**Clinical andrology: from evidence-base to ethics**

**The ‘E’ quintet in clinical andrology**

Frank Comhaire

Ghent University Hospital, Center for Medical and Urological Andrology, De Pintelaan, 185, B-9000 Gent, Belgium. E-mail: Frank.Comhaire@rug.ac.be

This debate was previously published on Webtrack, July 13, 2000

The management of the infertile man should be founded on consensus-based medicine, i.e. the consensual opinion of experts considering evidence-based as well as empirical or experience-based medicine, the effective cumulative rate of successful deliveries, ethical and economic considerations. The apparent contradictions between conclusions from experience-based medicine and evidence-based medicine regarding the efficacy of varicocele treatment and tamoxifen treatment can be explained by scientific reasons. It is argued that the suggestion not to implement these treatments is ill founded because of flawed meta-analyses. The effective cumulative rate of successful deliveries and time to pregnancy as observed in cohort studies should be considered the ultimate touchstone of treatment efficacy. Based on the data of effective cumulative delivery rate, cost per successful delivery, and the known prevalence of aetiological diagnoses in infertile men, it is possible to estimate the number of deliveries that can be attained thanks to an investment of, e.g. 1 million Euro. This number is ~70–80 if IVF (including intracytoplasmic sperm injection) is chosen as first line treatment, and four times higher if conventional treatment (including intruterine insemination) is applied. It is concluded that the well thought out approach recommended by the World Health Organization should generally be implemented for the management of couples in whom infertility is (mainly) due to a male factor.

Key words: andrology/E quintet/infertility/male factor

**The ‘E’ quintet**

The value of any therapeutic approach, and of the treatment of male infertility in particular, must be assessed on the basis of five principles that can be summarized as the ‘E quintet’. The recommendations of evidence-based medicine (conscientious explicit and judicious use of current best evidence) must be compatible with empirical medicine (care based on knowledge acquired from experience and observation). The present trend is for established experts to convene and to consider all available evidence in order to formulate a consensual view (consensus-based medicine). The effect of treatment of infertility must be assessed by means of the effective cumulative delivery rate and time to pregnancy. Ethical considerations are of pivotal importance, and economic aspects are part of ethics. Here, I address specific concerns relating to the implementation of the E quintet in the modern approach to infertility due to a male factor.

**Problems with evidence-based medicine**

Clinical research in the field of andrological aspects of human reproduction is difficult. The end-point of male infertility treatment can be either the occurrence of pregnancy in the partner, or improvement of semen quality. The former largely depends on the fertility potential of the (usually unique) female partner, which has been found to be impaired in as many as half of the couples with ‘male factor’ infertility [World Health Organization (WHO), 1987]. ‘Optimizing’ the fertility potential of the female will increase the pregnancy rate of those couples. The latter end-point, i.e. change in semen quality, may be difficult to interpret because of poor reproducibility of basic semen analysis. Also, the correlation between sperm quality and fertility potential is not very strong, though it is clearly documented (Bostofte et al., 1990; Wichmann et al., 1994; Bonde et al., 1998).

To circumvent these difficulties in evaluating the effectiveness of treatment of the infertile male, several strategies can be adopted. In order to reduce the influence of the ‘female factor’, randomized trials including a large number of couples can be performed. These usually require a multicentre design, which may decrease their reliability due to differences in diagnostic accuracy and therapeutic (e.g. surgical) skills (Mastroianni, 1999). Some large-scale randomized trials have been performed in one single centre. These too should be considered with caution, since inclusion/exclusion criteria may not be strictly adhered to in order to permit recruitment of a sufficient number of cases. Randomization and informed consent may not reach international standards, and external peer review is uncommon. Problems with standardization of semen analysis can be overcome by exchanging video recordings and morphology slides (Giwercman et al., 1999), but these measures are rarely applied in single centre trials.

An example of contradictory conclusions from apparently similar trials is the effect of varicocele treatment. Recently, the results of several prospective and randomized trials have been published comparing the pregnancy rate after immediate varicocele treatment versus treatment being postponed for 12 months. The single centre trial (Nieschlag et al., 1995, updated in 1998), referred to as trial A, concludes ‘counselling to be as effective as occlusion of the vena spermatica’ for the treatment of varicocele associated infertility. The WHO multi-
centre trial on varicocele surgery (Hargreave, 1997, referred to as trial B) concludes immediate varicocele surgery to be an effective mode of treatment. Both studies include a similar number of couples of whom a similar proportion completed the protocol (trial A: 125 out of 203 = 61.6%, trial B: 135 out of 238 = 56.7%, difference between the studies not statistically significant). Also, the success rate in the group of cases randomized to immediate treatment was similar in the two studies: 29.0% in trial A compared with 31.3% in trial B. The difference occurred in the prevalence of pregnancies in the control groups, which was 14.4% during the 12-month follow-up in trial B, compared with 25.4% in trial A. As a result there was no difference between controls and treated cases in trial A, whereas trial B revealed a significant positive effect of immediate treatment [relative risk (RR): 2.32, confidence interval (CI): 1.43–3.77, \( P < 0.01 \)].

All couples in trial A were ‘counselled’, independent of whether the men received immediate treatment or varicocele surgery was postponed. Counselling included ‘monitoring each partner’s reproductive function by the gynaecologists’ (Nieschlag et al., 1995), which apparently included optimizing her fertilizing potential. Monitoring of the female seems to have improved the couple’s fertility, since the monthly conception rate in the counselled controls (2.6%) was significantly higher (\( P < 0.01 \)) than that of couples treated with placebo as part of other studies performed in the same centre (0.4%, Kamishke et al., 1998; Rolf et al., 1999). Hence, the conclusion that varicocele treatment is ineffective, which is suggested from the equivalence in trial A is invalid, since application of an effective treatment to the female partner made it impossible to assess the effect of treatment of the man. This conclusion is reinforced by the observation that the relatively high pregnancy rate in the ‘control group’ of trial A was attained in spite of unchanged semen variables, and that the occurrence of conceptions in this group was independent of the duration of infertility. The latter stands in contrast with results of several other publications (Collins et al., 1983; WHO, 1984; Comhaire, 1987; Snick et al., 1997). In addition, it is not permissible to simply ‘add up’ the results of trials A and B in meta-analysis (Kamishke and Nieschlag, 1999).

**Pitfalls of type II or \( \beta \) errors**

Type I or \( \alpha \) error means the coincidental finding of a statistically significant difference, in spite of the absence of a ‘real’ difference. This relates to the chosen limit of statistical significance, e.g. if the limit is set at 0.05, approximately one out of every 20 (5%) of statistical tests may give a significant result. In type II or \( \beta \) error, a ‘real’ difference remains undetected upon statistical testing, usually because of too few observations. There are many examples of incorrect conclusions because of type II error (Smith, 1994; Rolf et al., 1999), and caution must be taken not to interpret ‘no evidence of effect’ as ‘evidence of no effect’ (Savulescu et al., 1996).

A typical example of type II error is the ‘tamoxifen case’. Meta-analysis of randomized double-blind studies with anti-oestrogen treatment (either clomiphene citrate or tamoxifen) for male infertility taking pregnancy rate as an outcome, gives an odds ratio of 1.54 in favour of treatment. This is not significant at the 5% level, since the 95% CI includes the value 1 (CI: 0.99–2.40). The reviewers of the Cochrane group (Vandekerckhove et al., 1999) note that ‘the stronger treatment effect observed in the trials of lower methodological quality seemed not to be due to a higher pregnancy rate in the treated groups, but to the lower pregnancy rate in the controls of these trials compared with the controls of the better quality trials (7.0 versus 12.5%).’ This sentence tends to discredit the conclusions of certain trials. In fact, the finding can be expected on the basis of the hyperbolic regression between sperm concentration and fecundability (Bonde et al., 1998) (Figure 1). From these data it can be calculated that doubling sperm concentration, e.g. from 4 to \( 8 \times 10^6 \) spermatozoa/ml multiplies the monthly conception rate by 2.58, from 3.3 to 8.5%. However, doubling sperm concentration from 8 to \( 16 \times 10^6 \) or from 16 to \( 32 \times 10^6 \) spermatozoa/ml increases fecundability from 8.5 to 11.4% and from 11.4 to 15.1% respectively (multiplication factors 1.34 and 1.32). Treatment that doubles sperm concentration (as reported for tamoxifen) will therefore have a stronger effect on the probability of conception when initial sperm count is lower (e.g. geometric mean in our own patient material = \( 5.1 \times 10^6 / \) ml) than when initial sperm concentration is higher (mean \( 13.1 \times 10^6 / \) ml in 17 published trials). Since a lower initial sperm concentration is correlated with a lower fecundability, it is not surprising that the success rate of tamoxifen treatment in terms of pregnancy rates is stronger in trials with lower treatment-independent pregnancy rate.

In addition, several controlled and observational studies on the effect of tamoxifen on sperm concentration have remained inconclusive. However, the average number of cases included in these trials is only 18, compared with 68 in the trials revealing a significant increased sperm concentration (\( \chi^2 \) for difference between groups: \( P = 0.029 \)). This suggests that the negative outcome of certain trials is related to the small number of observations rather than to a lack of efficacy, and therefore to type II error.
Effective cumulative delivery rate (ECDR) and time to pregnancy (TTP)

The aim of infertility treatment is to attain delivery of healthy offspring in as many couples as possible and in as short a period of time as possible. In order to investigate these end points, cohort studies and the calculation of the effective number of pregnancies resulting in successful delivery attained after a well-defined period of time is recommended (Figure 2). Calculation of the theoretical probability of conception (De Vries et al., 1999) does not take into account the proportion of couples that have abandoned treatment, reducing their probability of attaining pregnancy. The discontinuation rate is highly variable between treatment modalities. It is high after failed initial IVF (Stolwijk et al., 1996; Roest et al., 1998; Osmanagaoglu et al., 1999) but low or non-existent during medical treatment with tamoxifen or after varicocele treatment respectively. Also, the time interval between subsequent attempts at IVF is often long, reducing the ECDR in spite of a high success rate per attempt. As a result, the ECDR after 12 months is not significantly different in couples treated by IVF from that of couples treated for varicocele or with tamoxifen (Comhaire et al., 1995).

Furthermore, the time needed to attain a defined number of pregnancies may be different between different modes of treatment, even if the total number of pregnancies that will ultimately occur is similar. This has been documented for different methods of sperm preparation and cycle monitoring of the female partner in intrauterine insemination (IUI) (Depypere et al., 1995). Also, in the WHO varicocele trial, the time needed to attain pregnancy in 20% of couples was significantly shorter if varicocele was treated immediately (160 days) compared with 316 days in couples where varicocele treatment was postponed for 12 months. This proves that immediate varicocele treatment is more effective than delayed treatment.

Ethics and economic aspects

Recently, the results of a very large multi-centre trial on IUI have been published (Guzick et al., 1999) and discussed (Mastroianni, 1999). Patients who were treated by ovulation induction and IUI had a higher pregnancy rate (33%) than those who were treated with IUI only (18%) and those who were treated with ovulation induction and intracervical insemination (19%). Those in the intracervical insemination-only group had the lowest fecundity rate (10%). However, the live birth rate after IUI with or without ovulation induction was not significantly different (Figure 3) because of the lower birth rate per pregnancy in the stimulated group. Furthermore, 24 out of 25 multiple pregnancies occurred in the latter group. It is generally accepted that multiple pregnancies must be avoided since they are hazardous to both the mother and the offspring. Therefore, it should be concluded that ovulation induction with or without IUI is ethically unacceptable, but that IUI without ovulation induction is an effective mode of treatment for couples with unexplained infertility.

Ethical rules compel medical doctors to prescribe the best available established (standard) treatment in all cases. This implies that new treatment modalities, such as antioxidants (Rolf et al., 1999) or recombinant FSH (Kamishke et al., 1998), should only be used in controlled trials if treatment of established effectiveness is also given. Since ‘counselling’ was shown to be a more effective treatment than placebo (vide supra), it may be mandatory to counsel all couples included in double-blind trials. Hence, the implementation of ethical rules makes it extremely difficult to design strictly randomized trials, although there still is an urgent need for such trials based on biologically well-founded hypotheses and designs (Mahmoud and Comhaire, 2000).

Finally, ethics imply that the financial resources available for health care must be employed in the most cost-effective manner. This is particularly the case in providing effective, practical and inexpensive treatments for underprivileged people, or couples who cannot accept assisted reproductive...
Figure 4. Estimated cost per successful delivery resulting from different modes of treatment of male infertility. TLC = tender loving care (controls, counselling); VAR = varicocele treatment by means of transcatheter embolization or surgery on an outpatient basis; Tamox = treatment with tamoxifen; IUI-Percoll = intrauterine insemination of spermatozoa selected on a density gradient column (1 = first cycle, 2 = second cycle, 3 = third cycle); IVF = conventional in-vitro fertilization; ICSI = IVF with intracytoplasmic sperm injection (Comhaire, 1995).

This may turn out to be extremely difficult in couple infertility caused by a ‘male factor’. The (negative) outcome of certain randomized trials or meta-analyses contrasts with the positive evidence empirically collected from carefully performed observational studies. It may be counterproductive simply to reject the empirical evidence as unreliable; rather attempts should be made to explain scientifically the cause of these contradictions. In addition, ethical aspects must play a pivotal role in the approach taken to cure the infertile male and economic factors must also be considered. It is concluded that the well thought-out andrological approach recommended by WHO (Rowe et al., 2000) should be implemented by all doctors who are involved with the management of infertile couples.

Conclusions

While implementation of the rules of evidence-based medicine should be the preferred approach to any medical treatment, this may turn out to be extremely difficult in couple infertility caused by a ‘male factor’. The (negative) outcome of certain randomized trials or meta-analyses contrasts with the positive evidence empirically collected from carefully performed observational studies. It may be counterproductive simply to reject the empirical evidence as unreliable; rather attempts should be made to explain scientifically the cause of these contradictions. In addition, ethical aspects must play a pivotal role in the approach taken to cure the infertile male and economic factors must also be considered. It is concluded that the well thought-out andrological approach recommended by WHO (Rowe et al., 2000) should be implemented by all doctors who are involved with the management of infertile couples.

References


Karande, V.C., Korn, A., Morris, R. et al. (1999) Prospective randomized trial comparing the outcome and cost of in vitro fertilization with that of a...


