World Health Organization reference values for human semen characteristics‡

Trevor G. Cooper¹,¹⁰, Elizabeth Noonan², Sigrid von Eckardstein³, Jacques Auger⁴, H.W. Gordon Baker⁵, Hermann M. Behre⁶, Trine B. Haugen⁷, Thinus Kruger⁸, Christina Wang⁹, Michael T. Mbizvo³,†, and Kirsten M. Vogelsong³,†

¹Centre of Reproductive Medicine and Andrology of the University, Domagkstrasse 11, D-48129 Munster, Germany ²Fred Hutchinson Cancer Research Center, SCHARP Statistical Center for HIV/AIDS Research and Prevention, Seattle, WA, USA ³UNDP/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 ⁴Service d’Histologie-Embryologie, Biologie de la Reproduction/CECOS, Pavillon Cassini, Hôpital Cochin, Paris, France ⁵Department of Obstetrics and Gynaecology, University of Melbourne, Royal Women’s Hospital, Carlton, VIC, Australia ⁶Center for Reproductive Medicine and Andrology, University Hospital Halle, Martin Luther University, Halle, Germany ⁷Faculty of Health Sciences, Oslo University College, Oslo, Norway ⁸Reproductive Biology Unit, Stellenbosch University, Tygerberg, South Africa ⁹Harbor-UCLA Medical Center and Los Angeles Biomedical Research Institute, Torrance, CA, USA

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

TABLE OF CONTENTS

- Introduction
- Materials and Methods
  - Study populations
  - Analytical methods and quality control
  - Identification of data and handling of the datasets
  - Statistical analysis
- Results
  - Ages of men providing semen samples
  - Reference values for human semen
  - Statistical differences in semen characteristics among the various populations
- Discussion
  - Choice of reference limits
  - Comparison of the current with published reference limits
  - Comparisons of semen characteristics among different populations of men
  - Significance of lower reference limits
  - Limitations of the current reference values

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.

© World Health Organization [2009]. All rights reserved. The World Health Organization has granted Oxford University Press permission for the reproduction of this article.
**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—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 $20 \times 10^6$/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 proven 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 Frensili, 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 indeterminate 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>
<thead>
<tr>
<th>Category</th>
<th>City, Country, Continent</th>
<th>N</th>
<th>Semen Volume</th>
<th>Sperm Concentration</th>
<th>Sperm Motility</th>
<th>Sperm Morphology</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unscreened</td>
<td>Sydney, Australia, AU</td>
<td>225</td>
<td>S</td>
<td>N</td>
<td>37</td>
<td>Q</td>
<td>Turner et al. (2003)</td>
</tr>
<tr>
<td>Unscreened</td>
<td>Melbourne, Australia, AU</td>
<td>41</td>
<td>C</td>
<td>N</td>
<td>RT</td>
<td>S8</td>
<td>WHO (1996), McLachlan et al. (2000)</td>
</tr>
<tr>
<td>Unscreened</td>
<td>Edinburgh, UK, EU</td>
<td>84</td>
<td>W</td>
<td>N</td>
<td>37</td>
<td>D</td>
<td>Brady et al. (2004, 2006), Hay et al. (2005), Walton et al. (2007)</td>
</tr>
<tr>
<td>Unscreened</td>
<td>Manchester, UK, EU</td>
<td>24</td>
<td>P</td>
<td>N</td>
<td>37</td>
<td>P8</td>
<td>Unpublished results</td>
</tr>
<tr>
<td>Unscreened</td>
<td>Los Angeles, USA, AM</td>
<td>332</td>
<td>P</td>
<td>N</td>
<td>37</td>
<td>P8</td>
<td>Gonzalo et al. (2002), Qoubaitary et al. (2006), Wang et al. (2006)</td>
</tr>
<tr>
<td>Unscreened</td>
<td>Santiago, Chile, AM</td>
<td>60</td>
<td>P,S</td>
<td>M,N</td>
<td>37</td>
<td>P</td>
<td>von Eckardstein et al. (2003), Unpublished results</td>
</tr>
<tr>
<td>Unscreened</td>
<td>Münster, Germany, EU</td>
<td>199</td>
<td>GC</td>
<td>N</td>
<td>37</td>
<td>P</td>
<td>Büchter et al. (1999), Kamischke et al. (2000a, b, 2001a, b, 2002), Unpublished results</td>
</tr>
<tr>
<td>Fathers TTP</td>
<td>Melbourne, Australia, AU</td>
<td>206</td>
<td>C</td>
<td>N</td>
<td>RT</td>
<td>S8</td>
<td>Stewart et al. (2009)</td>
</tr>
<tr>
<td>Fathers TTP</td>
<td>Paris, France; Turku, Finland; Copenhagen, Denmark; Edinburgh, Scotland, EU</td>
<td>900</td>
<td>W</td>
<td>N,B,T,Ma</td>
<td>37</td>
<td>S8</td>
<td>Auger et al. (2001), Jørgensen et al. (2001), Slama et al. (2002)</td>
</tr>
<tr>
<td>Fathers TTP</td>
<td>Columbia, USA, AM</td>
<td>593</td>
<td>W</td>
<td>N</td>
<td>37</td>
<td>P8</td>
<td>Swan et al. (2003)</td>
</tr>
<tr>
<td>Fathers TTP</td>
<td>Oslo, Norway, EU</td>
<td>89</td>
<td>W</td>
<td>N</td>
<td>37</td>
<td>P8</td>
<td>Haugen et al. (2006)</td>
</tr>
<tr>
<td>Fathers TTP</td>
<td>Copenhagen, Denmark, EU</td>
<td>165</td>
<td>C</td>
<td>M,B,T</td>
<td>37</td>
<td>P</td>
<td>Bonde et al. (1998), Jensen et al. (2001)</td>
</tr>
<tr>
<td>Fathers no TTP</td>
<td>Davis, USA, AM</td>
<td>606</td>
<td>P</td>
<td>Mi</td>
<td>37</td>
<td>P8</td>
<td>Guzick et al. (2001)</td>
</tr>
<tr>
<td>Fathers no TTP</td>
<td>Münster, Germany, EU</td>
<td>58</td>
<td>GC</td>
<td>N</td>
<td>37</td>
<td>P</td>
<td>Kamischke et al. (2001a), Unpublished result</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Szeged, Hungary, EU</td>
<td>11 + 5</td>
<td>GC</td>
<td>M</td>
<td>RT</td>
<td>P</td>
<td>WHO (1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Singapore, AS</td>
<td>3 + 1</td>
<td>P</td>
<td>M</td>
<td>RT</td>
<td>EN</td>
<td>WHO (1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Sydney, Australia, AU</td>
<td>61 + 23</td>
<td>S</td>
<td>N</td>
<td>37</td>
<td>Q</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Melbourne, Australia, AU</td>
<td>45 + 18</td>
<td>W</td>
<td>N</td>
<td>RT</td>
<td>S</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Turku, Finland, EU</td>
<td>21 + 7</td>
<td>C</td>
<td>B</td>
<td>37</td>
<td>H</td>
<td>WHO (1990)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Manchester, UK, EU</td>
<td>22 + 5</td>
<td>P</td>
<td>N</td>
<td>37</td>
<td>P</td>
<td>WHO (1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Bicêtre, France, EU</td>
<td>11 + 4</td>
<td>P</td>
<td>B</td>
<td>37</td>
<td>S</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Los Angeles, USA, AM</td>
<td>16 + 5</td>
<td>S</td>
<td>N</td>
<td>RT</td>
<td>B</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Beijing, China, AS</td>
<td>56 + 2</td>
<td>P</td>
<td>N</td>
<td>RT</td>
<td>P</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Nanjing, China, AS</td>
<td>56 + 6</td>
<td>P</td>
<td>N</td>
<td>RT</td>
<td>P</td>
<td>WHO (1990, 1996)</td>
</tr>
</tbody>
</table>
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 case of morphology assessments where variability among centres could be attributed to differences in methodologies.

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 \( i \) and letting \( v_i \) be the estimated variance of \( b_i \). The pooled estimate of the fifth centile (Poolb) is 

\[
(\hat{b}_i/v_i + b_2/v_2 + \cdots)/(1/v_1 + 1/v_2 + \cdots) = w_1b_1 + w_2b_2 + \cdots
\]

where 

\[
w_1 = (1/v_1)/(1/v_1 + 1/v_2 + \cdots).
\]

The estimate of the standard error (Poolb) is \( w_1v_1 + w_2v_2 + \cdots \). 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.

### Table I

<table>
<thead>
<tr>
<th>Category¹</th>
<th>City, Country, Continent²</th>
<th>N³</th>
<th>Semen Volume⁴</th>
<th>Sperm Concentration⁵</th>
<th>Sperm Motility⁶</th>
<th>Sperm Morphology⁷</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>scr-no TTP</td>
<td>Seattle, USA, AM</td>
<td>41+22</td>
<td>P</td>
<td>N</td>
<td>RT</td>
<td>P</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>scr-no TTP</td>
<td>Chengdu, China, AS</td>
<td>29+17</td>
<td>P</td>
<td>N</td>
<td>RT</td>
<td>P</td>
<td>WHO (1990, 1996)</td>
</tr>
<tr>
<td>Screened</td>
<td>Melbourne, Australia, AU</td>
<td>84</td>
<td>P</td>
<td>N</td>
<td>37</td>
<td>H8</td>
<td>McLachlan et al. (2000, 2002)</td>
</tr>
<tr>
<td>Screened</td>
<td>Manchester, UK, EU</td>
<td>29</td>
<td>P</td>
<td>N</td>
<td>37</td>
<td>P8</td>
<td>Unpublished results</td>
</tr>
<tr>
<td>Screened</td>
<td>Beijing, China, AS</td>
<td>263</td>
<td>C</td>
<td>N</td>
<td>RT</td>
<td>P</td>
<td>Gu et al. (2003)</td>
</tr>
</tbody>
</table>

¹TTP, time (<12 to >12 months) to pregnancy; no TTP, no time to pregnancy recorded; scr, screened; ²AU, Australasia; AM, Americas; EU, Europe; AS, Asia; ³Number of samples [where two values are recorded for WHO (1990, 1996) studies, they relate to populations of screened men and fathers, respectively (scr + no TTP)]; ⁴GC, 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; ⁵B, Bürker–Türk chamber; M, Malassez chamber; Ms, Malassez chamber; ⁶Microcell chamber (data from these chambers were not used in the analyses); N, Neubauer chamber; ⁷T, Thoma chamber; ⁸37°C, 37°C, RT, room temperature; ⁹Stains: B, Bryan–Leishman; D, DiffQuik; H, Haematoxylin and Eosin; P, Papanicolaou; Q, QuickDip; S, Shorr; ¹⁰centres providing normal sperm morphology data.
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, Erkrah, 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 (±SD) age of 31 ± 5 years (range 18–53) and only 10 men were over age 45 whereas the ‘fertile men of unknown TTP’ were of age 33 ± 5 years (20–52) and 12 were over 45. The ‘unscreened’ men were of age 33 ± 7.8 years (17–67) and 54 men were over 45; the ‘screened’ men (age 32 ± 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 5th 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 (PettiClerc 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

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

*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 5th centile provides the lower reference limit.

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

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

*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 5th 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., 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
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 × 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 (Töpf-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 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.
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 $\approx 100$ 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 $20 \times 10^6$/ml for sperm concentration and $40 \times 10^6$ 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

<table>
<thead>
<tr>
<th>Group</th>
<th>Median (and interquartile range) of the number of spermatozoa ($10^6$) per ejaculate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>Progressively motile</td>
</tr>
<tr>
<td></td>
<td>Morphologically normal</td>
</tr>
<tr>
<td>Fathers TTP $\leq$ 12 months</td>
<td>255 (142–422)</td>
</tr>
<tr>
<td>Fathers with no TTP</td>
<td>162 (87–277)*</td>
</tr>
<tr>
<td>Unscreened (general) population</td>
<td>196 (101–336)*</td>
</tr>
<tr>
<td>Screened for normozoospermia</td>
<td>180 (104–315)*</td>
</tr>
</tbody>
</table>

*Within columns, significantly different from TTP $\leq 12$ ($P < 0.05$).
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 $2 \times 10^6$ 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 ejaculation 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 concentrations and percentages of motile and normally formed spermatozoa, but higher percentages of progressively motile spermatozoa than the fathers with TTP $\leq 12$ months. However, total numbers of all, progressively motile and morphologically normal spermatozoa per ejaculate were lower than those from the TTP $\leq 12$ 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 $\leq 12$ 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 its 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 duc tal potency.

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.
participated in writing portions of the article. J.A., T.B.H., C.W. and H.W.G.B. contributed to study design, provided data and participated in editing the article. H.M.B., T.K. and M.T.M. participated in editing the article. K.M.V. provided technical assistance during data collection and analysis and participated in editing the article.

Acknowledgements

We thank Drs D.J. Handelsman, P.F.A. Van Look and T.M.M. Farley for useful comments on the manuscript, all the investigators who contributed to the database and all the men who provided semen samples. The editorial group of the fifth edition of the ‘WHO laboratory manual for the examination and processing of human semen’ (forthcoming), is thanked for recognizing the need for this study and identifying sources of data.

Funding

The study was investigator-initiated. Parts of this study were funded by the UNDP/UNFPA/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), World Health Organization (WHO). WHO provided technical assistance in study design and data analysis, but had no role in data collection or the initial drafting of the report. Decisions regarding the interpretation of the data and review and revision of the manuscript were made on the basis of discussions between WHO and the authors. The corresponding author and the sponsor had full access to the data and the corresponding author had final responsibility for submitting the manuscript.

References

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


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.


Appendix

R. Anderson, University of Edinburgh, Scotland, United Kingdom; Auger J., Hôpital Cochin, Paris, France; H.W.G. Baker, University of Melbourne, Carlton, Victoria, Australia; J.P. Bonde, Department of Occupational Medicine, University Hospital, Aarhus, Denmark; Y.-Q. Gu, National Research Institute for Family Planning, Beijing, China; D.J. Handelsman, ANZAC Research Institute, Sydney, New South Wales, Australia; T.B. Haugen, Oslo University College, Oslo, Norway; A. Kamischke, Institute of Reproductive Medicine of the University, Münster, Germany; R. McLachlan, Prince Henry’s Institute of Medical Research, Clayton, Victoria, Australia; M.C. Meriggiola, University of Bologna, Bologna, Italy; G. Noé, Instituto Chileno de Medicina Reproductiva, Santiago, Chile; J.W. Overstreet, University of California, Davis, CA, USA; N.E. Skakkebaek, University Department of Growth and Reproduction, Rigshospitalet, Copenhagen, Denmark; S.H. Swan, University of Rochester, Rochester, NY, USA; C. Wang, Harbor-UCLA Medical Center, Torrance, CA, USA; F. Wu, University of Manchester, Manchester, United Kingdom; and the World Health Organization, Geneva, Switzerland.

Submitted on January 23, 2009; resubmitted on September 7, 2009; accepted on September 21, 2009.