human
reproduction
update

World Health Organization reference values for human semen characteristics*‡

Trevor G. Cooper[1,10], Elizabeth Noonan[2], Sigrid von Eckardstein[3], Jacques Auger[4], H.W. Gordon Baker[5], Hermann M. Behre[6], Trine B. Haugen[7], Thinus Kruger[8], Christina Wang[9], Michael T. Mbizvo[3,†], and Kirsten M. Vogelsong[3,†]

[1]Centre of Reproductive Medicine and Andrology of the University, Domagkstrasse 11, D-48129 Münster, Germany [2]Fred Hutchinson Cancer Research Center, SCHARP Statistical Center for HIV/AIDS Research and Prevention, Seattle, WA, USA [3]UNDP/UNFPA/WHO/ World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Reproductive Health and Research, WHO, CH-1211 Geneva 27, Switzerland [4]Service d'Histologie-Embryologie, Biologie de la Reproduction/CECOS, Pavillon Cassini, Hôpital Cochin, Paris, France [5]Department of Obstetrics and Gynaecology, University of Melbourne, Royal Women's Hospital, Carlton, VIC, Australia [6]Center for Reproductive Medicine and Andrology, University Hospital Halle, Martin Luther University, Halle, Germany [7]Faculty of Health Sciences, Oslo University College, Oslo, Norway [8]Reproductive Biology Unit, Stellenbosch University, Tygerberg, South Africa [9]Harbor-UCLA Medical Center and Los Angeles Biomedical Research Institute, Torrance, CA, USA

[10]Correspondence address. Tel: +49-251-835-6449; Fax: +49-251-835-6093; E-mail: tgcooper@gmx.de

TABLE OF CONTENTS

BACKGROUND: Semen quality is taken as a surrogate measure of male fecundity in clinical andrology, male fertility, reproductive toxicology, epidemiology and pregnancy risk assessments. Reference intervals for values of semen parameters from a fertile population could provide data from which prognosis of fertility or diagnosis of infertility can be extrapolated.

*Dedicated to the memory of Professor GMH Waites (1928–2005).

†These authors (M.T.M., K.M.V.) are staff members of the World Health Organization. The authors alone are responsible for the views expressed in this publication; these views do not necessarily represent the decisions or policies of the World Health Organization.

‡The list of authors who contributed data to this study is given in the Appendix.

METHODS: Semen samples from over 4500 men in 14 countries on four continents were obtained from retrospective and prospective analyses on fertile men, men of unknown fertility status and men selected as normozoospermic. Men whose partners had a time-to-pregnancy (TTP) of ≤12 months were chosen as individuals to provide reference distributions for semen parameters. Distributions were also generated for a population assumed to represent the general population.

RESULTS: The following one-sided lower reference limits, the fifth centiles (with 95th percent confidence intervals), were generated from men whose partners had TTP ≤ 12 months: semen volume, 1.5 ml (1.4−1.7); total sperm number, 39 million per ejaculate (33−46); sperm concentration, 15 million per ml (12−16); vitality, 58% live (55−63); progressive motility, 32% (31−34); total (progressive + non-progressive) motility, 40% (38−42); morphologically normal forms, 4.0% (3.0−4.0). Semen quality of the reference population was superior to that of the men from the general population and normozoospermic men.

CONCLUSIONS: The data represent sound reference distributions of semen characteristics of fertile men in a number of countries. They provide an appropriate tool in conjunction with clinical data to evaluate a patient's semen quality and prospects for fertility.

Key words: human semen / reference values / infertility diagnosis / fertile men

Introduction

The 'WHO manual for the examination of human semen and sperm[semen]-cervical mucus interaction' (WHO, 1987, 1992, 1999) is widely used as a source of standard methodology for laboratories engaged in semen analyses. However, the interpretation and application of previous WHO 'normal' or 'reference' values for semen parameters used thus far have limitations, since the data were derived from imprecisely defined reference populations and obtained from laboratories with unknown comparability with respect to analytical methodologies. These values were limited by the lack of available data on semen variables in recent fathers, and did not define true reference ranges or limits. There has been no consensus around the suitability of these values, as some centres consider the cited values for characteristics of sperm concentration, morphology and motility too high, whereas others consider them too low.

If too high, a high percentage of fertile men would be classified as subnormal, especially when morphology, sperm concentration or motility is considered (Barratt et al., 1988; Chia et al., 1998; Nallella et al., 2006; Pasqualotto et al., 2006; Gao et al., 2007, 2008). Healthy men may also be investigated for infertility, or inappropriately treated by Assisted Reproduction Technologies, as a result of their lower semen quality if reference limits are too high (Bostofte et al., 1983; Lemcke et al., 1997).

On the other hand, a sperm concentration of 20×10^6/ml, the 'normal' or 'reference' value cited by WHO (1987, 1992, 1999), has been considered too low for a lower reference limit because the probability of pregnancy is essentially linear with sperm concentrations up to $40–50 \times 10^6$/ml (Bonde et al., 1998; Slama et al., 2002). Conversely, sperm concentrations above this value are repeatedly observed in infertile patients (Nallella et al., 2006). There may be no upper limit of any semen characteristics since pregnancy rates increase with superior sperm morphology and motility (Garrett et al., 2003). The then-current normal morphology value of WHO (1987) was considered inadequate by Check et al. (1992) as it did not distinguish between fertile and infertile men whose partners were healthy. With uncertain reference values, over- or under-diagnosis may result. Although much of the investigation conducted to date has considered the WHO 'normal' or 'reference' values as

cut-off limits separating fertile from infertile populations, doubts have been raised about the validity of this approach (Bartoov et al., 1993; Barratt et al., 1995).

This article considers which men are most suitable for providing a reference population, presents data from such a population, mentions the possible limitations of the results obtained and discusses how the reference intervals could be interpreted as useful reference limits. The present analysis benefits from the availability and incorporation of multi-country data from recent fathers with known time-to-pregnancy (TTP). The development and application of clear reference ranges should help reduce the incidence of misdiagnosis of fertility problems and improve clinical care.

Individuals considered suitable for providing reference semen values have included unselected populations, that is, men of unproven fertility (Irvine et al., 1996; Paulsen et al., 1996; Lemcke et al., 1997; Junqing et al. 2002); men from couples presenting with infertility (MacLeod and Wang, 1979; Bostofte et al., 1983; Berling and Wolner-Hanssen, 1997; Andolz et al., 1999); candidates for semen donation, some proven fertile (Leto and Frensilli, 1981; Auger et al., 1995; Bujan et al., 1996; Van Waeleghem et al., 1996) and men presenting for vasectomy (Sultan Sheriff, 1983; Fisch et al., 1996). Whereas the first group may be considered drawn from the general population, semen donors may be, and vasectomy candidates most probably are, of proven fertility, although paternity may not have been recent relative to provision of the semen sample analysed. The majority of men have indefinable fertility status at any one moment: therefore a reference range comprising recently fertile men is defined by men whose semen variables may not reflect those of the general population. This is unusual among clinical laboratory tests and clearly presents a major challenge in defining a valid population reference range for human semen.

The present study examined semen quality in groups of men from the general population (having unknown fertility status) as well as fathers. For the investigation of male factor infertility, the most relevant reference group is that of proven fertile men, since for valid comparisons of patient data with the reference values, the patient should sufficiently resemble the reference individuals in all respects other than those under investigation (PetitClerc and Solberg, 1987; Solberg, 1987), in this case fertility. The selection criteria determining

which individuals are included in the reference population would ideally include proof of paternity, but this is rarely requested or obtained.

Where semen samples are sought from fertile men, approaching the pregnant woman is likely to lead to the identification of the true biological father; but whether he provides a sample may depend on his cultural and social background, as well as his doubts about paternity, biasing the study population unpredictably. Several prospective cross-sectional studies have established baseline values of human semen quality from standardized methodology in relation to fertility (Zinaman et al., 2000; Auger et al., 2001; Jørgensen et al., 2001; Swan, 2003; Eustache et al., 2004; Slama et al., 2004; Haugen et al., 2006; Iwamoto et al., 2006; Pal et al., 2006; Stewart et al., 2009). To avoid the collection bias associated with selecting fertile men, obtaining whole population data has been suggested as ideal. Although theoretically attractive, this is practically unachievable owing to the potentially embarrassing or personal nature of reproductive studies per se (Handelsman et al., 1985), the attitudes of those seeking care (Tielemans et al., 2002) and self-selection of those who are willing to participate (Handelsman, 1997).

The increasing acceptance of WHO standard methodology for semen analysis by laboratories performing clinical studies worldwide means that reference distributions can be generated from a combined analysis of these data. This article presents semen characteristics of, and provides reference intervals and limits generated from, a population of men who had fathered a child within 1 year of trying to induce a pregnancy. The 95% reference intervals for a range of semen variables and the lower (2.5th centile and 5th centile) reference limits, have been generated, in line with clinical chemistry standards. Data from populations of fathers with unknown TTP and men with unknown fertility status are also presented, to indicate that ranges may be different for men with untested fertility examined for other purposes such as male contraception studies, or recruited from the general population. The present analyses were performed on behalf of, and with financial and technical support from, WHO; the data are to be included in the forthcoming fifth edition of the 'WHO laboratory manual for the examination and processing of human semen'.

Materials and Methods

Study populations

Reference values can be subject-based (sequential samples from single individuals) or population-based (single samples from a group of usually healthy donors) (PetitClerc and Solberg, 1987; Solberg, 1987). In this study, data from a population of fertile men were analysed. The men were heterogeneous for definitions of fertility, having a currently or formerly pregnant partner with known TTP up to and including 12 months. This is a population of fertile men from partnerships of high or normal fecundity 'Fathers with TTP ≤ 12 months'. Data from 1953 semen samples from five studies in eight countries on three continents were combined and analysed (Table I for location of laboratories and methods used).

TTP is a well-known and standardized epidemiological index (Joffe, 2000), defined as the number of months (or cycles) from stopping contraception to achieving a pregnancy and was reported in the publications of the original prospective and retrospective studies cited here. The subset of fertile men with TTP ≤ 12 months was selected to provide reference values for human semen, since infertility is currently defined as a failure

to conceive after at least 12 months of unprotected intercourse (Rowe et al., 1993, 2000).

Semen data from three other groups of men were examined for comparison:

(i) 'unscreened' men were men from the general population or young healthy men applying to donate samples for trials of hormonal contraception. This is a mixed population of men of unknown fertility, assumed to be representative of the general population. Data from 965 semen samples from seven studies in five countries on three continents were combined and analysed (Table I).

(ii) 'screened' men were those whose samples satisfied the then-current WHO criteria for normozoospermia. This is a mixed population of men with unknown fertility history, being either volunteers who were screened prior to participation in male contraceptive trials or men attending infertility clinics. Data are presented to reveal any effects of pre-selection of samples and to represent the population that conformed to previous 'normal' or 'reference' values. A total of 934 data points from four studies in four countries on three continents and from two multinational WHO studies (WHO, 1990, 1996; Table I) were combined and analysed.

(iii) 'fertile men with unknown TTP' were those whose partners gave birth prior to the provision of the semen sample, but with no reported TTP. This is a population of fertile men with partnerships of probably all ranges of fecundity: high, normal, moderately or severely impaired. A total of 817 data points from two studies in two countries on two continents and from two multinational WHO studies (WHO, 1990, 1996; Table I) were combined and analysed.

Analytical methods and quality control

For results to be acceptable as reference values, the conditions under which the samples were obtained and processed for analysis should be known and laboratory results should be produced using adequately standardized methods under sufficient quality control (Solberg, 2004). All laboratories generating the data analysed here used standardized methods for semen analysis, i.e. procedures in the edition of the 'WHO manual for the examination of human semen and sperm[semen]-cervical mucus interaction' current at the time of the original studies (WHO, 1987, 1992, 1999). The various editions of the manual provided similar methods for assessing sperm concentration, motility and morphology but provided different criteria for categorising morphology. As the manual provides a choice of methods for measuring semen volume, counting spermatozoa and staining morphology slides, the actual methods used by each laboratory are listed in Table I.

No external quality control (EQC) for semen analysis was available for the early studies included here, but most of the later studies were done by laboratories employing both internal and EQC according to accepted practices. Data that were combined to calculate the reference distributions were provided by laboratories that practiced rigorous internal quality control (IQC) and EQC.

Identification of data and handling of the datasets

As semen analysis is difficult to perform by general clinical laboratories, and formal quality control has only recently been introduced into andrology laboratories (Cooper et al., 1999, 2002), data were obtained from laboratories that were known to provide assessments according to standardized methodologies. A systematic review of the literature was not performed to identify all data on semen quality in various populations. Laboratories and data were identified through the known literature and

Table I Location of and methods used by laboratories providing data for this study

Category[1]	City, Country, Continent[2]	N[3]	Semen Volume[4]	Sperm Concentration[5]	Sperm Motility[6]	Sperm Morphology[7]	Reference
Unscreened	Sydney, Australia, AU	225	S	N	37	Q	Turner et al. (2003)
Unscreened	Melbourne, Australia, AU	41	C	N	RT	S[8]	WHO (1996), McLachlan et al. (2000)
Unscreened	Edinburgh, UK, EU	84	W	N	37	D	Brady et al. (2004, 2006), Hay et al. (2005), Walton et al. (2007)
Unscreened	Manchester, UK, EU	24	P	N	37	P[8]	Unpublished results
Unscreened	Los Angeles, USA, AM	332	P	N	37	P[8]	Gonzalo et al. (2002), Qoubaitary et al. (2006), Wang et al. (2006)
Unscreened	Santiago, Chile, AM	60	P,S	M,N	37	P	von Eckardstein et al. (2003), Unpublished results
Unscreened	Münster, Germany, EU	199	GC	N	37	P	Büchter et al. (1999), Kamischke et al. (2000a, b, 2001a, b, 2002), Unpublished results
Fathers TTP	Melbourne, Australia, AU	206	C	N	RT	S[8]	Stewart et al. (2009)
Fathers TTP	Paris, France; Turku, Finland; Copenhagen, Denmark; Edinburgh, Scotland, EU	900	W	N,B,T,Ma	37	S[8]	Auger et al. (2001), Jørgensen et al. (2001), Slama et al. (2002)
Fathers TTP	Columbia, USA, AM	593	W	N	37	P[8]	Swan et al. (2003)
Fathers TTP	Oslo, Norway, EU	89	W	N	37	P[8]	Haugen et al. (2006)
Fathers TTP	Copenhagen, Denmark, EU	165	C	M,B,T	37	P	Bonde et al. (1998), Jensen et al. (2001)
Fathers	no TTP Davis, USA, AM	606	P	Mi	37	P[8]	Guzick et al. (2001)
Fathers	no TTP Münster, Germany, EU	58	GC	N	37	P	Kamischke et al. (2001a), Unpublished result
scr+no TTP	Stockholm, Sweden, EU	37 + 23	W	N	37	P	WHO (1990, 1996)
scr+no TTP	Szeged, Hungary, EU	11 + 5	GC	M	RT	P	WHO (1996)
scr+no TTP	Singapore, AS	3 + 1	P	M	RT	EN	WHO (1996)
scr+no TTP	Sydney, Australia, AU	61 + 23	S	N	37	Q	WHO (1990, 1996)
scr+no TTP	Melbourne, Australia, AU	45 + 18	W	N	RT	S	WHO (1990, 1996)
scr+no TTP	Turku, Finland, EU	21 + 7	C	B	37	H	WHO (1990)
scr+no TTP	Edinburgh, UK, EU	60 + 15	P	N	37	P	WHO (1990, 1996)
scr+no TTP	Manchester, UK, EU	22 + 5	P	N	37	P	WHO (1996)
scr+no TTP	Bicêtre, France, EU	11 + 4	P	B	37	S	WHO (1990, 1996)
scr+no TTP	Los Angeles, USA, AM	16 + 5	S	N	RT	B	WHO (1996)
scr+no TTP	Beijing, China, AS	56 + 2	P	N	RT	P	WHO (1990, 1996)
scr+no TTP	Nanjing, China, AS	56 + 6	P	N	RT	P	WHO (1990, 1996)

Continued

Table I *Continued*

Category[1]	City, Country, Continent[2]	N[3]	Semen Volume[4]	Sperm Concentration[5]	Sperm Motility[6]	Sperm Morphology[7]	Reference
scr+no TTP	Seattle, USA, AM	41 + 22	P	N	RT	P	WHO (1990, 1996)
scr+no TTP	Chengdu, China, AS	29 + 17	P	N	RT	P	WHO (1990, 1996)
Screened	Melbourne, Australia, AU	84	P	N	37	H[8]	McLachlan *et al.* (2000, 2002)
Screened	Manchester, UK, EU	29	P	N	37	P[8]	Unpublished results
Screened	Bologna, Italy, EU	89	P	N	RT	P[8]	Meriggiola *et al.* (1996, 2003, 2005)
Screened	Beijing, China, AS	263	C	N	RT	P	Gu *et al.* (2003)

[1]TTP, time (≤12 to >12 months) to pregnancy; noTTP, no time to pregnancy recorded; scr, screened; [2]AU, Australasia; AM, Americas; EU, Europe; AS, Asia; [3]Number of samples [where two values are recorded for WHO (1990, 1996) studies, they relate to populations of screened men and fathers, respectively (scr+no TPP)]; [4]GC, collected in graduated cylinder; W, from weight (assuming density 1 g/ml); P, drawn into a pipette from the collection vessel; S, taken into a syringe from the collection vessel; C, transferred to a cylinder from the collection vessel; [5]B, Bürker−Türk chamber; M, Makler chamber; Ma, Malassez chamber; Mi, Microcell chamber (data from these chambers were not used in the analyses); N, Neubauer chamber; T, Thoma chamber; [6]37, 37°C; RT, room temperature; [7]Stains: B, Bryan−Leishman; D, DiffQuik; H, Haematoxylin and Eosin; P, Papanicolaou; Q, Quickdip; S, Shorr; [8]centres providing normal sperm morphology data.

personal communication with investigators and the editorial group of the fifth edition of the 'WHO laboratory manual for the examination and processing of human semen' (forthcoming). The data representing the reference population of fertile men were derived from all known and identified prospective or retrospective studies designed with time to pregnancy as an outcome and in which WHO-recommended methodologies for semen analysis were employed. Data that were inadvertently omitted may be of similar quality. Although acknowledging that there were differences in results among centres, it was not possible to attribute this variability to the known methodological differences, apart from morphology, or possible geographical differences of the study populations, or the different size of the datasets. Data on ethnicity were not always available.

Datasets were provided by the centres responsible for generating them or by WHO (Table I). The analysis was designed as an integrated analysis combining primary data from various primary studies, which meets the definition of an individual patient data meta-analysis. The data from individual semen donors were compiled and analysed. Where data have been published, the relevant publications are marked with an asterisk in the reference list. All data were supplied as Excel spreadsheets and hand-checked for missing values and typographical errors before statistical analysis. Data on semen volume, sperm concentration, total sperm number per ejaculate, motility, vitality and normal morphology were included only if they were generated from complete semen samples, obtained following 2–7 days of sexual abstinence. This range was used because this is the interval recommended by WHO and it has thus become a standard practice. The relationship between abstinence time and semen analysis results within this time frame is well-known.

Semen analysis results from only one sample per man (as recommended by PetitClerc and Solberg, 1987; Solberg, 1987), the first where several were given, were included in the analyses, so as not to over-represent certain men by averaging values. As a result, the variation observed is likely to reflect inter-individual variation. Sperm concentration was measured by haemocytometer (mainly improved Neubauer, but some laboratories used Bürker−Türk or Malassez) on diluted and fixed samples. Only total motility (WHO grades a + b + c) and progressive motility (WHO grades a + b combined) were included, for more accuracy and consistency in results (Cooper and Yeung, 2006).

Although all centres reported using WHO procedures, the recommended methodologies have changed over time, and many centres have experienced difficulties with the subjective assessments of morphology. Data on normal sperm morphology were only included if results were reported as determined according to the so-called 'strict' (Tygerberg) method (WHO, 1992, 1999). Data from four studies (fathers TTP indicated in Table I) were combined to provide the reference distribution for fertile men. To obviate among-centre differences, morphology slides for the two multicentre studies (Auger *et al.*, 2001; Swan *et al.*, 2003 and ongoing studies) were read centrally. The other two studies were single-centre studies (Haugen *et al.*, 2006; Stewart *et al.*, 2009). All four studies involved EQC for sperm morphology.

Sperm vitality data, assessed by the eosin−nigrosin method in semen from partners of pregnant women with TTP ≤ 12 months was obtained from two centres, in France and Australia, but was not included as an endpoint in any studies of 'unscreened' men reflecting the general population.

Statistical analysis

Different paradigms used by statistical packages are known to influence the reference limits of human semen (Egeland and Haugen, 2007). In a preliminary analysis, SAS (SAS Institute, San Francisco, CA USA) was used to generate and compare the variance-weighted 2.5th and 5th centiles. For statistical comparison of lower reference limits, the values were weighted by letting b_i be the fifth centile estimate from study 1 and letting v_i be the estimated variance of b_i. The pooled estimate of the fifth centile (Poolb) is $(b_1/v_1 + b_2/v_2 + \cdots)/(1/v_1 + v_2 \cdots) = w_1b_1 + w_2b_1 + \cdots$ where $w_i = (1/v_i)/(1/v_1 + 1/v_2 + \cdots)$. The estimate of the standard error (Poolb) is $w_1^2v_1 + w_2^2v_2 + \ldots$. The 95% confidence intervals were calculated using the normal approximation, weighted 5th centile ± (1.96 × pooled standard error). The fifth centile and variance of each study were obtained by quantile regression. As the weighted values were not fundamentally different from those obtained in a non-weighted analysis, the final analyses were performed on non-weighted raw data. Preliminary analyses did not suggest significant differences among centres, except in the case of morphology assessments where variability among centres could be attributed to differences in methodologies.

In this study, Stata Version 9 (Statacorp, College Station, TX, USA) was used to generate the centiles of the raw data, and Sigma Stat (v3.5, SysStat Software GmbH, Erkrath, Germany) was used to compare the datasets from all the groups. As no transformation method produced Gaussian distributions of the data, non-parametric tests were used. Wilcoxon's Rank Sum Test was performed with multiple comparisons against the reference population (fathers TTP ≤ 12 months); in Dunn's *post hoc* test, $P < 0.05$ was considered significant. For the fathers in partnerships with TPP ≤ 12 months, both the central 95% and one-sided 95% of the population-based distribution are presented as potential reference intervals, i.e. the 2.5th and 5.0th centiles are both provided as possible lower reference limits. The 95% confidence intervals of both lower reference limits are presented. Graphical presentation of primary data in the form of box-and-whisker plots and histograms of the distributions of values are provided (SigmaPlot, Version 10.0, SysStat Software).

Results

Ages of men providing semen samples

The age range of all of the men who provided samples was 17–67 years, which covers the usual ages of men being investigated for infertility or requiring contraception. The fathers with partners with TTP ≤ 12 months had a mean ($\pm$ SD) age of 31 $\pm$ 5 years (range 18–53) and only 10 men were over age 45 whereas the 'fertile men of unknown TTP' were of age 33 $\pm$ 5 years (20–52) and 12 were over 45. The 'unscreened' men were of age 33 $\pm$ 7.8 years (17–67) and 54 men were over 45; the 'screened' men (age 32 $\pm$ 6 years; 19–50) included three men over 45 years. The data may not be representative of the normal distributions in younger or older men.

Reference values for human semen

Men whose partners had a TTP ≤ 12 months were chosen as the reference group from which the reference values for human semen from fertile men were determined. The distribution of data for various semen characteristics in this population is given in Table II. For a conventional two-sided distribution, the 2.5th centile, which constitutes the lower reference limit in most clinical laboratory tests, could be proposed for the lower limit of semen characteristics; for a one-sided distribution, the fifth centile is the lower reference limit. Both of these lower reference limits, and their 95% confidence intervals, are given in Table II. Table III presents the same data for men of unknown fertility from the general population, for comparison. All parameters were routinely measured according to standard methodologies, with the exception of total sperm number per ejaculate, which is derived, for each individual semen sample, by multiplying the sperm concentration by the volume of the whole ejaculate. This relationship does not hold for the population-based centiles, as the parameters of sperm concentration and semen volume are not correlated in the population.

Statistical differences in semen characteristics among the various populations

Semen from fathers with partners' TTP ≤ 12 months had significantly greater semen volume, sperm concentration and percentage of normal forms than those of the other three groups, although the percentages of all total motile and progressively motile spermatozoa were significantly lower in this group than in some others (Fig. 1). Frequency histograms of semen volume, sperm concentration, total sperm numbers per ejaculate, percentages of total and progressively motile and of morphologically normal spermatozoa are presented in Fig. 2 for the four populations. Despite the lower percentage of progressively motile spermatozoa in fathers with TTP ≤ 12 months, the larger semen volumes and total sperm numbers provided this population with higher total numbers of progressively motile spermatozoa than found in the screened and unscreened groups; the total number of morphologically normal spermatozoa was also greater in the reference population than in the other fathers and unscreened groups (Table IV).

Discussion

Semen analysis is usually performed to help determine why a couple is having difficulty conceiving, to follow the course of a treatment affecting testicular or accessory gland function, following vasectomy or in a research context. Reference values for the composition for semen, akin to those provided in clinical chemistry for blood values, would be helpful in both clinical and research settings. The current work presents an assembly of human semen variables from the most plausible reference group (fathers in partnerships with TTP ≤ 12 months) to form normative human population data, obtained from laboratories using standardized and controlled methods in eight countries on three continents. The conventional statistically accepted 95% reference interval, and 2.5th and 5th centile lower reference limits from two- and one-sided distributions, respectively, were calculated from over 1900 semen samples for semen volume, sperm concentration, total sperm numbers per ejaculate, sperm motility and sperm morphology; fewer samples were analysed for vitality. In this study, the standardized methods used will have minimized analytical error, so the large range in values observed in each group of men likely reflects biological variation (Castilla *et al.*, 2006). Despite the methodological differences over time and among centres (see below), the values presented here are considered to represent global semen characteristics of fertile men.

The men in the reference population are characterized by not only larger semen volumes and higher concentrations and numbers of spermatozoa in their ejaculate, but also by a higher total number of motile and morphologically normal cells per ejaculate than found in the other groups.

Choice of reference limits

Clinical reference values are required for comparison with values obtained from the patient being assessed, among other reasons. The observed values are used to make a clinical decision by comparing them with reference distributions and reference intervals (PetitClerc and Solberg, 1987), in addition to a number of other bioclinical aspects of both partners. Descriptive statistics of semen from fertile and infertile men have included the mean and standard deviation, although these are not appropriate for percentages or for concentrations, where transformation of the data is necessary before statistical analysis can be performed (Berman *et al.*, 1996; Handelsman, 2002). Non-parametric descriptions of semen analysis data have

Table II Distribution of values, lower reference limits and their 95% CI for semen parameters from fertile men whose partners had a time-to-pregnancy of 12 months or less

	N	Centiles										
		2.5	(95% CI)	5	(95% CI)	10	25	50	75	90	95	97.5
Semen volume (ml)	1941	1.2	(1.0–1.3)	1.5	(1.4–1.7)	2	2.7	3.7	4.8	6	6.8	7.6
Sperm concentration (10^6/ml)	1859	9	(8–11)	15	(12–16)	22	41	73	116	169	213	259
Total number (10^6/Ejaculate)	1859	23	(18–29)	39	(33–46)	69	142	255	422	647	802	928
Total motility (PR + NP, %)*	1781	34	(33–37)	40	(38–42)	45	53	61	69	75	78	81
Progressive motility (PR, %)*	1780	28	(25–29)	32	(31–34)	39	47	55	62	69	72	75
Normal forms (%)	1851	3	(2.0–3.0)	4	(3.0–4.0)	5.5	9	15	24.5	36	44	48
Vitality (%)	428	53	(48–56)	58	(55–63)	64	72	79	84	88	91	92

*PR, progressive motility (WHO, 1999 grades a + b); NP, non-progressive motility (WHO, 1999 grade c).
The values are from unweighted raw data. For a two-sided distribution the 2.5th and 97.5th centiles provide the reference limits; for a one-sided distribution the fifth centile provides the lower reference limit.

included the median alone (MacLeod, 1950, 1951; MacLeod and Gold, 1951a, b; Page and Houlding, 1951; MacLeod and Wang, 1979; Wang et al., 1985), together with the interquartile ranges (MacLeod and Gold, 1951a, b; Chia et al., 1998; Nallella et al., 2006; Pal et al., 2006; Pasqualotto et al., 2006), none intended to be a reference limit, and the 15th centile, suggested as a reference limit in one case because 15% of men in the population were infertile (Junqing et al., 2002).

Where lower reference limits for semen variables from fertile men have been proposed, there is surprisingly no agreement on which value to take, and proposals have included the 10th centile (Rehan et al., 1975; Jouannet et al., 1981; Menkveld et al., 2001; van der Merwe et al., 2005), the 5th centile (MacLeod, 1951, Sultan Sheriff, 1983; Barratt et al., 1988; Jørgensen et al., 2001; Andersen et al., 2002; Slama et al., 2002; Gao et al., 2007, 2008) and the 2.5th centile (Cooper et al., 1991). In their studies, Rehan et al. (1975) reported both the 16th and 10th centiles, Ombelet et al. (1997) the 10th and 5th centiles and Haugen et al. (2006) the 10th, 5th and 2.5th centiles. Other approaches taken to provide cut-off values distinguishing fertile from infertile men are classification and regression (Guzick et al., 2001) and receiver operating characteristics (ROC)

curves (Ombelet et al., 1997; Gunalp et al., 2001; Menkveld et al., 2001; Nallella et al., 2006).

In setting reference limits for clinical chemistry it is widely accepted that 95% of the data should be included in the reference interval. For a two-sided distribution, the 2.5th and 97.5th centiles of a reference distribution should form the lower and upper reference limits, respectively (Dybkaer and Solberg, 1987; PetitClerc and Solberg, 1987; Horn and Pesce, 2003; Solberg, 2004, 2006). However, the justification for setting the reference limits on the composition of blood is not necessarily relevant to setting limits for semen. Most blood components are well-regulated to prevent too high or too low concentrations reaching target tissues, so upper and lower reference limits are necessary. In contrast, the composition of semen is not controlled by strict feedback systems and is confounded by a variety of factors including accessory gland emptying and previous sexual activity.
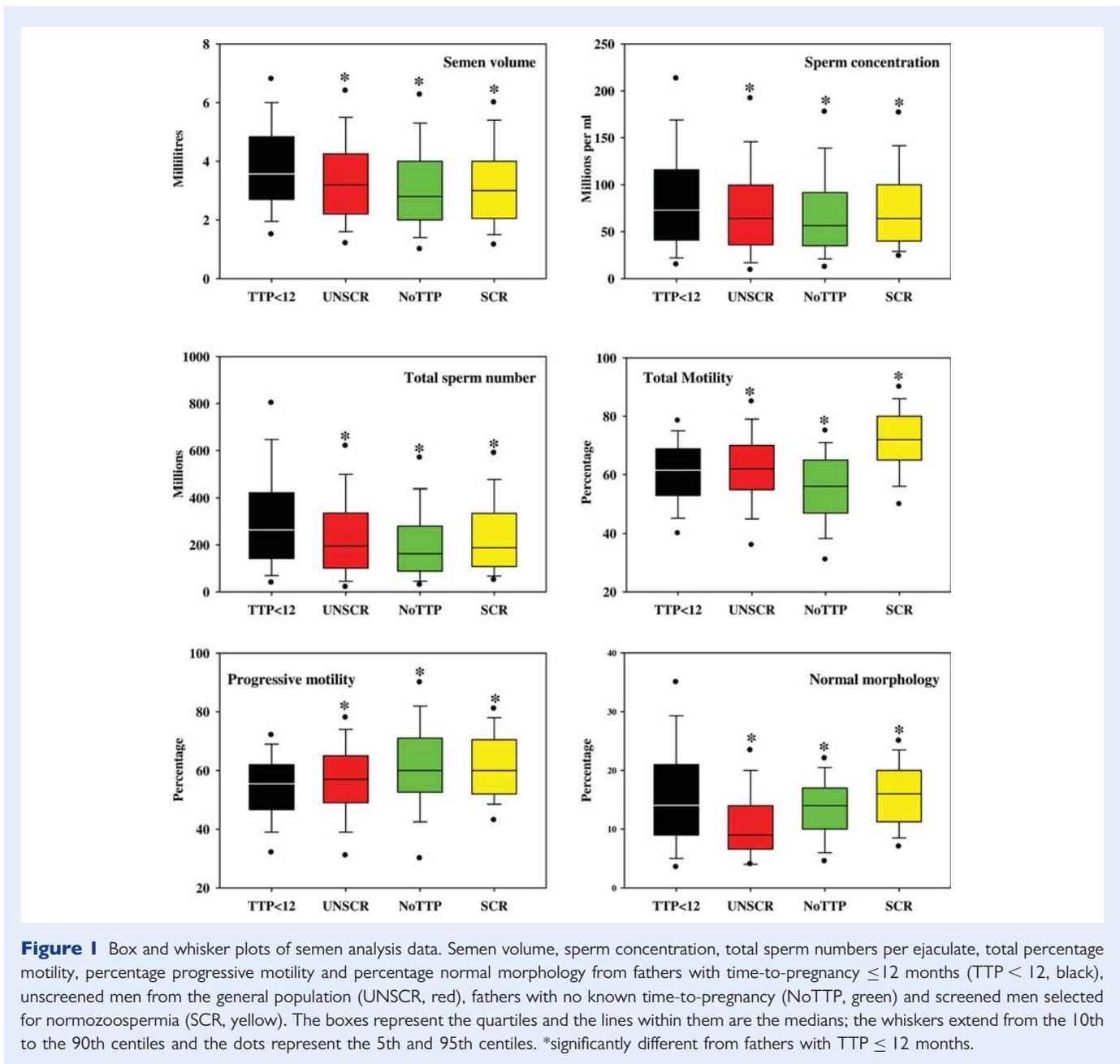
One-sided distributions are deemed appropriate when one side of the distribution is clinically irrelevant (Horn and Pesce, 2003; Solberg, 2006). An analogous situation to that of semen analysis may be that of urinary secretion of metabolites of styrene, hydroxypyrene or α-naphthol. To determine the upper limits of excretion, 95% one-sided

Table III Distribution of values, lower reference limits and their 95% CI for semen parameters from the general population of unscreened men

	N	Centiles										
		2.5	(95% CI)	5	(95% CI)	10	25	50	75	90	95	97.5
Semen volume (ml)	929	0.8	(0.7–1.0)	1.2	(1.0–1.3)	1.6	2.2	3.2	4.2	5.5	6.4	7
Sperm concentration (10^6/ml)	930	4	(1–6)	9	(6–11)	17	36	64	100	192	192	237
Total number (10^6/Ejaculate)	928	11	(3–14)	20	(14–29)	45	101	196	336	619	619	772
Total motility (PR + NP, %)*	928	26	(14–32)	36	(32–39)	45	55	62	70	85	85	88
Progressive motility (PR, %)*	708	20	(7–27)	31	(26–34)	39	49	57	65	78	78	81
Normal forms (%)	137	3.5	(2.0–4.5)	4.7	(3.8–5.5)	7	10.5	14	16	23.2	23.2	30

*PR, progressive motility (WHO, 1999 grades a + b); NP, non-progressive motility (WHO, 1999 grade c).
The values are from unweighted raw data. For a two-sided distribution the 2.5th and 97.5th centiles provide the reference limits; for a one-sided distribution the fifth centile provides the lower reference limit.

Figure 1 Box and whisker plots of semen analysis data. Semen volume, sperm concentration, total sperm numbers per ejaculate, total percentage motility, percentage progressive motility and percentage normal morphology from fathers with time-to-pregnancy ≤12 months (TTP < 12, black), unscreened men from the general population (UNSCR, red), fathers with no known time-to-pregnancy (NoTTP, green) and screened men selected for normozoospermia (SCR, yellow). The boxes represent the quartiles and the lines within them are the medians; the whiskers extend from the 10th to the 90th centiles and the dots represent the 5th and 95th centiles. *significantly different from fathers with TTP ≤ 12 months.

upper reference limits are computed (Murer *et al.*, 1994; Hansen *et al.*, 1995), since lower limits are irrelevant. One-sided limits are used for neonatal serum thyroid stimulating hormone levels, where action is taken only if values are too high (Koduah *et al.*, 2004).

Thus one-sided lower reference limits may be appropriate for the various semen parameters described here, since 'too high' values appear to be clinically irrelevant. Despite older reports that polyzoospermia (sperm concentration $>250 \times 10^6$/ml) is associated with subfertility and increased spontaneous miscarriage rates, the nature of the defect is unclear. Sperm penetration through cervical mucus (Glezerman *et al.*, 1982) and fusion with zona-free hamster oocytes (Chan *et al.*, 1986) are normal, although a lower sperm acrosin content (Schill and Feifel, 1984) and lower acrosome reaction rates than in controls (Töpfer-Petersen *et al.*, 1987) suggest defective

acrosomal function. One report indicates that there is no reason to believe that high sperm numbers or percentages of progressively motile or morphologically normal spermatozoa are harmful to fertility (Tournaye *et al.*, 1997).

Comparison of the current with published reference limits

This analysis represents a sound empirical estimation of lower reference limits, together with their confidence intervals, which prove much wider than previously assumed. The combined data come from various regions in the world where ethnic or other factors may differ and influence the distribution. Despite the use of an 'elite' population, the data nevertheless provide an appropriate and
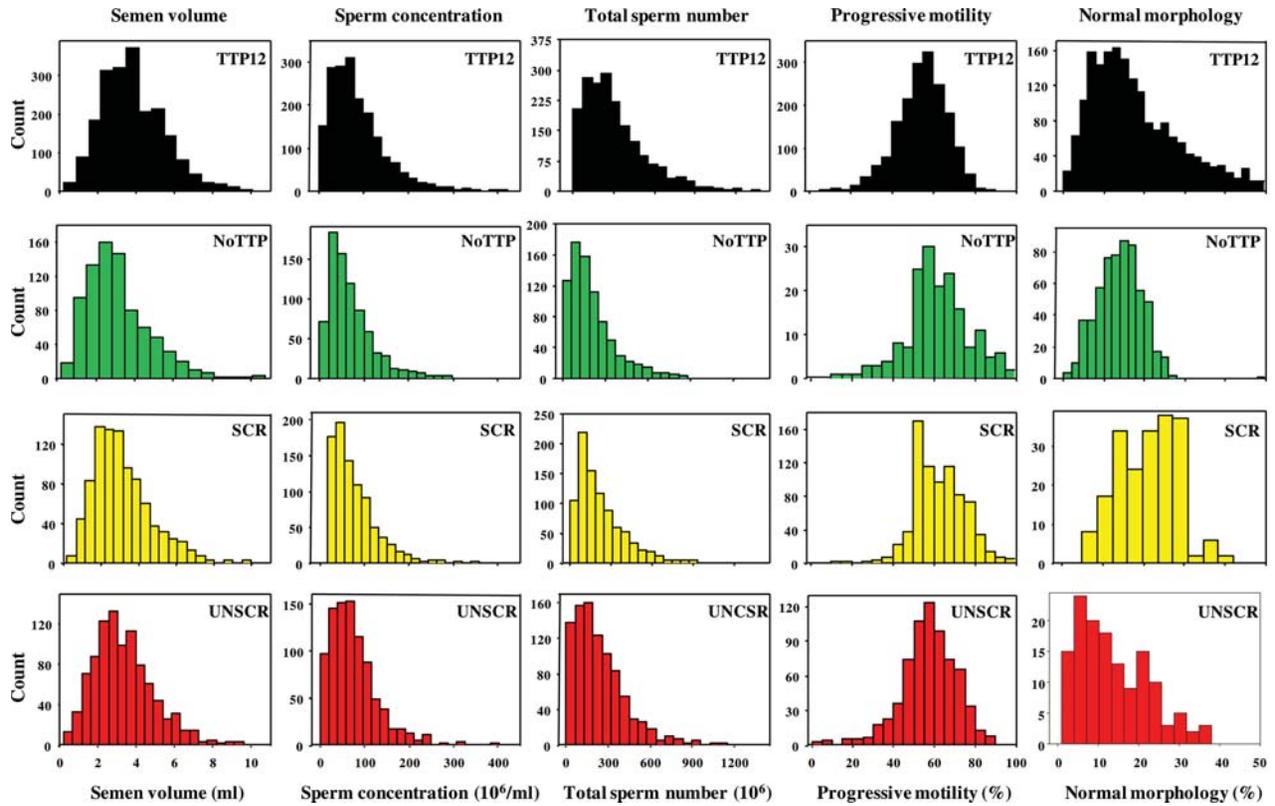
Figure 2 Frequency histograms of semen analysis data from fathers, the general population and men screened for normozoospermia. Distribution of semen volumes (ml, First Column), sperm concentration (10⁶/ml, Second Column), total sperm numbers (10⁶, Third Column), progressively motile spermatozoa (%, Fourth Column) and morphologically normal spermatozoa (%, Fifth Column) in ejaculates from fathers with time to pregnancy 12 months or less (TTP12, Top Row, black), fathers with no known time to pregnancy (NoTTP, Second Row, green), men screened for normozoospermia (SCR, Third Row, yellow) and unscreened men from the general population (UNSCR, Bottom Row, red).

relevant reference interval, with lower limits being suitable for use in conjunction with clinical data to evaluate a patient's semen quality and prospects for fertility. Apart from total sperm number per ejaculate, the lower limits of these distributions are lower than the previously presented 'normal' or 'reference' values (WHO, 1987, 1992, 1999).

One of the earliest published assessments of sperm concentration in human semen was by Macomber and Sanders (1929) who reported a median of ∼100 million spermatozoa per millilitre, using blood pipettes and an unidentified counting chamber. Systematic studies were started with the examination of semen from men whose partners were currently pregnant (MacLeod, 1950, 1951; MacLeod and Gold, 1951a) and an interesting discrepancy between results of different centres that surfaced since then has been reviewed by Zukerman et al. (1977) and MacLeod and Wang (1979), especially concerning what should be taken as discriminating values for fertility. The generally accepted values of 20×10^6/ml for sperm concentration and 40×10^6 spermatozoa per ejaculate, used as 'normal'

Table IV Total numbers of all, progressively motile and morphologically normal spermatozoa per ejaculate from fathers and men from screened and unscreened populations

Group	Median (and interquartile range) of the number of spermatozoa (10⁶) per ejaculate		
	Total	Progressively motile	Morphologically normal
Fathers TTP ≤ 12 months	255 (142–422)	145 (76–242)	37 (15–72)
Fathers with no TTP	162 (87–277)*	140 (69–274)	26 (11–58)*
Unscreened (general) population	196 (101–336)*	113 (54–201)*	29 (14–46)*
Screened for normozoospermia	180 (104–315)*	107 (59–205)*	38 (20–60)

*Within columns, significantly different from TTP ≤ 12 ($P < 0.05$).

or 'reference' in WHO's manuals for semen analysis, appear to stem from MacLeod's early work (MacLeod, 1950, 1951; MacLeod and Gold, 1951a, b), where there is much discussion of the fertility of men with less than 20×10^6 spermatozoa per millilitre. This value is close to the fifth centile, judging from the text and graphs (MacLeod and Gold, 1951a, b; MacLeod, 1951). The values for the fifth centile determined in the present analysis are close to these historical values, except for normal morphology where another classification system was used.

The fifth centile for semen volume from fertile men reported here is similar to the lower reference limits reported for fertile men in Norway (fifth centile: Haugen *et al.*, 2006) and Germany (2.5th centile: Cooper *et al.*, 1991). The lower reference limit (fifth centile) for total sperm number per ejaculate is in agreement with those of MacLeod and Gold (1951a) and close to that determined by Cooper *et al.* (1991: 2.5th centile), but lower than in reports from Ombelet *et al.* (1997) and Haugen *et al.* (2006), both using fifth centiles. The limit for sperm concentration lies between those of MacLeod and Gold (1951a) and Menkveld *et al.* (2001) using 5th and 10th centiles as cut-offs and Haugen *et al.* (2006) and Guzick *et al.* (2001) using the fifth centile and classification-and-regression, respectively; both Ombelet *et al.* (1997: 5th centile) and Gunalp *et al.* (2001: ROC) reported lower sperm concentrations as lower reference limits. The reference limit (fifth centile) reported here for progressive motility is in line with reports from Cooper *et al.* (1991), Gunalp *et al.* (2001) and Haugen *et al.* (2006).

Morphology data seemed to be centre-dependent, and highly dependent on the method used to determine the percentage of normal forms, indicating that these differences are procedural and demanding that the data selected for analysis should be limited to those centres adhering to strict guidelines on categorisation (WHO, 1999). Similar lower reference limits for normal sperm morphology were presented by all authors using the same strict application of criteria (Ombelet *et al.*, 1997; Guzick *et al.*, 2001; Gunalp *et al.*, 2001; Menkveld *et al.*, 2001). The low proportions of normal spermatozoa, as defined by those selected in endocervical mucus, will inevitably produce very low reference limits for a fertile population. Indeed, such was found in the present analysis, with 3 and 4% normal forms as the 2.5th and 5th centiles, respectively. With this method, similar low values of 3–5% normal forms have been found by ROC analysis (Pater, 2005) to be optimal cut-off values between fertile and infertile men whose spermatozoa were used for *in vitro* fertilization (Coetzee *et al.*, 1998), intrauterine insemination (Van Waart *et al.*, 2001) and in spontaneous pregnancies (van der Merwe *et al.*, 2005).

Comparisons of semen characteristics among different populations of men

In this study, data from semen analyses obtained from fertile men in partnerships with TTP of 12 months or less (the reference population) were compared with data from fathers with unknown TTP, men of unknown fertility status and men whose semen characteristics conformed to previous WHO reference values. Semen quality from the reference population was superior to that of the other groups used for comparison, as judged from the primary data of semen volume, sperm concentration and percentage of normal forms. The percentage of progressively motile spermatozoa was lower than that in all other groups; however, the

greater total sperm number in this group ensured that the total number of progressively motile spermatozoa was higher in the reference population than in the unscreened and screened populations, and the total number of morphologically normal forms was higher than that in the unscreened men and fathers with no known TTP. A high number of motile human spermatozoa is known to increase their entry into cervical mucus *in vitro* (Katz *et al.*, 1980).

The other group of fathers, in partnerships with unknown time to pregnancy, had significantly lower semen volumes, sperm concentrations and percentages of motile and normally formed spermatozoa, but higher percentages of progressively motile spermatozoa, than the reference population. The derived values of total numbers of all and normal spermatozoa were lower than, whereas the total numbers of progressively motile spermatozoa did not differ from, those from the reference population.

Semen from unscreened men, assumed to represent the general population and originally considered as a possible reference group (see Introduction), had significantly lower semen volume, sperm concentration and percentage of normal forms but a higher percentage of motile spermatozoa than the fathers with TTP ≤ 12 months. However, total numbers of all, progressively motile and morphologically normal spermatozoa per ejaculate were lower than those from the TTP ≤ 12 months fathers. This is consistent with the anticipated inclusion in this population of men with mixed and poor semen quality, infertile men as well as fathers. Choice of these men as the reference population would have provided mainly lower values for the lower reference limits than those obtained from the fathers in couples with TTP ≤ 12 months.

Although obtained from men selected according to previous WHO criteria to be normozoospermic, semen from the screened population also displayed significantly lower semen volumes and sperm concentrations but higher percentages of motile and normal forms than the reference population. Total and motile sperm numbers per ejaculate were lower, but total numbers of normal forms were not different from those of the reference population.

Significance of lower reference limits

Previous semen reference values were presumed to reflect an end-point for the diagnosis, or at least for the further investigation, of male infertility. However, such an end-point is uncertain for several reasons. In particular, the condition diagnosed is not strictly male infertility but rather the possible or probable contribution of one or more semen variables to a multi-factorial condition or disease, namely, a couple's inability to conceive within a given time period. Thus, male fertility only partially contributes to the outcome of interest, together with that of female fecundity (te Velde *et al.*, 2000). The prognostic value of semen components such as sperm number, motility and morphology, as surrogate markers of male fertility, is also confounded in several ways; the fertility potential of a man is influenced by sexual activity, the function of accessory sex glands and other, defined as well as yet unrecognized, conditions and routine semen analysis itself has it own limitations, and does not account for putative sperm dysfunctions such as immature chromatin or a fragmented DNA.

Interpretation of the reference ranges requires an understanding that they provide a description of the semen characteristics of recent fathers. The reference limits should not be over-interpreted

to distinguish fertile from infertile men accurately, but they do represent semen characteristics associated with a couple's achieving pregnancy within 12 months of unprotected sexual intercourse. The reference limits provided here are derived from semen samples from men whose partners conceived spontaneously; as such, the limits provide only a standardized guide regarding a man's fertility status. As fathers constitute a select group of individuals, they may differ in semen values from other normal healthy men. Semen characteristics are highly variable within and among men and these parameters are not the sole determinants of a couple's fertility. Semen parameters within the 95% reference interval do not guarantee fertility nor do values outside those limits (in isolation from other clinical data) necessarily indicate male infertility or pathology. Indeed (by definition) 5% of the fertile men providing the reference data have values outside the 95% reference interval. A man's semen characteristics need to be interpreted in conjunction with his clinical information. The reference limits provided here are from semen samples initiating natural conceptions and as such indicate whether a man may need infertility treatment, but they should not be used to determine the nature of that treatment.

Limitations of the current reference values

The data included in the present analysis were obtained from laboratories using WHO methods for various studies of apparently fertile men and volunteers from the general population. It is difficult to get men to volunteer for reproductive studies that involve semen analysis and the selection biases involved are well recognized. Generally, the acceptance rates following requests to donate semen are low, in the range of 13–19% (Bonde et al., 1998; Andersen et al., 2000; Jouannet et al., 2001; Jensen et al., 2002; Swan et al., 2003; Eustache et al., 2004; Muller et al., 2004). Such low rates may invalidate attempts to extrapolate data to the general population, as the majority of men are not represented by the groups volunteering to provide reference semen samples. The data may be made more representative by permitting samples to be provided at home where donation rates are higher, at 32–54% (Larsen et al., 1998; Hjollund et al., 2000; Jørgensen et al., 2001; Andersen et al., 2002; Cohn et al., 2002), but at the expense of introducing more variables before semen analysis begins, such as the handling and temperature of the sample during transit to the laboratory and the increased time before analysis. The extent of this bias may be large (Handelsman, 1997) but is contested (Eustache et al., 2004; Muller et al., 2004; Hauser et al., 2005).

Whether or not differences exist between the semen quality of men who are willing to provide semen samples and those who are not, can be addressed indirectly by studying semen characteristics from initial responders to advertisements and those subsequently recruited. These comparisons indicate significant differences in semen quality between initial and later responders (Cohn et al., 2002). There may be a greater incidence of previous unfavourable pregnancy outcomes in the partners of volunteers compared with non-volunteers, as shown in a French study (Muller et al., 2004). On the other hand, the comparability of semen characteristics of study and non-study subjects recruited from infertility clinics (Hauser et al., 2005), of serum testosterone between donors and non-donors (Andersen et al., 2000) and of characteristics of the pregnancies between semen donors and non-donors argue against there being major differences between the

populations of men who provide semen samples for research and those who do not.

The studies included in the present analysis were conducted in different regions of the world with some areas over-represented, such as Northern Europe, and others, such as Africa, parts of Europe and Central and South America, under-represented. There were some differences between the results of the different studies but the origin of these differences is unclear. It is possible that they represent real biological differences among men in different regions, or that they are laboratory-dependent biases of measurement, despite their adherence to the WHO manual methods. The studies were conducted over many years, during which time the WHO standardized methods changed for assessing sperm motility and morphology and for performing quality control. The earlier studies were performed without formal quality assurance activities whereas the later studies were conducted with internal and EQC, and whereas the laboratories reputedly performed well, not all laboratories reported QC data for analysis and adjustment of the results.

Assumptions were made that a single semen sample can be taken to represent each man and that the first of multiple ejaculates is representative. The present analysis may be limited in precision by the inclusion of samples obtained after an abstinence period of 2–7 days. This range is allowed because of the practical difficulties in obtaining semen samples following a prescribed period of abstinence. To define reference intervals specific to more precise periods of abstinence may be desirable, but would require a much larger sample size. In a healthy man, the total number of spermatozoa emitted in an ejaculate will depend not only on the time of abstinence, but also on the volume of his testes, the size of his epididymal sperm reserve and the extent of ductal patency.

Further studies will be required to confirm the validity of global reference ranges. Prospective studies will need to be designed to avoid possible among-laboratory variations in methodology and might include centralized assessment of sperm concentration on preserved samples (Jonckheere et al., 2005), video recordings for sperm motility and morphology or automated computer assisted semen analysis using the same standardized equipment. If regional differences are revealed, their mechanism and significance for fertility will need to be studied before it can be decided whether there should be specific reference values for different ethnic groups or regions. It may be that laboratories have to produce their own local reference ranges for semen parameters. A future, confirmatory, analysis would include a systematic review of laboratories using more highly standardized techniques (such as those presented in the forthcoming fifth edition of the 'WHO laboratory manual for the examination and processing of human semen') and reporting successful participation in external and IQC programmes, and would take geographical and ethnic origins into account. It will be of interest to determine the success of various clinical management protocols that incorporate the reference limits into research and practice guidelines.

Authors' Role

T.G.C. initiated and designed the study, conducted the data collection, participated in the data analysis and interpretation, wrote the article and prepared the tables and figures. E.N. performed the statistical analyses and contributed to the drafting and editing of the article. S.E.

References

* Indicates a publication containing data included in the present analysis.

Andersen AG, Jørgensen N, Andersson AM, Carlsen E, Skakkebaek NE, Jensen TK, Keiding N, Swan SH. Serum levels of testosterone do not provide evidence of selection bias in studies of male reproductive health. *Epidemiology* 2000;**11**:232–234.

Andersen AG, Ziebe S, Jørgensen N, Petersen JH, Skakkebaek NE, Andersen AN. Time to pregnancy in relation to semen quality assessed by CASA before and after sperm separation. *Hum Reprod* 2002;**17**:173–177.

Andolz P, Bielsa MA, Vila J. Evolution of semen quality in North-eastern Spain: a study in 22,759 infertile men over a 36 year period. *Hum Reprod* 1999;**14**:731–735.

Auger J, Kunstmann JM, Czyglik F, Jouannet P. Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med* 1995;**332**:281–285.

Auger J, Eustache F, Andersen AG, Irvine DS, Jørgensen N, Skakkebaek NE, Suominen J, Toppari J, Vierula M, Jouannet P. Sperm morphological defects related to environment, lifestyle and medical history of 1001 male partners of pregnant women from four European cities. *Hum Reprod* 2001;**16**:2710–2717.*

Barratt CL, Dunphy BC, Thomas EJ, Cooke ID. Semen characteristics of 49 fertile males. *Andrologia* 1988;**20**:264–269.

Barratt CL, Naeeni M, Clements S, Cooke ID. Clinical value of sperm morphology for in-vivo fertility: comparison between World Health Organization criteria of 1987 and 1992. *Hum Reprod* 1995;**10**:587–593.

Bartoov B, Eltes F, Pansky M, Lederman H, Caspi E, Soffer Y. Estimating fertility potential via semen analysis data. *Hum Reprod* 1993;**8**:65–70.

Berling S, Wolner-Hanssen P. No evidence of deteriorating semen quality among men in infertile relationships during the last decade: a study of males from Southern Sweden. *Hum Reprod* 1997;**12**:1002–1005.

Berman NG, Wang C, Paulsen CA. Methodological issues in the analysis of human sperm concentration data. *J Androl* 1996;**17**:68–73.

Bonde JP, Ernst E, Jensen TK, Hjollund NH, Kolstad H, Henriksen TB, Scheike T, Giwercman A, Olsen J, Skakkebaek NE. Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *Lancet* 1998;**352**:1172–1177.*

Bostofte E, Serup J, Rebbe H. Has the fertility of Danish men declined through the years in terms of semen quality? A comparison of semen qualities between 1952 and 1972. *Int J Fertil* 1983;**28**:91–95.

Brady BM, Walton M, Hollow N, Kicman AT, Baird DT, Anderson RA. Depot testosterone with etonogestrel implants result in induction of azoospermia in all men for long-term contraception. *Hum Reprod* 2004;**11**:2658–2667.*

Brady BM, Amory J, Perheentupa A, Zitzmann M, Hay C, Apter D, Anderson R, Bremner W, Huhtaniemi I, Nieschlag E *et al*. A multi-centre study investigating subcutaneous etonogestrel implants with injectable testosterone decanoate as a potential long-acting male contraceptive. *Hum Reprod* 2006;**21**:285–294.*

Büchter D, von Eckardstein S, von Eckardstein A, Kamischke A, Simoni M, Behre HM, Nieschlag E. Clinical trial of transdermal testosterone and oral levonorgestrel for male contraception. *J Clin Endocrinol Metab* 1999;**84**:1244–1249.*

Bujan L, Mansat A, Pontonnier F, Mieusset R. Time series analysis of sperm concentration in fertile men in Toulouse, France between 1977 and 1992. *BMJ* 1996;**312**:471–472.

Castilla JA, Alvarez C, Aguilar J, González-Varea C, Gonzalvo MC, Martínez L. Influence of analytical and biological variation on the clinical interpretation of seminal parameters. *Hum Reprod* 2006;**21**:847–51.

Chan SYW, Tang LCH, Tang GWK, Ho PC, Wang C. Spermatozoal fertilizing capacity in polyzoospermia: a preliminary study. *Andrologia* 1986;**18**:208–213.

Check JH, Bollendorf A, Press M, Blue T. Standard sperm morphology as a predictor of male fertility potential. *Arch Androl* 1992;**28**:39–41.

Chia SE, Tay SK, Lim ST. What constitutes a normal seminal analysis? Semen parameters of 243 fertile men. *Hum Reprod* 1998;**13**:3394–3398.

Coetzee K, Kruger TF, Lombard CJ. Predictive value of normal sperm morphology: a structured literature review. *Hum Reprod Update* 1998;**4**:73–82.

Cohn BA, Overstreet JW, Fogel RJ, Brazil CK, Baird DD, Cirillo PM. Epidemiologic studies of human semen quality: considerations for study design. *Am J Epidemiol* 2002;**155**:664–671.

Cooper TG, Yeung CH. Computer-aided evaluation of assessment of 'grade a' spermatozoa by experienced technicians. *Fertil Steril* 2006;**85**:220–224.

Cooper TG, Jockenhovel F, Nieschlag E. Variations in semen parameters from fathers. *Hum Reprod* 1991;**6**:859–866.

Cooper TG, Atkinson AD, Nieschlag E. Experience with external quality control in spermatology. *Hum Reprod* 1999;**14**:765–769.

Cooper TG, Björndahl L, Vreeburg J, Nieschlag E. Semen analysis and external quality control schemes for semen analysis need global standardization. *Int J Androl* 2002;**25**:306–311.

Dybkaer R, Solberg HE. Approved recommendation on the theory of reference values. Part 6. Presentation of observed values related to reference values. *Clinica Chemica Acta* 1987;**170**:S33–S42.

Egeland T, Haugen TB. [Nuts and bolts of percentiles]. *Tidsskr Nor Laegeforen* 2007;**127**:733–734. Norwegian.

Eustache F, Auger J, Cabrol D, Jouannet P. Are volunteers delivering semen samples in fertility studies a biased population? *Hum Reprod* 2004;**19**:2831–2837.

Fisch H, Goluboff ET, Olson JH, Feldshuh J, Broder SJ, Barad DH. Semen analyses in 1,283 men from the United States over a 25-year period: no decline in quality. *Fertil Steril* 1996;**65**:1009–1014.

Gao J, Gao ES, Yang Q, Walker M, Wu JQ, Zhou WJ, Wen SW. Semen quality in a residential, geographic and age representative sample of healthy Chinese men. *Hum Reprod* 2007;**22**:477–484.

Gao J, Gao ES, Walker M, Yang Q, Wu JQ, Zhu QX, Wen SW. Reference values of semen parameters for healthy Chinese men. *Urol Int* 2008; **81**:256–262.

Garrett C, Liu DY, Clarke GN, Rushford DD, Baker HW. Automated semen analysis: 'zona pellucida preferred' sperm morphometry and straight-line velocity are related to pregnancy rate in subfertile couples. *Hum Reprod* 2003;**18**:1643–1649.*

Glezerman M, Bernstein D, Zakut C, Misgav N, Insler V. Polyzoospermia: a definite pathologic entity. *Fertil Steril* 1982;**68**:605–608.

Gonzalo IT, Swerdloff RS, Nelson AL, Clevenger B, Garcia R, Berman N, Wang C. Levonorgestrel implants (Norplant II) for male contraception clinical trials: combination with transdermal and injectable testosterone. *J Clin Endocrinol Metab* 2002;**87**:3562–3572.*

Gu YQ, Wang XH, Xu D, Peng L, Cheng LF, Huang MK, Huang ZJ, Zhang GY. A multicenter contraceptive efficacy study of injectable testosterone undecanoate in healthy Chinese men. *J Clin Endocrinol Metab* 2003;**88**:562–568.*

Gunalp S, Onculoglu C, Gurgan T, Kruger TF, Lombard CJ. A study of semen parameters with emphasis on sperm morphology in a fertile population: an attempt to develop clinical thresholds. *Hum Reprod* 2001;**16**:110–114.

Guzick DS, Overstreet JW, Factor-Litvak P, Brazil CK, Nakajima S, Coutifaris C, Carson SA, Cisneros P, Steinkampf MP, Hill JA et al. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med* 2001;**345**:1388–1393.*

Handelsman DJ. Estimating familial and genetic contributions to variability in human testicular function: a pilot twin study. *Int J Androl* 1997; **20**:215–221.

Handelsman DJ. Optimal power transformations for analysis of sperm concentration and other semen variables. *J Androl* 2002; **23**:629–634.

Handelsman DJ, Dunn SM, Conway AJ, Boylan LM, Jansen RP. Psychological and attitudinal profiles in donors for artificial insemination. *Fertil Steril* 1985;**43**:95–101.

Hansen AM, Christensen JM, Sherson D. Estimation of reference values for urinary 1-hydroxypyrene and alpha-naphthol in Danish workers. *Sci Total Environ* 1995;**163**:211–219.

Haugen TB, Egeland T, Magnus O. Semen parameters in Norwegian fertile men. *J Androl* 2006;**27**:66–71.*

Hauser R, Godfrey-Bailey L, Chen Z. Does the potential for selection bias in semen quality studies depend on study design? Experience from a study conducted within an infertility clinic. *Hum Reprod* 2005; **20**:2579–2583.

Hay CJ, Brady BM, Zitzmann M, Osmanagaoglu K, Pollanen P, Apter D, Wu FC, Anderson RA, Nieschlag E, Devroey P et al. A multicenter phase IIb study of a novel combination of intramuscular androgen (testosterone decanoate) and oral progestogen (etonogestrel) for male hormonal contraception. *J Clin Endocrinol Metab* 2005; **90**:2042–2049.*

Hjollund NH, Jensen TK, Bonde JP, Henriksen TB, Andersson AM, Kolstad HA, Ernst E, Biwercman AJ, Skakkebaek NE, Olsen J. [Stress and fertility]. A follow-up study among couples planning the first pregnancy. *Ugeskr Laeger* 2000;**162**:5081–5086. Danish.

Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta* 2003; **334**:5–23.

Irvine S, Cawood E, Richardson D, MacDonald E, Aitken J. Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 557 men in Scotland over 11 years. *Br Med J* 1996;**312**:467–471.

Iwamoto T, Nozawa S, Yoshiike M, Hoshino T, Baba K, Matsushita T, Tanaka SN, Naka M, Skakkebaek NE, Jørgensen N. Semen quality of 324 fertile Japanese men. *Hum Reprod* 2006;**21**:760–765.

Jensen TK, Slama R, Ducot B, Suominen J, Cawood EH, Andersen AG, Eustache F, Irvine S, Auger S, Jouannet P et al. Regional differences in waiting time to pregnancy among fertile couples from four European cities. *Hum Reprod* 2001;**16**:2697–2704.*

Jensen TK, Carlsen E, Jørgensen N, Berthelsen JG, Keiding N, Christensen K, Petersen JH, Knudsen LB, Skakkebaek NE. Poor semen quality may contribute to recent decline in fertility rates. *Hum Reprod* 2002;**17**:1437–1440.

Joffe M. Time trends in biological fertility in Britain. *Lancet* 2000; **355**:1961–1965.

Jonckheere K, Callewaert N, Grootenhuis AJ, Kersemaekers WM, Cooper TG. Feasibility study for the centralized measurement of sperm concentration. *Int J Androl* 2005;**28**:311–315.

Jørgensen N, Andersen AG, Eustache F, Irvine DS, Suominen J, Petersen JH, Andersen AN, Auger J, Cawood EH, Horte A et al. Regional differences in semen quality in Europe. *Hum Reprod* 2001; **16**:1012–1019.*

Jouannet P, Cyzglik F, David G, Mayaux MJ, Spira A, Moscato ML, Schwartz D. Study of a group of 484 fertile men. Part I: distribution of semen characteristics. *Int J Androl* 1981;**4**:440–449.

Jouannet P, Wang C, Eustache F, Kold-Jensen T, Auger J. Semen quality and male reproductive health: the controversy about human sperm concentration decline. *APMIS* 2001;**109**:333–344.

Junqing W, Qiuying Y, Jianguo T, Wei Y, Liwei B, Yuxian L, Yumei Z, Kangshou Y, Weiqun L, Lu C et al. Reference value of semen quality in Chinese young men. *Contraception* 2002;**65**:365–368.

Kamischke A, Diebäcker J, Nieschlag E. Potential of norethisterone enanthate for male contraception: pharmacokinetics and suppression of pituitary and gonadal function. *Clin Endocrinol* 2000a;**53**:351–358.

Kamischke A, Plöger D, Venherm S, von Eckardstein S, von Eckardstein A, Nieschlag E. Intramuscular testosterone undecanoate with or without oral levonorgestrel: a randomized placebo-controlled feasibility study for male contraception. (Erratum published in Clin Endocrinol 53 661). *Clin Endocrinol* 2000b;**53**:43–52.

Kamischke A, Simoni M, Schrameyer K, Lerchl A, Nieschlag E. Is inhibin B a pharmacodynamic parameter for FSH in normal men? *Eur J Endocrinol* 2001a;**144**:629–637.

Kamischke A, Venherm S, Plöger D, von Eckardstein S, Nieschlag E. Intramuscular testosterone undecanoate and norethisterone enanthate in a clinical trial for male contraception. *J Clin Endocrinol Metab* 2001b; **86**:303–309.

Kamischke A, Heuermann T, Krüger K, von Eckardstein S, Schellschmidt I, Rübig A, Nieschlag E. An effective hormonal male contraceptive using testosterone undecanoate with oral or injectable norethisterone preparations. *J Clin Endocrinol Metab* 2002;**87**:530–539.

Katz DF, Overstreet JW, Hanson FW. A new quantitative test for sperm penetration into cervical mucus. *Fertil Steril* 1980;**33**:179–186.

Koduah M, Iles TC, Nix BJ. Centile charts 1: new method of assessment for univariate reference intervals. *Clin Chem* 2004;**50**:901–906.

Larsen SB, Abell A, Bonde JP. Selection bias in occupational sperm studies. *Am J Epidemiol* 1998;**147**:681–685.

Lemcke B, Behre HM, Nieschlag E. Frequently subnormal semen profiles of normal volunteers recruited over 17 years. *Int J Androl* 1997; **20**:144–152.

Leto S, Frensilli F. Changing parameters of donor semen. *Fertil Steril* 1981;**36**:766–770.

MacLeod J. The male factor in fertility and infertility; an analysis of ejaculate volume in 800 fertile men and in 600 men in infertile marriage. *Fertil Steril* 1950;**1**:347–361.

MacLeod J. Semen quality in 1000 men of known fertility and in 800 cases of infertile marriage. *Fertil Steril* 1951;**2**:115–39.

MacLeod J, Gold RZ. The male factor in fertility and infertility. II. Spermatozoon counts in 1000 men of known fertility and in 1000 cases of infertile marriage. *J Urol* 1951a;**66**:439–449.

MacLeod J, Gold RZ. The male factor in fertility and infertility. III. An analysis of motile activity in the spermatozoa of 1000 fertile men and 1000 men in infertile marriage. *Fertil Steril* 1951b;**2**:187–204.

MacLeod J, Wang Y. Male fertility potential in terms of semen quality: a review of the past, a study of the present. *Fertil Steril* 1979;**31**:103–116.

Macomber D, Sanders MD. The spermatozoa count: its value in the diagnosis, prognosis and concentration in fertile and infertile men. *N Engl J Med* 1929;**200**:981–984.

McLachlan RI, McDonald J, Rushford D, Robertson DM, Garrett C, Baker HW. Efficacy and acceptability of testosterone implants, alone or in combination with a 5α-reductase inhibitor, for male hormonal contraception. *Contraception* 2000;**62**:73–78.*

McLachlan R I, O'Donnell L, Stanton PG, Balourdos G, Frydenberg M, de Kretser DM, Robertson DM. Effects of testosterone plus medroxyprogesterone acetate on semen quality, reproductive hormones, and germ cell populations in normal young men. *J Clin Endocrinol Metab* 2002;**87**:546–556.*

Menkveld R, Wong WY, Lombard CJ, Wetzels AM, Thomas CM, Merkus HM, Steegers-Theunissen RP. Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort towards standardization of *in-vivo* thresholds. *Hum Reprod* 2001;**16**:1165–1171.

Meriggiola MC, Bremner WJ, Paulsen CA, Valdiserri A, Incorvaia L, Motta R, Pavani A, Capelli M, Flamigni C. A combined regimen of cyproterone acetate and testosterone enanthate as a potentially highly effective male contraceptive. *J Clin Endocrinol Metab* 1996;**81**:3018–3023.*

Meriggiola MC, Costantino A, Cerpolini S, Bremner WJ, Huebler D, Morselli-Labate AM, Kirsch B, Bertaccini A, Pelusi C, Pelusi G. Testosterone undecanoate maintains spermatogenic suppression induced by cyproterone acetate plus testosterone undecanoate in normal men. *J Clin Endocrinol Metab* 2003;**88**:5818–5826.*

Meriggiola MC, Costantino A, Saad F, D'Emidio L, Labate AM, Bertaccini A, Bremner W, Rudolph I, Ernst M, Kirsch B *et al.* Norethisterone enanthate plus testosterone undecanoate for male contraception: effects of various injection intervals on spermatogenesis, reproductive hormones, testis and prostate. *J Clin Endocrinol Metab* 2005;**90**:2005–2014.*

Muller A, De La Rochebrochard E, Labbe-Decleves C, Jouannet P, Bujan L, Mieusset R, Le Lannou D, Guerin JF, Benchaib M, Slama R *et al.* Spira A. Selection bias in semen studies due to self-selection of volunteers. *Hum Reprod* 2004;**19**:2838–2844.

Murer AJ, Christensen JM, Midtgaard T. Determination of the urinary metabolites of styrene: estimation of the method evaluation function and evaluation of reference values in Danish subjects. *Int Arch Occup Environ Health* 1994;**65**:313–318.

Nallella KP, Sharma RK, Aziz N, Agarwal A. Significance of sperm characteristics in the evaluation of male infertility. *Fertil Steril* 2006;**85**:629–634.

Ombelet W, Bosmans E, Janssen M, Cox A, Vlasselaer J, Gyselaers W, Vandeput H, Gielen J, Pollet H, Maes M *et al.* Semen parameters in a fertile versus subfertile population: a need for change in the interpretation of semen testing. *Hum Reprod* 1997;**12**:987–993.

Page EW, Houlding F. The clinical interpretation of 1000 semen analyses among applicants for sterility studies. *Fertil Steril* 1951;**2**:140–151.

Pal PC, Rajalakshmi M, Manocha M, Sharma RS, Mittal S, Rao DN. Semen quality and sperm functional parameters in fertile Indian men. *Andrologia* 2006;**38**:20–25.

Pasqualotto FF, Sobreiro BP, Hallak J, Athayde KS, Pasqualotto EB, Lucon AM. High percentage of abnormal semen parameters in a prevasectomy population. *Fertil Steril* 2006;**85**:954–960.

Pater C. The blood pressure 'uncertainty range'—a pragmatic approach to overcome current diagnostic uncertainties (II). *Curr Control Trials Cardiovasc Med* 2005;**6**:5–21.

Paulsen CA, Berman NG, Wang C. Data from men in greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. *Fertil Steril* 1996;**65**:1015–1020.

PetitClerc C, Solberg HE. Approved recommendations (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *Clinica Chemica Acta* 1987;**170**:S3–S12.

Qoubaitary A, Meriggiola C, Ng CM, Lumbreras L, Cerpolini S, Pelusi G, Christensen PD, Hull L, Swerdloff RS, Wang C. Pharmacokinetics of testosterone undecanoate injected alone or in combination with norethisterone enanthate in healthy men. *J Androl* 2006;**27**:853–867.*

Rehan N, Sobrero AJ, Fertig JW. The semen of fertile men: statistical analysis of 1300 men. *Fertil Steril* 1975;**26**:492–502.

Rowe PJ, Comhaire FH, Hargreave TB, Mellows HJ. *WHO Manual for the Standardized Investigation and Diagnosis of the Infertile Couple*. Cambridge: Cambridge University Press, 1993, 83 p.

Rowe PJ, Comhaire FH, Hargreave TB, Mahmoud AM. *WHO Manual for the Standardized Investigation, Diagnosis and Management of the Infertile Male*. Cambridge: Cambridge University Press, 2000, 91 p.

Schill W-B, Feifel M. Low acrosin activity in polyzoospermia. *Andrologia* 1984;**16**:589–591.

Slama R, Eustache F, Ducot B, Jensen TK, Jørgensen N, Horte A, Irvine S, Suominen J, Andersen AG, Auger J *et al.* Time to pregnancy and semen parameters: a cross-sectional study among fertile couples from four European cities. *Hum Reprod* 2002;**17**:503–515.*

Slama R, Kold-Jensen T, Scheike T, Ducot B, Spira A, Keiding N. How would a decline in sperm concentration over time influence the probability of pregnancy? *Epidemiology* 2004;**15**:458–465.

Solberg HE. International federation of clinical chemistry. scientific committee, clinical section. expert panel on theory of reference values and international committee for standardization in haematology standing committee on reference values. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of references values. *Clin Chim Acta* 1987;**165**:111–118.

Solberg HE. The IFCC recommendation on estimation of reference intervals. The RefVal program. *Clin Chem Lab Med* 2004;**42**:710–714.

Solberg HE. Establishment and use of reference values. In: Burts CA, Ashwood ER, Bruns DE (eds). *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics.* Philadelphia: Elsevier Inc, 2006. 425–448.

Sultan Sheriff D. Setting standards of male fertility I. Semen analyses in 1500 patients—a report. *Andrologia* 1983;**15**:687–692.

Stewart TM, Liu DY, Garrett C, Jørgensen N, Brown EH, Baker HWG. Associations between andrological measures, hormones and semen quality in ferile Australian men: inverse relationship between obesity and sperm output. *Hum Reprod* 2009;**24**:1561–1568.

Swan SH. Do environmental agents affect semen quality? *Epidemiology* 2003;**14**:261–262.

Swan SH, Brazil C, Drobnis EZ, Liu F, Kruse RL, Hatch M, Redmon JB, Wang C, Overstreet JW. Geographic differences in semen quality of fertile U.S. males. *Environ Health Perspect* 2003;**111**:414–420.*

te Velde ER, Eijkemans R, Habbema HD. Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction. *Lancet* 2000;**355**:1928–1929.

Tielemans E, Burdorf A, te Velde E, Weber R, van Kooij R, Heederik D. Sources of bias in studies among infertility clients. *Am J Epidemiol* 2002;**156**:86–92.

Töpfer-Petersen E, Völcker C, Heissler E, Schill W-B. Absence of acrosome reaction in polyzoospermia. *Andrologia* 1987;**19**:225–228.

Tournaye H, Staessen C, Camus M, Verheyen G, Devroey P, Van Steirteghem A. No evidence for a decreased fertilizing potential after *in-vitro* fertilization using spermatozoa from polyzoospermic men. *Hum Reprod* 1997;**12**:2183–2185.

Turner L, Conway AJ, Jimenez M, Liu PY, Forbes E, McLachlan R I, Handelsman DJ. Contraceptive efficacy of a depot progestin and androgen combination in men. *J Clin Endocrinol Metab* 2003; **88**:4659–4667.*

van der Merwe FH, Kruger TF, Oehninger SC, Lombard CJ. The use of semen parameters to identify the subfertile male in the general population. *Gynecol Obstet Invest* 2005;**59**:86–91.

Van Waart J, Kruger TF, Lombard CJ, Ombelet W. Predictive value of normal sperm morphology in intrauterine insemination (IUI): a structured literature review. *Hum Reprod Update* 2001;**7**:495–500.

Van Waeleghem K, De Clercq N, Vermeulen L, Schoonjons F, Comhaire F. Deterioration of sperm quality in young healthy Belgian men. *Hum Reprod* 1996;**11**:325–329.

von Eckardstein S, Noe G, Brache V, Nieschlag E, Croxatto H, Alvarez F, Moo-Young A, Sivin I, Kumar N, Small M et al. International Committee for Contraception Research, The Population Council. A clinical trial of 7 alpha-methyl-19-nortestosterone implants for possible use as a long-acting contraceptive for men. *J Clin Endocrinol Metab* 2003; **88**:5232–5239.*

Walton MJ, Kumar N, Baird DT, Ludlow H, Anderson RA. 7α-methyl-19-nortestosterone (MENT) vs testosterone in combination with etonogestrel implants for spermatogenic suppression in normal men. *J Androl* 2007;**28**:679–688.*

Wang C, Chan SY, Leung A, Ng RP, Ng M, Tang LC, Ma HK, Tsoi WL, Kwan M. Cross-sectional study of semen parameters in a large group of normal Chinese men. *Int J Androl* 1985;**8**:257–274.

Wang C, Wang XH, Nelson AL, Lee KK, Cui YG, Tong JS, Berman N, Lumbreras L, Leung A, Hull L et al. Levonorgestrel implants enhanced the suppression of spermatogenesis by testosterone implants: comparison between Chinese and non-Chinese men. *J Clin Endocrinol Metab* 2006;**91**:460–470.*

World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction*, 2nd edn. Cambridge: Cambridge University Press, 1987, 80 p.

World Health Organization. Task Force on Methods for the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia in normal men. *Lancet* 1990;**336**:955–959.*

World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction*, 3rd edn. Cambridge: Cambridge University Press, 1992, 107 p.

World Health Organization. Task Force on Methods for the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia and oligozoospermia in normal men. *Fertil Steril* 1996;**65**:821–829.*

World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction*, 4th edn. Cambridge: Cambridge University Press, 1999, 128 p.

Zinaman MJ, Brown CC, Selevan SG, Clegg ED. Semen quality and human fertility: a prospective study with healthy couples. *J Androl* 2000; **21**:145–153.

Zukerman Z, Rodriguez-Rigau LJ, Smith KD, Steinberger E. Frequency distribution of sperm counts in fertile and infertile males. *Fertil Steril* 1977;**28**:1310–1313.

Appendix