Clinical relevance of routine semen analysis and controversies surrounding the 2010 World Health Organization criteria for semen examination

Sandro C. Esteves
ANDROFERT, Andrology & Human Reproduction Clinic, Campinas, SP, Brazil

ABSTRACT

Semen analysis is the cornerstone of infertility evaluation as it provides information on the functional status of the seminiferous tubules, epididymis and accessory sex glands. The methods on how the human semen should be evaluated are provided by the World Health Organization, which periodically releases manuals that include specific protocols and reference standards. In 2010, the WHO published new criteria for human semen characteristics that were markedly lower than those previously reported. In this review initially it is discussed the limitations of semen analysis as a surrogate measure of a man’s ability to father a pregnancy. Secondly, it is analyzed methodology issues that could explain why the newly released reference values were different from those earlier reported. Thirdly, it is speculated on the likely effects of the 2010 WHO criteria in the management of male infertility. Due to the several inherent limitations of semen analysis as a surrogate marker of male infertility, physicians should exercise caution when interpreting results. A template for semen analysis reports that incorporates the distribution of the semen characteristics of recent fathers in centiles rather than solely the minimum thresholds could aid clinicians to better understand how a given patient results compare with the reference population. Importantly, a male infertility evaluation must go far beyond a simple semen analysis, as it has to be complemented with a proper physical examination, a comprehensive history taking, and relevant endocrine, genetic, and other investigations.

Key words: Infertility, Male; Semen Analysis; Andrology; Diagnosis; Spermatozoa; Reference Values; Therapeutics

INTRODUCTION

The World Health Organization (WHO) periodically releases manuals for the laboratory examination and processing of human semen. While laboratories use these manuals as a practical guide of standardized methods for performing semen analyses clinicians rely on the reference of normal limits for interpreting semen analysis results. The first manual, published in 1980, summarized the clinical experience and research from the previous eighty years. In its subsequent updates in 1987, 1992, 1999 and 2010, WHO manuals provided substantial improvements on how to assess the seminal parameters. The reference values that were thought to be compatible with normal male fertility have also changed (Table-1) (1–4).

In its latest fifth edition (WHO 2010) the semen analysis reference values are markedly lower than those of previous editions. Much debate has taken place thereafter, and a series of reports has questioned the validity of the newly released reference values (5–9).
In this review, it is discussed the controversy surrounding the new 2010 WHO criteria for semen analyses. First, we point out the importance and limitations of the routine semen analysis in the workup of male infertility. Then, we present the 2010 WHO cutoff values for human semen characteristics and how they compare with previous references. Third, we critically discuss the methods used for generating these new limits and present our hypotheses to explain these lowered limits. Subsequently, we analyze the likely effect of the 2010 WHO cutoff values on the clinical management of men with unexplained infertility. Lastly, we propose a practical approach to report semen analysis results for those contemplating adopting the 2010 WHO cutoff values for semen characteristics.

**Importance and limitations of semen analysis for male infertility evaluation**

Semen analysis is the most widely used biomarker to predict male fertility potential (10). It provides information on the functional status of the seminiferous tubules, epididymis and accessory sex glands, and its results are often taken as a surrogate measure of a man’s ability to father a pregnancy. Routine semen analysis include: (a) physical characteristics of semen, including liquefaction, viscosity, pH, color and odor; (b) specimen volume; (c) sperm concentration; (d) sperm motility and progression; (e) sperm morphology; (f) leukocyte quantification; and (g) fructose detection in cases where no spermatozoa are found and ejaculate volume is low (11).

Owed to its widespread availability, health care providers usually use semen analysis alone as the main marker to determine male partner referral for further investigation. However, semen characteristics that discriminate between infertile and fertile men are not well defined, and results fall within the accepted reference ranges in up to 40% of those suffering from infertility (12–14). Not only sperm production varies widely in same men but also conventional semen analysis neither tests for the diverse array of biological properties spermatozoa express as an eminently specialized cell nor accounts for putative sperm dysfunctions such as immature chromatin or fragmented DNA. In addition, there is a wide

### Table 1 - Cut-off reference values for semen characteristics as published in consecutive WHO manuals.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume (mL)</td>
<td>ND</td>
<td>≥ 2</td>
<td>≥ 2</td>
<td>≥ 2</td>
<td>1.5</td>
</tr>
<tr>
<td>Sperm count (10⁶/mL)</td>
<td>20-200</td>
<td>≥ 20</td>
<td>≥ 20</td>
<td>≥ 20</td>
<td>15</td>
</tr>
<tr>
<td>Total sperm count (10⁹)</td>
<td>ND</td>
<td>≥ 40</td>
<td>≥ 40</td>
<td>≥ 40</td>
<td>39</td>
</tr>
<tr>
<td>Total motility (% motile)</td>
<td>≥ 60</td>
<td>≥ 50</td>
<td>≥ 50</td>
<td>≥ 50</td>
<td>40</td>
</tr>
<tr>
<td>Progressive motility²</td>
<td>≥ 2³</td>
<td>≥ 25%</td>
<td>≥ 25% (grade a)</td>
<td>≥ 25% (grade a)</td>
<td>32% (a + b)</td>
</tr>
<tr>
<td>Vitality (% alive)</td>
<td>ND</td>
<td>≥ 50</td>
<td>≥ 75</td>
<td>≥ 75</td>
<td>58</td>
</tr>
<tr>
<td>Morphology (% normal forms)</td>
<td>80.5</td>
<td>≥ 50</td>
<td>≥ 75</td>
<td>≥ 75</td>
<td>58</td>
</tr>
<tr>
<td>Leukocyte count (10⁶/mL)</td>
<td>&lt; 4.7</td>
<td>&lt; 1.0</td>
<td>&lt; 1.0</td>
<td>&lt; 1.0</td>
<td>&lt; 1.0</td>
</tr>
</tbody>
</table>

¹Lower reference limits generated from the lower fifth centile value; ²Grade a = rapid progressive motility (> 25μm/s); grade b = slow/stagnant progressive motility (5-25μm/s); Normal = 50% motility (grades a+b) or 25% progressive motility (grade a) within 60 min of ejaculation; ³Forward progression (scale 0-3); ⁴Arbitrary value; ⁵Value not defined but strict criterion is suggested; ⁶Strict (Tygerberg) criterion; ND = not defined.

variation on how laboratories perform semen analysis. In this section, we will continue to discuss the major drawbacks of semen analysis for male infertility evaluation.

Biological intra-individual variability of semen parameters

The semen parameters from same individuals are highly variable. Many conditions including the duration of ejaculatory abstinence, activity of the accessory sex glands, analytical errors, and inherent biological variability account for such discrepancies (15-18). In one study the within-subject variability of twenty healthy subjects assessed over a 10-week follow-up ranged from 10.3% to 26.8% (15). Sperm concentration showed the highest within-subject variation (26.8%), followed by morphology (19.6%) and progressive motility (15.2%) whereas vitality had the lowest variation (10.3%). The utility of population-based reference values is related to the individual variability of a particular characteristic. Reference values of characteristics with attributable individuality, including the ones routinely assessed in the semen, are generally of limited utility. It means that individual subjects could present results that were very unusual for them, and such results might have been accounted when establishing the reference thresholds. For the aforesaid reasons and other uncontrolled factors such as the regression towards the mean, it is impossible to take the results of a single semen specimen as a surrogate for a man’s ability to father a child unless when at extreme low levels (19). Regression towards the mean is the phenomenon in which a variable would tend to be closer to the average on a second measurement if it were extreme in its first measurement. This uncontrolled factor should be contemplated when designing studies involving semen analysis because following an extreme random event the next random event is less likely to be extreme. It has been shown that sperm concentration and motility were significantly higher in the second test in men with previous abnormal semen analyses results (20). Regression towards the mean could be reduced in its magnitude if we used means of multiple samples (two or three in the case of semen analysis). Hence, it is prudent that clinicians request at least two semen specimens following 2-5 days of ejaculatory abstinence to allow a better understanding of the baseline semen quality status of a given individual (21-23).

Sperm dysfunctions not tested in the routine semen analysis

Up to 30% of men with difficulties to father a child have no demonstrable abnormalities after an initial male infertility workup. Additional tests have been developed to unravel functional disorders and other sperm abnormalities that cannot be identified by conventional semen analysis (11,24). Some of these tests include the hypo-osmotic swelling test, computer-assisted sperm analysis, antisperm antibody test, sperm penetration assay, hemizona assay, reactive oxygen species (ROS) test, and sperm chromatin integrity test (25). Despite being available, there are inherent difficulties to set up these tests including cost of equipment and technical complexity. In addition, their predictive value in assessing the male fertility status is either variable or unknown (26). Not surprisingly, many couples with unexplained infertility choose assisted reproduction techniques (ART) because of their widespread availability and overall success despite the male infertility cause (27). Yet, the assessment of sperm oxidative stress (OS) and DNA integrity has gained clinical importance in recent years. Oxidative stress, which is present anywhere from 30% to 80% of infertile men, is a result of the generation of ROS from contaminating leukocytes, defective sperm and antioxidant depletion (28,29). ROS target sperm DNA molecules and ultimately affect the quality of the genetic material transmitted from the parents to the offspring. Damage to sperm DNA integrity can also result from apoptosis during spermiogenesis, alterations in chromatin remodeling during spermiogenesis, as well as exposure to environmental toxicants and gonadotoxins such as chemotherapy and radiotherapy (30). Abnormal levels of DNA damage are observed in approximately 5% and 25% of infertile men with normal and abnormal semen analysis, respectively (31-33). Therefore, some authors propose that the assessment of both conditions might be included to the male infertility workup algorithm (34,35).
Evidence of poor standardization in semen analysis among laboratories

Accuracy, the degree to which the measurement reflects the true value, and precision, the reproducibility of the results, are vitally important for clinicians who rely upon the values provided by the laboratory to direct the further work-up, diagnosis and counseling of the infertile male (36). When both accuracy and precision are assured, the clinician is able to rely upon the semen analysis results to provide adequate counseling to the infertile couple. However, data from surveys of laboratory practice in the United States and the UK indicate that semen analysis techniques are still poorly standardized.

Among 536 clinical laboratories in the United States only about 60% reported abstinence and indicated the criteria adopted for sperm morphology assessments. Moreover, fewer than half of them performed quality control for commonly assessed parameters such as sperm counts, motility and morphology (37). A survey involving thirty-seven laboratories in the UK about the methods used to assess sperm morphology revealed that only 5% complied with all WHO guidelines (38). In the aforementioned study, participating laboratories had high inter-observer variability when evaluating the same specimen. These data were corroborated by another study in which interlaboratory coefficient of variation was as high as 34% for sperm concentration, 20% for total sperm motility, 40% for sperm vitality, and 70% for sperm morphology (strict criteria) (39). Discrepancies were also seen in laboratories enrolled in quality control programs, thus indicating that there is a need of global standardization among the laboratories and the providers of external quality control (40).

Owed to its complex nature, semen analysis should ideally be carried out in a dedicated Andrology laboratory attired with experienced technicians, internal and external quality control, validation of test systems, quality assurance during all testing processes, and proper in place communication with clinicians and patients (41). Despite being nonspecific for identifying male factor infertility etiologies, semen analysis is often the gateway test from which multiple expensive and often invasive treatments are based. Therefore, the importance of a reliable Andrology laboratory cannot be underestimated.

The 2010 WHO Criteria for Semen Analysis

The WHO department of reproductive health and research workgroup made important changes in the 2010 laboratory manual for the examination of human semen and sperm-cervical mucus interaction (4). While the WHO workgroup reviewed and updated in great detail all the methods delineated in previous manuals it incorporated new protocols and tests. One of its main features was the inclusion of new references ranges and limits that were markedly lower than those reported in previous manuals.

Data characterizing the semen quality of fertile men provided the reference ranges for the manual (42). For the first time, semen analysis results from recent fathers with known time-to-pregnancy (TTP), defined as months (or cycles) from stopping contraception to achieving a pregnancy, were analyzed. Raw data obtained from five studies of seven countries on three continents were pooled then assessed (43-48). Approximately 1,900 men who had fathered a child within one year of trying to initiate a pregnancy provided each one semen sample for sperm counts, motility and volume assessments. Data on sperm morphology were extracted from four studies comprising approximately 1,800 men whereas sperm vitality, assessed by the eosin-nigrosin method was obtained from approximately 400 men of two countries (43,45,47,48). The mean (± SD) male age was 31 (± 5) years (range 18-53) and only ten men were over 45 years old. Participating laboratories practiced internal and external quality control and used standardized methods for semen analysis according to the WHO manual for the examination of human semen current at the time of the original studies (42).

The 95% interval for sperm volume, count, motility, vitality, and morphology were generated, and the fifth centile was proposed as the lower cutoff limits (42). Of note, although the fifth centile was considered as a reference limit for ‘normality’, it merely represented an arbitra-
rily chosen distribution point commonly used in clinical chemistry. It was then assumed that values below these limits would come from a different population. The assessment of progressive motility according to grades, as recommended by the previous WHO manuals, was replaced by categorizing motile sperm as being ‘progressive’ or ‘non-progressive’. In addition, the strict criteria for morphology assessment was incorporated at last as the standard method. The lower limits of these distributions were lower than the values presented in previous editions except for the total sperm number per ejaculate (1-4). Leukocyte reference values (< 1x10⁶/mL) were not determined and remained the same as in previous manuals.

In the 2010 WHO manual, the nomenclature to describe deviations from reference values, using words rather than numbers, remained the same as in previous manuals.

Controversies Surrounding the Validity of the 2010 WHO Thresholds

The lower reference limits in the 2010 WHO manual aimed to provide evidence-based thresholds that may aid clinicians in estimating the relative fertility of a given patient. Besides the aforesaid limitations of routine semen analysis in evaluating the male reproductive potential, methodological concerns arise from a careful examination of the studies that generated the current reference values. In a recent review, we critically analyzed these issues and concluded that it was unsound to assume that the 2010 reference standards represented the distribution of fertile men across the globe (6). The group of studied men represented a limited population of individuals who lived in large cities in the North hemisphere but for a small subset of men from Australia. Of note it was the absence of men from densely populated areas in Asia, Middle East, Latin America and Africa, which represent the areas where most men live nowadays. This fact precludes the examination of regional and racial discrepancies that could account for semen quality variability. The selection criteria were arbitrary as stated by Cooper et al. ‘laboratories and data were identified through the known literature and personal communication with investigators and the editorial group of the fifth edition of the WHO laboratory manual’ (42). Not surprisingly, there was a significant overlap of authorship in the included studies. In addition, a single semen specimen of each man was included for the pooled analysis, thus limiting the appraisal of the already discussed large intra-individual biological variability (6).

Some authors have claimed that the lowered 2010 WHO thresholds are resulting from the declines in sperm count caused by endocrine disruptors and other environmental pollutants, such as insecticides, and pesticides (49-51). I, otherwise, conjecture that the observed discrepancies are likely to be associated with the patient selection criteria, the higher laboratory quality control standards and the methods used for semen assessment, such as the strict criteria for morphology determination. It means that methodology issues related to data generation might explain the discrepancies in the reference thresholds among WHO guidelines (52).

Collectively, these findings cast to doubt on the validity of the proposed reference range and cutoff limits about universally represent the distribution of semen results of fertile men.

Impact of the 2010 WHO Criteria for Semen Analysis in the Management of Male Infertility

Clinicians involved in the care of infertile couples still rely on the semen analysis results to determine a management plan. Semen parameters outside reference limits are taken into account not only to define male infertility but also to recommend further evaluation and treatment. One example is unexplained infertility which is based on the absence of female infertility, and the presence of at least two normal semen analysis and no identifiable causes after a thorough work-up including history, physical examination, and endocrine laboratory testing (24). The adoption of the new WHO reference values will likely lead to more men being classified as “fertile,” which is of particular importance for gynecologists who rely on semen analysis alone as a surrogate measure for male fertility. In a recent study up to 15% of men with at least one parameter below the 1999 WHO reference values...
were reclassified as "normal" by having all parameters at or above the 2010 WHO thresholds (8). We have also contemplated our own data involving 982 men seeking evaluation for infertility that had abnormal semen analysis results based on the 1990 WHO criteria. We found that approximately 39% of these men would be reclassified as "normal" by the new 2010 criteria. Morphology itself accounted for over 50% of the reclassifications (unpublished data). Patient referral for evaluation could then be postponed or not undertaken if fertility status would be based on semen analyses alone. Albeit it is ambiguous yet whether this re-classification will lead to a more cost-effective evaluation, it is also possible that it could delay the definitive diagnosis and management of the infertile couple and lead to a more pronounced infertility condition with ageing.

The current guidelines for male infertility evaluation also rely on the concept of semen abnormality for patient management. The American Urological Association (AUA) defines that the initial male evaluation should include a reproductive history and two properly performed semen analyses, and that an extended evaluation is warranted in the presence of semen abnormalities in the initial evaluation (53). In contrast, the European Association of Urology (EAU) recommends undertaking a male examination in individuals with abnormal semen analysis results (54). Surprisingly, a single seminal evaluation would then be sufficient if the semen analysis results were normal according to the EAU. These recommendations understate the limitations of the semen analysis results and do not discuss the paradigm shift that is likely to occur in referrals and management on the face of the recent changes in the WHO reference thresholds.

Similarly, the recommendation for treatment has also been based on the results of routine semen analysis. Current guidelines for varicocele propose that treatment should be offered to men with clinical varicoceles in the presence of abnormal semen analyses (55-58). Application of the new WHO reference values might lead to patients earlier deemed to be candidates for varicocele repair now be considered ineligible for treatment if their semen parameters are above the fifth centile. This may create a situation where health care providers might not reimburse treatment if semen parameters were above the new thresholds. As stated by Esteves et al. 'the concern is that by denying these men a varicocele repair we may prevent them from achieving a substantial improvement in semen parameters and a greater chance of spontaneous pregnancy'. Of note, the most recent Practice Committee report on varicocele by the American Society for Reproductive Medicine (ASRM) acknowledged the limitations of routine semen analysis and included the presence of an abnormal sperm function test as an indication for treatment (58). Yet, another example is sperm morphology results in which infertility specialists have relied on to recommend treatment modalities owed to their relationship with in vivo and in vitro fertilization (59). The thresholds of sperm morphology (strict criteria; Tygerberg method) were lowered to 4% in the 2010 WHO criteria compared with 14% in the previous 1999 standards (3,4). Infertility specialists recommend intracytoplasmic sperm injection (ICSI) instead of conventional IVF or intrauterine insemination (IUI) on the face of morphology results of below 4% owed to the markedly lower pregnancy outcomes of these two treatment methods when using semen with low proportion of normal sperm (60,61). Interestingly, the distribution of semen analysis results of fertile men in centiles, as shown by the new WHO standards, clearly shows that though 5% of the studied men had morphology values below the 4% cutoff point they still could initiate an unassisted pregnancy within twelve months of unprotected intercourse (6,42).

In summary, these considerations raise the question on how the 2010 WHO references thresholds would affect the current male infertility practice. It should be noted however, that reference values, as proposed by the WHO, merely represent the distribution of semen parameters of a limited group of recent fathers. Physicians treating infertile couples should exercise circumspection when interpreting the results of routine semen analysis. Semen analysis alone is only a tool among several others for determining clinical care. The male infertility evaluation must
go far beyond a simple semen analysis, as it has to be complemented with a proper physical examination, a comprehensive history taking, and relevant endocrine, genetic, and other investigations (10,11).

A Proposal of a New Template for Semen Analysis Report

Semen analysis reports usually present the specimen data and include the cutoff limits as a reference for interpretation (62). Despite having updated to the 2010 WHO criteria, our Andrology laboratory has changed the way results are reported. We have included the 95% reference interval of the semen characteristics from recent fathers, as generated by the WHO workgroup, instead of only providing the lower reference limits.

The reason why we have included the 95% reference interval was because we believe it is clinically useful to determine in which centile the patient specimen fits in comparison with the reference standards instead of simply classify the specimen as ‘normal’ or ‘abnormal’. This approach would be more realistic and clinicians would have a better understanding of the patient’s seminal profile by comparing the specimen results with the reference group. I therefore propose that laboratories willing to adopt the 2010 WHO references include the full reference interval, as presented in Figure-1.

Our laboratory has also abandoned nomenclature; i.e., using words such as ‘oligozoospermia’, ‘asthenozoospermia’, ‘teratozoospermia’, etc., to describe deviations from reference limits. We think it is unsound to use such terminology for several reasons. First, these definitions are confusing. For instance, clinicians erroneously have used oligospermia and oligozoospermia as synonymous. While the former refers to ‘volume’, the latter refers to ‘number’. Second, owed to the changes in the reference values according to the WHO manual edition, patients formerly classified as having oligozoospermia (sperm count less than the reference value) could now fall within the accepted reference limits. Such labeling poses a problem for researchers and clinicians who might erroneously interpret that the seminal profile has changed because of a given intervention. Lastly, nomenclature cannot determine the magnitude of the effect, and as a remedy other words such as ‘mild’, ‘severe’, and ‘moderate’ have been introduced as qualifiers. Our opinion is that semen results should be reported numerically to allow proper temporal comparison among analyses of same individuals, as well as comparison across different studies irrespective of the time they have been published.

CONCLUSIONS

The 2010 WHO semen analysis criteria are likely to have a significant effect on the management of male infertility, including reclassification of ‘normal’ and ‘abnormal’ semen analyses reports, deferment of patient referral for proper evaluation and recommendation for treatment. These new reference limits were derived from a limited number of semen samples used to initiate natural conceptions. Albeit values below the thresholds may indicate a need for infertility treatment they cannot be used to determine the nature of that treatment. Several methodological shortcomings are associated with the new references standards that might explain why references were lowered in comparison with previous WHO guidelines. Semen parameters within the reference interval do not guarantee fertility nor do values outside those limits necessarily imply male infertility or pathology. Physicians treating infertile couples should exercise circumspection when interpreting the results of routine semen analysis. Semen analysis alone is usually insufficient for the diagnosis because it does not account for sperm dysfunction, such as immature chromatin, oxidative stress and DNA damage. Semen quality must be interpreted within the context of the patient’s clinical information. The male infertility evaluation must go far beyond a simple semen analysis, as it has to be complemented with a proper physical examination, a comprehensive history taking, and relevant endocrine, genetic, and other investigations.
Figure 1 - Modified semen analysis report template. The main difference between the routinely used templates and this one is the inclusion of the ‘centile’ distribution of semen characteristics from the reference population rather than solely the lower thresholds. The patient values (left column) are then compared with the reference limits thus aiding the clinician to appreciate how a given patient seminal profile fits within the centile distribution.

![Semen Analysis Report](image)

*Reference values obtained from a population of approximately 2,000 recent fathers from five countries (time to pregnancy from stopping contraception to obtaining natural conception ≤12 months). Values may not be representative of the Brazilian population. Values outside the reference limits do not necessarily imply infertility; they should be interpreted in conjunction with clinical information. Sperm analysis performed in accordance to the World Health Organization laboratory manual for the examination and processing of human semen, fifth edition, 2010.*
CONFLICT OF INTEREST

None declared.

REFERENCES


Correspondence address:
Sandro C. Esteves, MD, PhD
ANDROFERT, Andrology and Human Reproduction Clinic
Av. Dr. Heitor Penteado, 1464
Campinas, SP, 13075-460, Brazil
Fax: +55 19 3294-6992
E-mail: s.esteves@androfert.com.br