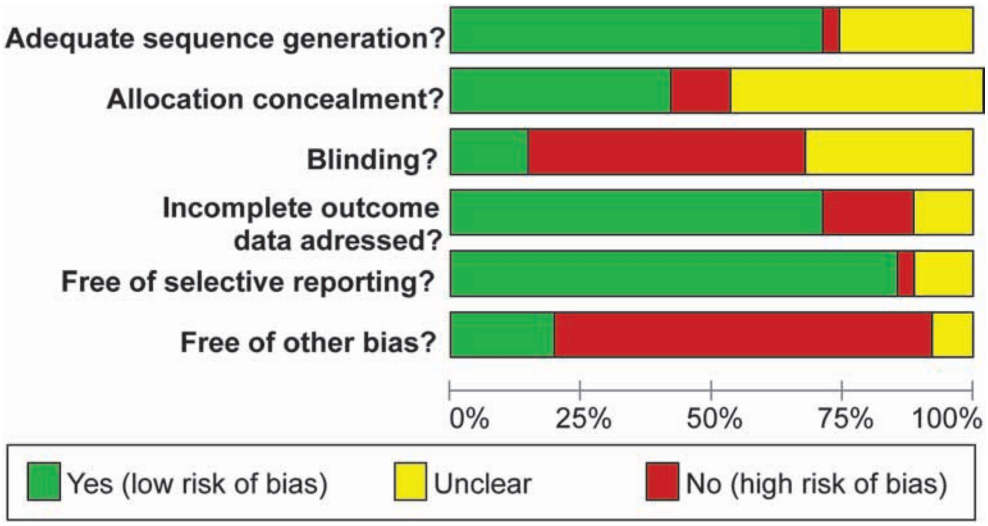

Fig. 2. — Methodological quality: risk of bias across studies

to the treatment and the control group has been consistently shown to be the single most important factor in assessing the quality of RCTs (Farquhar and Vail, 2006). Nevertheless several large studies assessing the use of allocation concealment in different topic areas and subfertility trials have reported this item infrequently (Jüni *et al.*, 1999; Moher *et al.*, 1995; Schulz *et al.*, 1994; Kjaergard *et al.*, 2001). This should be a major concern for trialists designing future RCTs in surgery. In contrast, while the absence of blinding is almost inherently associated with surgical trials, blinding has not been consistently shown to affect the estimation of the treatment effect magnitude (Jüni *et al.*, 1999; Moher *et al.*, 1995; Schulz *et al.*, 1994; Kjaergard *et al.*, 2001). The quality of the generation of the randomisation

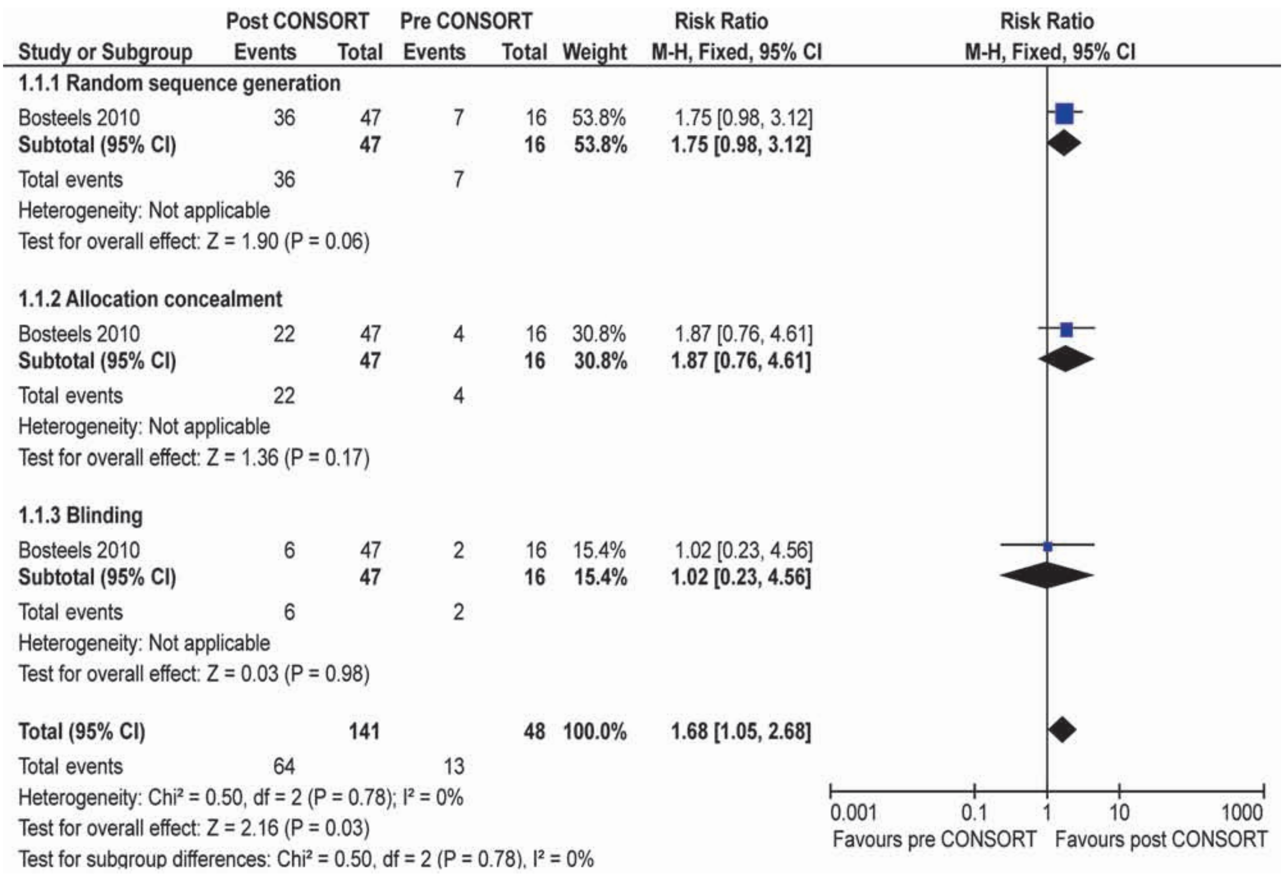

Fig. 3. — RCTs with adequate random sequence generation, allocation concealment and blinding before vs. after CONSORT (1996)

Table 1. — Grading of evidence of the randomised studies in reproductive surgery.

Topic under review	RCTs	number of participants	conclusions for primary outcomes (evidence category)
Laparoscopic treatment for subfertility associated with rAFS I/II endometriosis	2	437	Laparoscopic excision/ ablation and adhesiolysis improves the chance for live birth and ongoing pregnancy (E)
Treatment of endometriomata by excision or ablation.	2	88	The excision of endometriotic cysts significantly improves the chance for spontaneous conception at 12 months (EG)
Treatment of endometriomata prior to IVF	1	99	There is no evidence of an effect in favor of removing endometriomata prior to IVF (G).
Laparoscopic drilling for induction of ovulation in PCOS.	6	439	There is no evidence of a treatment effect of LOD (6-12 months follow-up) versus gonadotropin injections (3-6 cycles) for the ongoing pregnancy rates (EG).
	5	166	There are significantly fewer multiple pregnancies with LOD (E)
	5	181	There is no evidence of an effect of bilateral compared to unilateral LOD (EG).
Surgical treatment for tubal disease in women with hydrosalpinx due to undergo IVF	4	455	Laparoscopic salpingectomy for hydrosalpinges prior to IVF significantly improves the chances for pregnancy (all definitions) (E).
	2	209	Tubal occlusion is at least as effective as an alternative (EG)
Prevention of adhesions after previous reproductive surgery	1	74	There is no evidence of a treatment effect for second-look laparoscopy with adhesiolysis in improving pregnancy rates after failed tubal microsurgery(E).
	1	36	There is some benefit for the use of hyaluronic acid gel after laparoscopic myomectomy (G).
Surgical treatment of fibroids for subfertility	2	309	Hysteroscopic myomectomy doubles the pregnancy rate compared to expectant management in subfertile women with submucosal fibroids (EG).
	1	87	The removal of intramural or subserosal fibroids tends to increase the pregnancy rate, but the effect is not statistically significant (G).
Laparoscopy prior to IUI	1	154	There is no evidence of a treatment effect of laparoscopy prior to IUI (E).
Hysteroscopic removal of polyps	1	215	Hysteroscopic removal of polyps visible on ultrasound increases the pregnancy rates in women undergoing IUI (E)
Hysteroscopy in women with IVF failure	2	941	Hysteroscopy prior to IVF doubles the clinical pregnancy rates in patients with 2 failed IVF attempts (E).

sequence has similarly with the item of blinding not been shown to be of major importance in causing substantial bias (Jüni *et al.*, 2001).

Considering the outcome measures, the majority of trials in subfertility and reproductive surgery do not report live birth outcomes as their primary outcome. This problem has already been highlighted by others (Vail and Gardner, 2003). It could be argued that all future trials on the effectiveness of reproductive surgical interventions should report live birth rate as the primary outcome measure since it is the single most important outcome of interest for couples undergoing fertility treatment. Ideally, the cumulative live birth rate, using life table analysis, should be described, as it accounts for the time to pregnancy and allows to substract periods when the patient was not actively seeking to conceive. Time-to-event data are however troublesome for use in statistical pooling in meta-analyses. Moreover, the other outcome measures of interest in reproductive trials e.g. pregnancy and miscarriage rates should not be considered inferior since some conditions amenable to surgery may have an indirect impact on fertility, e.g. septate uterus which increases the probability of miscarriage.

Finally, the correct use of evidence statements should be encouraged. A common error observed in many studies is the confusion between "significant" and "important" or "clinically relevant". A result is statistically significant if the difference observed between the study and the control samples is sufficiently convincing to signify a real difference in the population of which the sample is representative. A result is important or clinically relevant if the magnitude of the effect estimate is large enough to constitute a real difference between a control and study intervention for a given outcome. Ideally, authors and trialists should predefine minimally important clinical differences, based on estimates or trade-offs by physicians and/ or patients of what really constitutes an important improvement of the outcome under study. If the sample size is large enough, a clinically unimportant or even trivial difference may signify a population difference, while in contrast clinically relevant differences may not be statistically significant if the sample size is too small. A second common error is the misinterpretation of a statistically non-significant finding. "Negative trials" do not exist! The correct expression of a conclusion is the absence of evidence of a particular effect and not the evidence of its absence (Altma and Bland, 1995; Alderson and Chalmers, 2003).

The methodological quality of surgical trials can be improved eighter through the training of surgeons in clinical epidemiology and evidence-based medicine or employing epidemiologists in surgical units where clinical research is being carried out (Urschel *et al.*, 2001; Madhok *et al.*, 2002). The evidence from our recent systematic review is consistent with this viewpoint.

In conclusion, true progress in the field of reproductive surgery needs a balanced combination of surgical skills, a drive for innovation together with the exposure of clinical research to the undoubtful validity of evidence-based medicine.

References

Alborzi S, Momtahan M, Parsanezhad ME *et al.* A prospective randomised study comparing laparoscopic ovarian cystectomy versus fenestration and coagulation in patients with endometriomas. Fertil Steril. 2004;82:1633-7.

Alderson P, Chalmers I. Survey of claims of no effect in abstracts of Cochrane Reviews. BMJ. 2003; 326:475.

Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995;311:485.

Begg C, Cho M, Eastwood S *et al.* Improving the quality of reporting of randomized controlled trials: the CONSORT statement. JAMA. 1996;276:637-9.

Beretta P, Franchi M, Ghezzi F *et al.* Randomised clinical trial of two laparoscopic treatments of endometriomas: cystectomy versus drainage and coagulation. Fertil Steril. 1998; 70:1176-80.

Bosteels J, Weyers S, Mathieu C *et al.* The effectiveness of reproductive surgery in the treatment of female infertility: facts, views and vision. F, V & V in ObGyn. 2010;2:232-52.

Britton A, Mc Kee M, Black N *et al.* Choosing between the randomised and non-randomised studies: a systematic review. Health Technol Assess. 1998;2:13.

Cochrane AL. 1931-1971: a critical review with particular reference to the medical profession. In: Teeliz Smith, G. (ed) Medicines for the year 2000. Office of Health Economics, London, UK.pp.1-11.

Devereaux PJ, Bhandari M, Clarke M *et al.* Need for expertise based randomised controlled trials. BMJ. 2005;330:88-91.

Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalence or error? Arch Surg. 2001;136:796-800.

Farquhar C,Vail A. Pitfalls in systematic reviews. Curr Opin Obstet Gynecol. 2006;18:433-9.

Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987;317:141-5.

Hart RJ, Hickey M, Maouris P *et al.* Excisional surgery versus ablative surgery for ovarian endometriomata. Cochrane Database Syst Rev. 2011;issue 3:CD004992.

Haynes BR, Sackett DL, Guyatt GH *et al.* Clinical epidemiology – how to do clinical practice research. Third edition, Lipincott, Williams and Wilkins, 2006.

Johnson NP, Proctor M, Farquhar CM. Gaps in the evidence for fertility treatment – an analysis of the Cochrane Menstrual Disorders and Subfertility Group database. Hum Reprod. 2003;18:947-54.

Johnson NP, Selman T, Zamora J *et al.* Gynaecologic surgery from uncertainty to science: evidence-based surgery is no passing fad. Hum Reprod. 2008;23:832-9.

Jüni P, Witschi A, Bloch R *et al.* The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282: 1054-60.

Jüni P, Altman DG, Egger M. Systematic reviews in healthcare: assessing the quality of controlled clinical trials. BMJ. 2001; 323:42-6.

Kapiteijn E, Kranenbarg EK, Steup WH *et al.* Total mesorectal excision (TME) with or without preoperative radiotherapy in the treatment of primary rectal cancer. Prospective randomised trial with standard operative and histopathological techniques. Eur J Surg. 1999; 165:410-20.

Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001;135:982-9.

Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in health care trials (Cochrane Methodology Review). Updated 22 august 2001. In: *The Cochrane Library*. Issue 3. Chichester, UK: John Wiley & Sons, 2004.

Madhok R, Handoll HHG, Bridgman S *et al.* Randomised trials in surgery. Letters to the Editor of the BMJ. BMJ. 2002; 325:658-9.

McCulloch P, Taylor I, Sasako M *et al.* Randomised trials in surgery: problems and possible solutions. BMJ. 2002;324: 1448-51.

McLeod RS. Issues in surgical randomized controlled trials. World J Surg. 1999;23:1210-4.

Moher D, Jadad A, Nichol G *et al.* Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995;16:62-73.

Moseley JB, O' Malley K, Petersen NJ *et al.* A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med. 2002;347:81-8.

Sackett DL, Rosenberg WMC, Muir Gray JA *et al.* Evidence based medicine: what it is and what it isn't. BMJ. 1996;312: 71-2.

Schulz KF, Chalmers I, Grimes DA *et al.* Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA. 1994;272:125-8.

Selman TJ, Johnson NP, Zamora J *et al.* Gynaecologic surgery from uncertainty to science: evolution of randomized control trials. Hum Reprod. 2008;23:827-31.

Urschel JD, Goldsmith CH, Tandan VR *et al.* Evidence-based Surgery Working Group. Users' guide to evidence-based surgery: how to use an article evaluating surgical interventions. Can J Surg. 2001;44:95-100.

Vail A, Gardner E. Common statistical errors in the design and analysis of subfertility trials. Hum Reprod. 2003;18:1000-4.

Vandekerckhove P, O'Donovan PA, Lilford RJ *et al.* Infertility treatment: from cookery to science. The epidemiology of randomized controlled trials. Br J Obstet Gynaecol. 1993; 100: 1005-36.