Conventional treatments of male infertility in the age of evidence-based andrology

A.Kamischke and E.Nieschlag¹

Institute of Reproductive Medicine of the University (WHO Collaborating Centre for Research in Human Reproduction), Domagkstrasse 11, D-48129 Münster, Germany

¹To whom correspondence should be addressed

The field of reproductive medicine is characterized by the rapid development of assisted fertilization. Today, fertilization can be achieved by cytoplasmic microinjection of single spermatozoa or spermatids into oocyte cytoplasm. In view of these new reproductive techniques, conventional diagnosis and treatment of male infertility seem to be rather outdated and limited. However, the final goal of reproductive medicine should be to improve reproductive function in order to provide each couple with the chance of conceiving offspring naturally. In this paper we analyse the effectiveness of current conventional treatments of male infertility in the light of high-quality studies and emphasize that future therapies must be evaluated in properly conducted studies.

Key words: andrology/clinical trials/hormonal treatment/male infertility

Introduction

The past 20 years have witnessed rapid growth in health care information. Every month, clinicians, policy makers and patients face a deluge of evidence published in 20 000 journals catering to the health care community. These reports are now so numerous and dispersed that it is unreasonable to expect people to read and retain all the information they contain. Because of this mass of information, most health care providers rely on reviews, which claim to provide an overview of results of primary research. Unfortunately, the quality of at least some reviews leaves much to be desired (Antczak-Bouckoms, 1995).

The most important factor influencing the ability of reviews and practised medicine to enhance personal experiences and the readers’ own views is the quality of the primary basic and clinical studies reviewed. Although many of the studies were conducted with the best of intentions, some did more harm than good. In experimental designs, randomized, controlled prospective clinical trials are regarded as the most reliable method of proving the effectiveness of diagnostic and therapeutic strategies. Too often, however, the results of high-quality experimental trials and high-quality, observational, quantitative epidemiological studies are ignored. There are several reasons for this, and important among them are the simple lack of awareness of trial results and the confusion which arises when several trials of the same topic provide inconclusive or conflicting results.

These problems are common to all clinical disciplines. Andrology, and the question of male infertility treatment in particular, present further problems. One reason is that the pathogenesis of many fertility disturbances has not yet been elucidated and therefore rational approaches to treatment are lacking. In ~30% of cases no obvious cause for abnormal seminal parameters can be found and the condition is classified as ‘idiopathic infertility’ (Nieschlag and Behre, 1997). Another large proportion of infertile men present with varicoceles, leukocytes in seminal plasma or sperm antibodies. Whether such findings are, however, causative of impaired fertility or just coincidental has not yet been convincingly proven. The second problem of conventional treatment of male infertility is the definition of the endpoints of therapy. Although the induction of pregnancy and the birth...
Conventional treatments of male infertility

of a child are clearly defined endpoints, they require the interaction of two individuals. Therefore, factors outside the andrological realm influence the success of andrological therapy and have to be considered and evaluated in couple diagnosis and treatment.

Good clinical trials: implications for evidence-based andrology

No valid laboratory model exists for male infertility, and conventional therapeutic strategies often rely on speculative concepts and clinical observations. Even when derived from pathophysiological laboratory findings, therapeutic concepts must be confirmed or challenged by systematic observations in clinical trials. Clinical trials should be conducted in agreement with the rules of good clinical practice, which provide an international ethical and scientific quality standard for designing, conducting, recording and reporting trials that involve the participation of human subjects [World Health Organization (WHO), 1993b; International Conference on Harmonisation (ICH), 1996]. Good clinical practice rules obligate the medical community to lead the way to appropriate clinical evaluation of diagnostic tests and trials of clinical treatment procedures.

The concept of evidence-based medicine provides this appropriate tool, which de-emphasizes intuition, unsystematic clinical experience and pathophysiological rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research (Evidence-Based Medicine Working Group, 1992). The concept of evidence-based medicine is rooted in five interlinked ideas: (i) Clinical decisions should be based on the best available scientific evidence. (ii) The clinical problem (rather than habits or protocols) should determine the type of evidence to be sought. (iii) Identifying the best evidence means using epidemiological and biostatistical ways of thinking. (iv) Conclusions derived from identifying and critically appraising evidence are useful only if put into action in managing patients or making health care decisions. (v) Performance should be constantly evaluated (Davidoff et al., 1995). Therefore, literature-based evidence, and not personal skills and authority born of experience, is the major determining factor(s) in the outcome of research-oriented medicine when therapeutic decisions must be taken.

Compared with other fields of medicine, andrology was especially late in applying the paradigm of evidence-based medicine in clinical practice (Olive, 1986). Evidence-based andrology is, notwithstanding the exponential increase in the number of published randomized, controlled clinical trials concerning infertility treatment over the past decade, only in its infancy (O’Donovan et al., 1993; Vandekerckhove et al., 1993; Leifke and Nieschlag, 1996).

Basis of good clinical trials

The basis of good clinical experimental trials is summarized in Table I.

Diagnostic methods

Applied diagnostic procedures should be established in an appropriate spectrum of patients who are subjected to an independent, blind comparison of diagnostic tests, thus avoiding any kind of bias or systematic error (Jaschke et al., 1994). High sensibility and specificity of the applied methods provide the basis for good laboratory practice, which is concerned with the organizational process and the conditions under which laboratory studies are planned, performed, monitored, recorded and reported [Organisation for Economic Co-Operation and Development (OECD), 1992].

Proper andrological diagnosis

Clinical problem solving must rely on an understanding of the underlying pathophysiology. Clinical experience and the development of clinical instincts (particularly with respect to diagnosis) are crucial and necessary skills of a competent physician. Together with careful history taking and physical examination, these tools provide much, and often the best, evidence for diagnosis and direct treatment decisions. Therefore, the more the experienced clinicians can dissect the processes used in diagnosis, the greater the benefit for the patients and clinical trials (Campbell, 1987; Evidence-Based Medicine Working Group, 1992).

In addition to semen analysis, a comprehensive history and physical examination with particular emphasis on detection of any genital abnormality,
Table I. Basis of good clinical experimental trials

<table>
<thead>
<tr>
<th>Topic</th>
<th>Criteria which should be fulfilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic methods</td>
<td>High sensibility and high specificity of applied methods</td>
</tr>
<tr>
<td></td>
<td>Avoiding bias or systematic error</td>
</tr>
<tr>
<td></td>
<td>Valid internal and external quality control</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Clinical experience</td>
</tr>
<tr>
<td></td>
<td>Careful history taken and physical examination</td>
</tr>
<tr>
<td>Design</td>
<td>Results should provide scientific and clinical relevance (outcome research)</td>
</tr>
<tr>
<td></td>
<td>Fulfilling rules of good clinical practice (WHO, 1993a,b)</td>
</tr>
<tr>
<td></td>
<td>Prevention of any systematic bias by</td>
</tr>
<tr>
<td></td>
<td>Inclusion of control groups</td>
</tr>
<tr>
<td></td>
<td>Randomized allocation to the groups</td>
</tr>
<tr>
<td></td>
<td>Double blinding and placebo medication</td>
</tr>
<tr>
<td></td>
<td>High precision and external validity</td>
</tr>
<tr>
<td></td>
<td>Statistical power estimation a priori</td>
</tr>
</tbody>
</table>

such as maldescended testes, absence of the vas deferens, occlusions, phimosis or varicocele, should be included in the patient’s work-up. Further evaluations should routinely include ultrasonography of the scrotal organs, as infertile patients often present various abnormalities. Endocrinological evaluation of reproductive hormones helps to identify the cause of infertility and is mandatory in the diagnosis of obstructive azoospermia and differentiation between hyper- and hypogonadal hypogonadism (e.g. Klinefelter syndrome versus Kallmann syndrome; ESHRE Capri Workshop, 1996; Behre et al., 1997).

Because of the subjective nature of much of andrology, it is difficult, but feasible, to implement quality guidelines into the sperm assessment methods, as has already been done for hormones (WHO, 1993a). A continuous record of technician performance is necessary and is the purpose of internal quality control (Cooper et al., 1992). It is equally obvious that, in addition to internal quality control, external quality control schemes which aim to ensure agreement between different centres are essential if men with the same problems are not to receive different diagnosis and treatment in different centres or countries as a result of different diagnoses (Cooper, 1996). Internal and external quality control can be achieved by standardizing the methods of sperm preparation and examination (WHO, 1992a) and use of equipment to set absolute values for sperm concentration (flow cytometry) and motility (computer-assisted sperm analysis, CASA). For sperm morphology, absolute values do not yet exist, but internal and external quality control could be achieved by repeated assessment of the same morphology slide by different technicians and centres, as a recent enquiry among the Training Centres for Andrology of the European Academy of Andrology (EAA) has shown (Cooper, 1996).

Apart from the minimum requirements of andrological patient work-up, various standardized and unstandardized diagnostic sperm function tests have been implemented in andrology without any diagnostic advance. Therefore, urgent consideration must be given to the development of sperm function tests that are able to provide a likelihood of pregnancy for given diagnostic situations within certain timeframes (ESHRE Andrology Special Interest Group, 1996). In general, useful diagnostic fertility tests can be classified into three categories: (i) abnormal test results that are consistently correlated with impaired fertility. In each case, when the test result is unequivocally abnormal, fertility is unarguably impaired without therapy; (ii) abnormal test results that are not consistently correlated with impaired fertility. For these diagnostic tests, abnormal results are frequently associated with subsequent fertility without therapy; (iii) abnormal test results that do not appear to be correlated with impaired fertility. For these diagnostic tests, either there are data that confirm the lack of a correlation.
with pregnancy, or such follow-up studies do not exist (ESHRE Capri Workshop, 1996).

Good laboratory practice and proper application of the diagnostic procedures are the tools of good clinical practice, which is the substantial basis for good clinical trials and evidence-based medicine/andrology (WHO, 1993a,b; ICH, 1996).

**Considerations for design of good clinical trials**

**Providing evidence**

Regardless of the aim of the clinical trial, it should provide evidence of scientific and clinical relevance (outcome research). The quality of the study and the resulting evidence for its conclusion depend on the quality of the diagnostic methods, proper diagnosis and on the design of the study. The quality of a trial can be defined as the extent to which its design and conduct are likely to have prevented systematic bias. Mainly, there are four kinds of bias to avoid: (i) selection bias (systematic differences in the comparison groups), (ii) performance bias (systematic differences in the care provided apart from the intervention being evaluated), (iii) exclusion bias (systematic differences in withdrawals from the trial) and (iv) detection bias (systematic differences in outcome assessment).

In addition to the quality of the clinical trial, there are two other important aspects: precision (i.e. to what extent the clinical trial avoids the likelihood of random errors) and external validity (i.e. to what extent the outcome results are generalizable or applicable). These rules for high-quality studies apply to both experimental treatment and observational studies. Some of the general principles and limits concerning the design of such studies are discussed below in the section ‘Conventional male infertility treatment’.

**Providing evidence by experimental treatment studies**

**Aim and validity of a study:** Precise definition of the question to be answered forms the basis for the ensuing clinical trial (WHO, 1993b; ICH, 1996). The patient’s disease, its treatment and outcome as well as the clinical context, e.g. the existence of ancillary treatment, co-morbidity and male-independent female factors, should be defined before patients are assigned to treatment groups (inclusion criteria). The broader the clinical context, the more generalizable will be the results (external validity) of the trial; however, this will also render it more likely that a possible effect will be masked or biased by unknown confounding factors. Homogenization by precise inclusion or exclusion criteria, rules for drop outs and stratification help to avoid this problem, but will also limit the validity of the results and narrow the spectrum of patients or clinical context.

**Control groups, randomization and crossover designs:** ‘Controlled’ studies compare one or more treatments to one or more control groups. Control groups are essential to exclude performing bias. Especially in conditions such as severe oligozoospermia, in which variable semen parameters are observed, spontaneous changes are likely to be mistaken for a genuine therapeutic effect. Furthermore, when performing a study, it should be taken into account that no medical intervention is without an effect (Kleijnen et al., 1994). In order to discriminate between intervention-associated specific and unspecific effects and avoid biases, untreated control groups should be included in clinical studies.

Randomization of treatment reduces selection bias by ensuring balance between known and unknown covariables and between comparison groups. The ideal process of allocation to the treatment groups is achieved if an assignment schedule generated using true randomization is administered by someone who is not responsible for recruiting participants (Altman and Dore, 1990). After the randomized assignment to the treatment groups has been revealed, no alteration of the assignment or decision about eligibility is possible.

‘Crossover’ refers to a study design in which more than one intervention was administered in a definite order, one following another. As at least the second treatment is not assignment randomized to the participant, crossover studies are not truly randomized. For example, the evaluation of drugs can be biased by carry-over effects of prolonged action. This is particularly pertinent when considering the duration of spermatogenesis and the factors influencing it. In addition, if pregnancy is the ultimate goal of therapy, responders will leave
A.Kamischke and E.Nieschlag

the first part of the study and thus bias the second part by dropping out.

**Blinding and placebo effects:** In a ‘blinded’ study, the participants and/or the investigators (single/double blind) evaluating its outcome are deliberately kept unaware of the group to which the participants have been assigned. The purpose of blinding is to minimize the bias that might result from knowing which intervention a participant is receiving (detection and performance bias; Gotzsche, 1994). Trials which blind outcome assessors regarding treatment allocation should be rated more highly than trials which do not include any blinding. The placebo effect is the difference in outcome between a placebo-treated group and an untreated control group in an unbiased experiment (non-specific effect). In clinical trials, placebo treatment versus experimental treatment reduces detection and performance bias and should be provided in a double blind manner (Kleijnen et al., 1994).

**Statistical considerations and multicentre studies**

The significance of investigations depends on the incidence of outcome parameters. Statistical power estimation (internal validity) prior to the recruitment phase should provide an estimate of the number of patients required to exclude for a low probability of false-positive (type I error) as well as for low false-negative (type II error) errors (Freiman et al., 1978). The ‘minimal important benefit’ defines the smallest relevant difference in function that a patient and clinician agree to be worth the risks and trouble of offering and accepting the treatment (Sackett and Cook, 1993; WHO, 1993b). Cumulative pregnancy rates on the basis of life table calculation are the best parameter for outcome research of male infertility treatment, although they may not always be a practicable and therefore reliable endpoint because of low pregnancy rates among infertile couples, interference by additional female factors and violations of the protocol by resorting to performance of assisted reproduction techniques.

Because pregnancies among infertile couples occur relatively infrequently, the number of patients that has to be included in a trial on the basis of a power estimate is usually larger than the number available in a given centre at a certain period and the number of patients that can be handled by the investigators in a reasonable time. Therefore, multicentre studies provide larger sample sizes and more clinical judgement than single-centre studies. Unfortunately, multicentre studies are often complicated by systematic inhomogeneity between the centres, and they depend on the assumption that the differences between the compared groups are constant from centre to centre (performing and detection bias). Because of the general lack of internal and external quality control in andrology laboratories, performing and detection bias are further common difficulties in multicentre studies.

**Meta-analysis**

Meta-analysis is a method of summarizing the results of a number of randomized trials. This statistical approach is no doubt useful if applied correctly. Meta-analyses could provide evidence if the results of several experiments are contradictory or inconclusive. In addition, even if the results of single high-quality studies are consistently non-significant, a meta-analysis might show an overall significant effect. A meta-analysis is only as good as the included trials (Whitehead and Jones, 1994). Meta-analysis of results from trials of variable quality can result in false-positive or false-negative conclusions if the less-rigorous studies that are included are biased towards overestimating or underestimating the effectiveness of the intervention being evaluated (Detsky et al., 1992).

Results of meta-analysis should be presented as odds ratios or relative risks. Odds ratios and relative risks greater than one (risk differences less than zero) always indicate that treatment is better. Where the 95% confidence interval is entirely above or below 1, the treatment group did significantly (at the 5% level) better or worse than the placebo control group (for more details, see Morris and Gardner, 1988). If confidence intervals for the results of the included studies overlap, differences between single studies should be considered statistically insignificant and therefore homogeneous (Walker et al., 1988). In cases of homogeneity, combined odd ratios for every treatment can be calculated (example in Figure 3).
**Interim conclusion**

Because of the above-mentioned considerations, the statistically valid, (placebo) controlled, randomized, double-blind clinical trial is today regarded as the most reliable method of proving the effectiveness of diagnostic and therapeutic strategies. This kind of study provides an excellent tool to show whether as yet unproven therapies are superior to no or existing treatments and are therefore recommended by drug and health authorities to evaluate the effectiveness of new drugs.

**Providing evidence by observational studies**

Observational designs are not in contrast to experimental studies or evidence-based andrology, but they have to be seen as complementary approaches. The principal observational epidemiological methods are non-randomized trials, cohort studies (prospective and retrospective) and case control studies (Black, 1996). Whether an experimental or observational approach to the problem is preferable depends mainly on the question to be answered.

Observational studies are adequate to demonstrate effectiveness when the intervention is dramatic and unknown confounding factors can be ignored, as in hormone substitution therapy in diabetes mellitus, hypothyroidism or hypogonadotrophic hypogonadism. Observational studies might also be preferable in evaluating interventions designed to prevent very rare events or in detecting rare adverse events as well as when the outcome of the intervention is far in the future (Black, 1996). For example, because of limitations in study size, the influence of therapy for maldescended testes in early infancy on subsequent development of testicular cancer and infertility might only be detectable by observational studies. Observational studies might sometimes also be more appropriate than experimental studies if the patient’s beliefs and wishes and the clinician’s attitudes, experiences and beliefs are crucial to determining the success of the intervention. A special problem with the placebo effect arises in experimental surgery studies, because placebo (sham) operations (i.e. for the treatment of varicocele) are unethical in humans (Johnson, 1994). Observational studies may sometimes provide a rationale for performing randomized controlled trials (Hennekens and Buring, 1993).

**Conventional male infertility treatment**

In this section, examples of therapies for male infertility are discussed in view of high-quality, properly designed, truly randomized, placebo-controlled, at least single-blind studies (with the exception of varicocele) which have pregnancy as the main outcome measure. Crossover studies were only included in this analysis if pregnancy rates could be separately calculated from the overall outcome for the first treatment phase. In areas where no high-quality studies could be identified or strong observational evidence existed, treatment options were discussed on the basis of the trials available. The studies were identified by computerized library searches (Medline and Cochrane) under the keywords comprising the special issues listed in the following subheadings. In addition, it is strongly recommended that optimization of the female reproductive functions, adequate deposition of the ejaculate and appropriate timing of sexual intercourse in relation to ovulation are the basis of all male infertility treatments (Wilcox et al., 1995).

**Secondary hypogonadism**

Testosterone serum concentrations, testicular growth and appearance of spermatozoa in the ejaculate are clearly definable outcome parameters in secondary hypogonadism, where specific supplementation of what the endocrine system endogenously lacks will lead to dramatic reconstitution of male reproductive functions. Therefore, observational studies were adequate to demonstrate the effectiveness of gonadotrophin-releasing hormone (GnRH) or gonadotrophin therapy (Black, 1996), and evidence for their effectiveness has been provided.

To date, however, it has not been possible to show conclusively whether human chorionic gonadotrophin/human menopausal gonadotrophin (HCG/HMG), recombinant gonadotrophins (Kliesch et al., 1995) or GnRH are superior in achieving paternity in patients with idiopathic hypogonadotrophic hypogonadism (IHH; Schopohl et al., 1991; Kliesch et al., 1994). This would...
require controlled randomized trials comparing the therapeutic entities. As the number of patients with IHH is small and its clinical appearance is heterogeneous, studies providing sufficient statistical power are almost impossible to perform in a single institution; thus treatment of IHH would be appropriate for prospective, randomized, controlled multicentre trials.

**Testicular maldescent**

A history of maldescended testes is frequent among infertile patients, although only a few cases attending an infertility clinic present with testes that are still in an abnormal position (Nieschlag et al., 1997). For prevention of infertility and to reduce the risk of testicular cancer, it was usual in the past to treat maldescended testes in advanced childhood or in puberty. In the course of the past 20 years it has become accepted that the position of the testes should be corrected by the end of the first year of life. There is evidence that descent can be achieved in uncomplicated cases by intranasal GnRH or intramuscular HCG therapy, although truly randomized controlled trials have not been performed. If endocrine therapy fails twice, an orchidopexy should be performed [Deutsche Gesellschaft für Endokrinologie (DGE), 1991]. Whether treatment of maldescent earlier in childhood, as is practiced today, is superior to later treatment is unclear, as the patients treated earlier are only beginning their reproductive phase of life and high-quality studies on this topic are lacking (Nieschlag, 1993). For treatment of maldescent-related infertility in general, the same treatment options as in idiopathic male infertility have to be considered.

**Varicocele**

Varicoceles are the most frequent physical finding in infertile men (WHO, 1992c). Based on these observational data and the speculative concept that varicoceles may cause testicular and epididymal damage while elevating intrascrotal temperature, occlusion of the spermatic vein by ligation or embolization has long been accepted as the treatment of choice (Takihara et al., 1991). Most studies have assessed treatment success by the physical disappearance of the Valsava-positive varicocele and the purported improvement of male fertility has hardly been assessed by controlled clinical trials aiming to appraise pregnancy rates. In addition, in most studies, female factors were not clearly defined by inclusion or exclusion criteria. Recent studies that corrected for these possible biases have suggested significant benefits for male fertility after surgical varicocele repair compared with unoperated controls (Madgar et al., 1994; Hargreave, 1995). In these latter studies, the control groups remained completely untreated, not even counselled, and thus the existence of unspecific (placebo) effects associated with any medical intervention was ignored. In contrast to these studies, a proper randomized, controlled, prospective trial recently showed no significant difference in the cumulative pregnancy rate over a 12-month period between the control and treatment groups when the control group was counselled throughout the follow-up time (Figure 1; Nieschlag et al., 1995). This study again emphasizes the importance of properly designed control groups and the significance of bias. In order to evaluate the magnitude of counselling effects, a design with three groups, i.e. untreated, counselled and surgically or radiologically treated varicoceles, would be preferable, but a sham-operated group cannot be included.

![Figure 1. Cumulative pregnancy rates over 12 months in 48 untreated patients (solid line) and 47 treated patients (dotted line) with varicocele (Nieschlag et al., 1995).](https://academic.oup.com/humrep/article-abstract/13/suppl_1/62/789096/162786086?ref=1535073355)
Conventional treatments of male infertility

for ethical reasons. Since it remains questionable whether interventional treatment is superior to no treatment or to counselling in terms of fertility, intervention can no longer be recommended.

Infections

Prevention of infections of the reproductive organs plays a more important role than treatment. Before antibiotics were introduced, occlusions of the efferent ducts as sequelae of venereal diseases were the most common reasons for male infertility and still prevail in some areas in sub-Saharan Africa (Nieschlag, 1993). The aetiology and evidence of effective antibiotic treatments in leukocytospermia and subclinical genital tract infections remain unclear. Randomized controlled studies have shown a high resolution rate in the placebo group. These studies result in the recommendation that patients with a single positive leukocytospermia test should not be given antibiotics without further evidence of a specific genital tract bacterial infection (Comhaire et al., 1986; Yanushpolsky et al., 1995). Because of the low number of good quality studies, the non-efficacy of antibiotic treatment of subclinical genital infections should be investigated further in truly randomized trials that have pregnancy or at least a convincing increase in sperm quality as the outcome.

Immunological male infertility

The concept of immunological infertility is not yet well established. The term refers mainly to the detection of sperm antibodies in the seminal fluid of men whose infertility is otherwise unexplained. Findings based mainly on laboratory test support the relevance for male fertility of the presence of immunoglobulin (Ig)G and IgA antibodies on the sperm surface (Eggert-Kruse et al., 1991; Mahmoud et al., 1996). A prospective and controlled study demonstrated significantly lower sperm motility and pregnancy rates when anti-sperm antibodies of the IgG, and particularly of the IgA, class were present in semen and yielded a positive mixed agglutination reaction (MAR) test >30% (Eggert-Kruse et al., 1995), while other studies showed no effects of anti-sperm antibodies on fertility (Collins et al., 1993).

Immunosuppressive agents such as glucocorticoids have been tried as treatment on the grounds of immunological considerations. Most studies concerning this issue were uncontrolled or not truly randomized, and only a few investigated pregnancy as an outcome (O'Donovan et al., 1993; Howards, 1995). Qualitatively adequate studies of corticosteroid treatment (Figure 3) have claimed both efficacy (Haas and Manganiello, 1987) and lack of efficacy (Bals-Pratsch et al., 1992).

As the clinical results are at least conflicting and it is still unclear whether and to what extent anti-sperm antibodies are causative for fertility problems in men, well-designed clinical studies should readdress this issue. Such studies require that laboratory findings concerning special subgroups of antibodies or their titre within seminal plasma should be better defined and that randomized treatments should be performed after a certain class or titre of antibodies is found in semen.

Idiopathic male infertility

One third of our patients, representing the largest group of those consulting for infertility, suffer from ‘idiopathic infertility’, which is characterized by subnormal semen parameters and possibly also by pathological hormone concentrations whose cause is unknown (Nieschlag and Behre, 1997). Despite the absence of clear pathophysiological concepts, several treatments have been recommended.

Non-hormonal therapies

Kallikrein is a glycoprotein involved in the enzymatic activation of kininogens that has been shown to have a positive effect on sperm motility in vitro and on stimulation of sperm metabolism and to improve cervical mucus penetration (Schill, 1978). Kallikrein administered orally has been widely used in andrological practices for idiopathic male infertility. Sixteen randomized trials have been performed with this drug, but only four were truly randomized and included a placebo-treated control group (Bedford and Elstein, 1989; Izzo et al., 1984; Glezerman et al., 1993; Keck et al., 1994). Two single studies have shown a significant positive effect of kallikrein on pregnancy rates (Schill, 1978; Micic et al., 1985). Unfortunately, both were methodologically impaired. The first (Schill, 1978) was only pseudorandomized and recorded pregnancies retrospectively over a post-treatment period.
of 12 months, which is too long for pregnancies to be ascribed only to kallikrein, which has a very short half-life. The second study (Micic et al., 1985) was not double blinded. The higher quality studies (Figure 3) in contrast could not demonstrate any benefits of kallikrein on sperm parameters or pregnancy rates (Bedford and Elstein, 1989; Izzo et al., 1984; Glezerma n et al., 1993; Keck et al., 1994). The implication of these studies for clinical andrology clearly highlights the importance of statistical and outcome-valid controlled studies and questions the benefit of kallikrein for idiopathic infertility (for review, see Vandekerckhove et al., 1995).

Other non-hormonal therapies for idiopathic male infertility include pentoxifylline (Micic et al., 1988) and the antioxidants vitamin C (Dawson et al., 1992) and vitamin E (Moilanen and Hovatta, 1995). Existing data for pentoxifylline (Micic et al., 1988) are not conclusive, because of uncertain randomization and insufficient data. Despite a known existing pathophysiological role for the antioxidants, most published studies are not truly randomized. The truly randomized studies with vitamin C yielded only marginal improvements, with low statistical power and low external validity (Dawson et al., 1992). Moreover, the vitamin E study of highest quality has uncertain inclusion (e.g. concomitant medication) and exclusion criteria (Moilanen and Hovatta, 1995) and a statistical power too low to provide sufficient evidence. As efficacy cannot be ruled out on the basis of existing data, the antioxidants should be evaluated in good quality trials.

**Hormonal treatment**

Since GnRH, gonadotrophins and testosterone are required for normal testicular function (for review, Weinbauer et al., 1997) and are effective in the treatment of hypogonadism, their application has also been tried for treatment of idiopathic male infertility, although no endocrine defect has been demonstrated as a cause of this type of infertility.

It was suggested that oligoasthenoteratozoospermia in patients with elevated follicle stimulating hormone (FSH) might be caused by too infrequent GnRH pulses and hence that sperm parameters might be improved by pulsatile GnRH therapy (Wagner and Warsch, 1984). This could not be confirmed by a longitudinal uncontrolled study where patients were treated with GnRH pulses (Bals-Pratsch et al., 1989). Because of the negative result compared with pretreatment semen values, a controlled study appears not to be necessary, as it would have been if the result had been positive.

Gonadotrophins were applied for the treatment
Conventional treatments of male infertility

Kallikrein treatment

COR \( n=144 \)

HCG/HMG treatment

COR \( n=834 \)

Androgen treatment

COR \( n=244 \)

Antiestrogenic treatment

COR \( n=54 \)

Immunological infertility

COR \( n=54 \)

Varicocele

\( \text{Bedford and Elstein, 1981 } n=23^a \)
\( \text{Izzo et al., 1984 } n=30 \)
\( \text{Keck et al., 1994 } n=91 \)

\( \text{Knuth et al., 1987 } n=38 \)

\( \text{Mauss, 1974 } n=209 \)
\( \text{Aafjes et al., 1983 } n=59^a \)
\( \text{Scottish Infertility Group, 1984 } n=328 \)
\( \text{WHO, 1989 } n=105 \)
\( \text{Pusch, 1986 } n=37 \)
\( \text{Comhaire, 1990 } n=24^a \)

\( \text{Ronnberg, 1990 } n=29 \)
\( \text{Torok, 1985 } n=54 \)
\( \text{Sokol et al., 1986 } n=20 \)
\( \text{WHO, 1992 } n=141 \)

\( \text{Balu-Pratsch et al., 1992 } n=19^b \)
\( \text{Haas and Mangianni, 1997 } n=35 \)

\( \text{Nieschlag et al., 1995 } n=95 \)

Figure 3. Combined odds ratios (COR) of randomized, placebo-controlled clinical trials on male infertility with the outcome expressed as pregnancies per patient. Results were presented as COR with 95% confidence interval (CI) of all single studies mentioned on the right side of the figure. \(^a\)If pregnancies could be clearly allocated to treatment phases in crossover studies, data were calculated only for the first treatment phase. \(^b\)No pregnancy in one of the groups; one was added to all groups for statistical purpose.

of idiopathic infertility on the hypothesis that elevation of gonadotrophin values may lead to stimulation of spermatogenesis. The HCG/HMG treatment was used for many years. In a review, 39 uncontrolled studies were summarized that reported average pregnancy rates of 8–14% (Schill, 1986). In contrast to these studies, a randomized, double-blind, placebo-controlled study of HCG/HMG treatment for normogonadotrophic oligo-asthenoteratozoospermic men could not demonstrate any beneficial effect on sperm parameters or pregnancy rate (Figure 2; Knuth et al., 1987). This study provides evidence that the HCG/HMG treatment schedule did not improve semen parameters beyond random fluctuations.

In monkeys it was shown that FSH treatment alone was able to stimulate spermatogenesis (van Alphen et al., 1988). Previous uncontrolled attempts to treat patients with idiopathic infertility with HMG, i.e. FSH alone, failed (Pescosolido et al., 1985), which was claimed to be due to a loss of biopotency of serum FSH and of residual HCG activity. The availability of recombinant FSH and positive results in an observational study with pure FSH have led to reconsideration of FSH treatment for male infertility (Acosta et al., 1992; Simoni and Nieschlag, 1995). Other groups have also confirmed a benefit from pure FSH treatment in various uncontrolled trials. As no improvements in classical semen parameters in the treatment group were observed in some trials, in a different uncontrolled trial the efficacy of rhFSH was considered to be due to the better fine structure of the sperm head subcellular organelles (Bartoov et al., 1994). As all conducted studies published so far were not properly controlled, the conclusion that FSH is really efficacious in male infertility has still not been convincingly proven. This stresses the need for double-blind, placebo-controlled studies, which are now under way (Baker et al., 1992; Simoni and Nieschlag, 1995).

The requirement of testosterone for normal spermatogenesis has also led to the use of androgens in treatment of idiopathic male infertility. Mestosterone was used for more than 20 years in the treatment of idiopathic infertility on the basis of...
A. Kamischke and E. Nieschlag

uncontrolled studies. Truly randomized, placebo-controlled, double-blind, studies, however, clearly showed the ineffectiveness (Figure 3) of mestosterone and oral testosterone undecanoate treatment in controlled studies (for review, see Vandekerckhove et al., 1996).

It is well known that oestrogens can have a negative feedback on hypophyseal gonadotrophin secretion. Anti-oestrogenic compounds competitively blocking oestrogens at the receptor site might reduce the inhibitory feedback and lead to elevation of gonadotrophin and testosterone serum concentrations. Assuming that an increase in luteinizing hormone and FSH may improve sperm production, anti-oestrogens (in particular tamoxifen) have been widely used. None of the truly randomized, double-blind and placebo-controlled studies demonstrated any significant beneficial effect of tamoxifen therapy on fertility (Figure 3) (for review, see Rolf et al., 1996). Nor could any significant therapeutic effect be demonstrated for clomiphene, which is the other widely used anti-oestrogenic compound (WHO, 1992b). Therefore, anti-oestrogenic treatment has to be considered obsolete, especially in view of its potentially toxic side effects.

Conclusions

The above-mentioned examples of male infertility treatment highlight the importance of high-quality studies in andrology. As such studies are complex and time consuming, these problems should be tackled by certain centres and institutions which have developed a reputation for the quality of their work over time. Such centres (of excellence, as they are sometimes called) or organizations (such as WHO) have a specific responsibility to maintain the quality of their work (Leifke and Nieschlag, 1996) and focus their efforts on achieving new therapeutic regimes for male infertility.

New therapeutic strategies for male infertility are urgently required. These might be derived from increasing knowledge about testicular paracrinology (for review, see Spiteri-Grech and Nieschlag, 1993) and the biology of spermatogenesis (e.g. Blendy et al., 1996) These factors, in addition to endocrine regulation, may offer a great opportunity for successful infertility treatment and should be thoroughly explored with the help of high-quality studies.

Acknowledgements

A. Kamischke is a recipient of a fellowship from the Federal Health Ministry (Bonn) and the Klinische Forschungsgruppe für Reproduktionsmedizin e.V (Münster). In addition, our own recent work reported here was supported by the Deutsche Forschungsgemeinschaft (DFG). The authors wish to thank Susan Nieschlag, M.A. for language editing.

References


Conventional treatments of male infertility


Vandekerckhove, P.O., Donavan, P.A., Lilford, R. and
Conventional treatments of male infertility


