ESHRE special interest group for andrology basic semen analysis course: a continued focus on accuracy, quality, efficiency and clinical relevance†

C.L.R. Barratt 1,*, L. Bjo¨rndahl 2, R. Menkveld 3, and D. Mortimer 4

1 Reproductive and Developmental Biology, Maternal and Child Health Science Laboratories, Centre for Oncology and Molecular Medicine, Ninewells Hospital, University of Dundee, Dundee, DD1 9SY, Scotland 2 Centre for Andrology and Sexual Medicine, Karolinska University Hospital and Karolinska Institutet, Huddinge, Stockholm 141 86, Sweden 3 Andrology Laboratory, Department of Obstetrics and Gynaecology, Tygerberg Academic Hospital and University of Stellenbosch, Tygerberg 7505, South Africa 4 Oooza Biomedical Inc, Box 93012 Caulfeild Village, West Vancouver, BC, Canada V7V 3G4

*Correspondence address. E-mail: c.barratt@dundee.ac.uk

Submitted on August 19, 2011; resubmitted on August 19, 2011; accepted on August 22, 2011

ESHRE has been running courses for basic semen analysis since 1994 and course material has been updated regularly in response to new findings and publications. Following publication of the 5th edition of the WHO laboratory manual, entitled WHO Laboratory Manual for the Examination and Processing of Human Semen (WHO5), the Subcommittee for training of the ESHRE Special Interest Group for Andrology evaluated potential amendments to its course. In respect of the updated ESHRE course, there are eight particular areas of discourse that are reviewed (i) maintaining the four-class differential motility count allowing distinction between rapid and slow progressive sperm for assisted reproduction technology. (ii) Maintaining the four-category assessment for sperm morphology with calculation of the teratozoospermic index. (iii) Continuing to advocate the use of three categories of results: ‘normal’, ‘borderline’ and ‘abnormal’ with respect to the clinical interpretation of the data. (iv) Presenting clear and unequivocal methods for performing assessments e.g. morphology. (v) Correcting the inconsistencies in WHO5, some of which are actually erroneous. (vi) Reducing the requirements for substantial extra work for what are unestablished improvements in accuracy and/or precision in the final results. (vii) Presentation of logical methods of sperm preparation. (viii) Discussion of the suddenly changed limits between fertile and subfertile men.

Key words: semen analysis / WHO / quality control / training

Background

ESHRE has been running courses for basic semen analysis (BSA) since 1994 (Bjo¨rndahl et al., 2002). The material in the original courses was based on published proven protocols and methods including, for example, Practical Laboratory Andrology (Mortimer, 1994). The course material has been updated regularly in response to new findings and publications by, for example, the WHO (1999), and laboratory manuals have been produced and updated by ESHRE and the Nordic Association for Andrology (Krist and Bjo¨rndahl, 2002; NAFA, 2002). Concordant with the ESHRE BSA courses, an external quality assurance scheme, based in the Karolinska University Hospital, was introduced in 1999 and has provided laboratory staff throughout Europe and further afield with an external quality assurance scheme that not only aids the monitoring of laboratory performance but also includes an effective quality improvement component (www.eshre.eu/ESHRE/English/Specialty-Groups/SIG/Andrology/External-Quality-Control/page.aspx/104). The BSA courses, more than 50 of which have been run in 14 countries to date, have consistently shown that the methods adopted are effective at training laboratory technicians (Bjo¨rndahl et al., 2002).

Following the meeting of the Editorial Committee in 2005, the 5th edition of the WHO laboratory manual, entitled WHO Laboratory Manual for the Examination and Processing of Human Semen was published in 2010 (‘WHO5’: WHO, 2010). Supplementary to this manual are peer-reviewed reference ranges for fertile men based on data from between 428 and 1941 semen analyses collected from fertile populations in several laboratories (Cooper et al., 2010). Since its inception in 1980, the usefulness of the various editions of the WHO laboratory manual has been discussed in a plethora of articles.
articles considering its methods, selection of techniques, reference ranges, etc.

Notwithstanding these discussions, the WHO manual remains a cornerstone in laboratory andrology, and has undoubtedly helped to develop international standards. Interestingly, while WHO5 is more comprehensive than previous versions, and has adopted an ‘evidence-based methodology’ where appropriate, it also contains a number of differences from its predecessors.

The issues
Following publication of WHO5, the ESHRE Subcommittee for training of the Special Interest Group for Andrology (SIGA) evaluated potential amendments to its BSA course. In respect of the updated ESHRE BSA course, there are eight particular areas that require explanation.

Sperm motility
Surprisingly, WHO5 abandons the distinction between slow- and rapid-progressive spermatozoa; a controversial decision. The reasoning behind this appears primarily based on the observation that poorly trained technicians cannot distinguish between the two categories repeatedly and reliably—but this should be obvious, since poorly trained technicians cannot do many things, e.g. count spermatozoa (Mortimer, 1994). Indeed, MacLeod and Gold (1951) noted 60 years ago that the quality of sperm motility was a prime factor to be considered in a semen analysis. In addition, they stated that the achievement of intra- and inter-observer standardization was essential in any method used to assess sperm motility, and also that observers must be properly trained. The arguments posited by the WHO have been refuted elsewhere (Björndahl, 2010; Eliasson, 2010) and, very importantly, there are clinical data both from manual sperm motility assessments and computer-aided sperm analysis showing the distinction of rapidly progressive spermatozoa to be biologically—and hence clinically—important. This evidence ranges from the ability of spermatozoa to penetrate cervical mucus (Aitken et al., 1985; Mortimer et al., 1986) and in vivo conceptions (Comhaire et al., 1988; Barratt et al., 1992), to clinical outcome studies in donor insemination (Irvine and Aitken, 1986), IUI (Bollendorf et al., 1996) and IVF (Bollendorf et al., 1996; Sifer et al., 2005). Even in regard to ICSI, the straight line velocity of the individual spermatozoa subsequently injected into the oocyte has been shown to have a significant effect on fertilization outcome (Van den Bergh et al., 1998). It is therefore both scientifically and clinically inappropriate to abandon the differentiation of rapid- and slow-progressive spermatozoa. The ESHRE BSA course teaches appropriate methods to become able to reliably distinguish between these categories which, combined, give the overall % progressive spermatozoa.

Sperm morphology
WHOS has fully adopted the Tygerberg Strict Criteria for normal sperm morphology (Menkveld et al., 1990). These criteria are based on the typical morphology of spermatozoa that are able to migrate through cervical mucus and bind to the zona pellucida, even though in ‘normal’ men only a small proportion of spermatozoa correspond to the typical morphology (Menkveld et al., 2011). As a consequence, an extra measure that includes the different types of abnormalities can provide additional useful information by identifying men with more severe disturbances in sperm form and related function, e.g. the Multiple Anomalies Index (MAI), see Jouannet et al. (1988) and the Teratozoospermia Index (TZI), see Menkveld et al. (1998), Mortimer et al. (1990) and Mortimer and Menkveld (2001). The TZI is an indirect indication of (1) the risk of what appeared to be normal spermatozoa actually having defects that were invisible at the level of observation and (2) just how badly affected spermogenesis was in the man, and hence how impaired his sperm fertilizing ability might be (Mortimer and Menkveld, 2001). The TZI can provide extra information in cases where there are very few morphological normal forms, introducing a dynamic range into the interpretation of the role of sperm morphology, especially when a perceived difference between 4 and 6% normal forms is considered to reflect a major difference in clinical significance. Such additional information would seem highly pertinent when interpreting sperm morphology assessments based on counts of just 200 spermatozoa, since even when 400 spermatozoa are counted, % normal forms values of 4 and 6% are statistically indistinguishable, having 95% confidence intervals of 2–6 and 4–9, respectively (see p. 307 of Björndahl et al., 2010).

Unfortunately, in WHO5 the assessment of multiple sperm defects has been relegated to ‘Optional Procedures’, although calculation of the TZI has been corrected to be out of four instead of three, as erroneously used in the 4th edition (WHO, 1999). Even if only % normal spermatozoa is reported, the actual assessment procedure should include all the characteristics/criteria needed for TZI since recording the prevalence of the four categories of morphological deviations is essential for quality control (internal and external) purposes which require the ability to analyze observers’, or laboratories’, differences from target values. In terms of clinical application of the TZI, the consensus-based WHO Manual for the Standardized Investigation, Diagnosis and Management of the Infertile Male (Rowe et al., 2000) commented that together with the introduction of the Tygerberg Strict Criteria in the 1999 WHO laboratory manual, the TZI had been included to provide additional information to facilitate discrimination of the extent of impairment of sperm functional potential in men with very low numbers of normal spermatozoa. Rowe et al. (2000) provided an illustrative case with 4% normal forms indicating that if the TZI was <1.7 successful fertilization may be expected in vitro without ICSI, with a TZI >1.9 ICSI may well be required in order to achieve fertilization. They further suggested that ‘extreme teratozoospermia’ could be defined as 0% morphologically normal spermatozoa combined with a TZI >1.7, and noted that further studies and more extensive clinical experience would permit better definition of more widely applicable criterion values in the future. Unfortunately, no mention was made as to how these values were obtained, or whether they were based on three (WHO, 1999) or four (WHO, 1992) abnormality classes, and no published work was referenced. As a general comment, authors publishing data on the TZI need to state clearly how the TZI was calculated.

Only the Menkveld and co-workers group have published studies defining cut-off values for TZI: one in vitro study which showed that the ‘acro-some index’ was the best predictor of an IVF rate of >50% (Menkveld et al., 1998; 2007a), and one in vivo study (Menkveld et al., 2001). These studies reported TZI cut-off values of 1.46 and 1.64 respectively [or 2.09 if a 50% prevalence of infertile males in a subfertile population undergoing assisted reproduction technology (ART) is assumed].
Consequently, either the TZI should have been excluded from WHO5, or else included with applicable reference values based on the four defect category TZI. Its calculation will remain an integral component of the ESHRE BSA course.

**Retention of the use of nomenclature terms**

WHO5 retains the use of nomenclature terms such as oligozoospermia. It must be reiterated that these terms simply classify the perceived quality of the semen and do not identify, or even suggest, biological cause or real fertility potential (Eliasson et al., 1970; Eliasson, 1977, 2010; Bostofte et al., 1981), and hence their continued use is not helpful. While there have been numerous debates and publications discussing possible reference values and discussion of such nomenclature (e.g. Eliasson, 1977; Guzick et al., 2001; Björndahl et al., 2010), probably the most useful approach is to provide three interpretation categories: normal, doubtful and pathological or not normal (Eliasson, 1977; Björndahl et al., 2010; also see ‘The delusion of suddenly changed limits between fertile and subfertile men’, below). The ESHRE BSA course discusses in detail the use of reference ranges and nomenclature, with an emphasis on the potential usefulness of the three category interpretation.

**Multiple methods and nonlinear method presentation**

WHO5 still includes multiple methods for performing some of the tests, with poor explanations of their relative merits or otherwise: e.g. determination of low sperm concentrations in semen, alternative stains for sperm morphology assessment (e.g. Diff-Quik™), and the use of eosin without a counter stain for sperm vitality assessment. Some of the methods in WHO5 are presented in an unnecessarily complex manner not amenable to easy use in the laboratory, e.g. determining sperm concentration. Both these issues diminish the manual’s practical usefulness and will impede its adoption.

A good laboratory manual always provides clear, step-by-step protocols for easy implementation and routine use in service laboratories. These protocols include considerations of a method’s limitations and issues affecting its implementation and routine application. Any deviations from these basic principles make it harder for a laboratory to adapt a method into its standard operating procedure manual.

**Inconsistencies and errors**

There are several inconsistencies in WHO5, some of which are actually erroneous. One method particularly affected by this is the determination of sperm vitality using eosin-nigrosin staining:

1. The cut-off to perform a vitality assessment has been changed from >50% immotile spermatozoa (WHO, 1992, 1999) to ‘less than about 40% progressively motile spermatozoa’ (WHO, 2010). The change is illogical since non-progressively motile spermatozoa are clearly still ‘live’.

2. The interpretation criteria for eosin staining has been changed arbitrarily so that ‘light pink heads are considered alive’ (WHO, 2010). This is contrary to papers on eosin exclusion staining for mammalian sperm vitality going back 60 years. The standard criterion is that any degree of pink colouration indicates that a spermatozoon is not ‘live’ (Mortimer, 1994) with the sole, strict, exception of the ‘leaky neck’ staining artefact where faint pink colouration might be seen in the very posterior region of the sperm head (Björndahl et al., 2003, 2004). The revised criterion in WHO5 is clearly wrong, and when applied with the technique described in WHO5, which is the one described by Björndahl et al. (2003), will affect the results obtained.

**Unnecessary extra work**

In WHO5, it is stated that both sperm vitality and sperm morphology assessments must be made in duplicate, evaluating 200 spermatozoa in each replicate ‘in order to achieve an acceptably low sampling error’ (WHO, 2010). These requirements represent substantial extra work for what are unestablished improvements in accuracy and/or precision in the final results. Indeed, Menkveld has previously established the adequacy of a single assessment of sperm morphology on 200 cells from a single slide (Menkveld et al., 1990), and with a binary endpoint such as vitality any possible improvement will be minimal.

In addition, the purportedly improved method for determining low values of sperm concentration represents substantial extra work for an, again unproven, improvement in accuracy or precision and cannot be expected to provide any increase in clinical value from either the diagnostic or prognostic perspective.

For each of these changes the WHO manual should have provided justifications for the substantial extra effort (and hence cost), e.g. statements of reductions in the uncertainty of measurement they would achieve.

**Illogical sperm preparation methods**

WHO5 still allows simple centrifugal washing of spermatozoa for ‘good quality’ semen samples. Unfortunately, one cannot be certain that an ejaculate is free from the attendant risks of reactive oxygen species damage (Aitken and Clarkson, 1987, 1988; Mortimer, 1991) without assessing both sperm morphology for spermatozoa with retained cytoplasm and verifying the absence of peroxidase positive leucocytes. To achieve both of theseobile between completion of semen liquefaction and the need to commence sperm preparation by 30 min post-ejaculation (Mortimer, 2000; Björndahl et al., 2010) is clearly impossible on a routine basis. Also, the WHO5-recommended density gradient method contains numerous errors. Although the method refers to the use of modern silane-coated colloidal silica, it still requires the addition of 10 ml of a 10 x medium to 90 ml of a ‘density-gradient medium’—yet all commercially available silanized colloidal silica sperm preparation products since 1997 are already isotonic. The only colloidal silica product that is not already isotonic is Percoll (which is polyvinyl alcohol-coated silica) and it was banned from clinical use by its manufacturer effective 1 January 1997 (see Mortimer, 2000). WHO5 perpetuates the incorrect colloid layers that have been in the WHO laboratory manual since WHO3, (WHO, 1992), using a 72% colloid-equivalent lower layer, which is too low in density (i.e. <1.1 g/ml). While this will provide an apparently higher yield, it only does so by allowing poorer quality spermatozoa into the pellet (see Mortimer, 2000; Björndahl et al., 2010).

Finally, WHO5 still recommends Ham’s F10 medium for all sperm preparation methods, 15 years after a clear recommendation that it not be used for this purpose due to its iron content (Gomez and Aitken, 1996).
The delusion of suddenly changed limits between fertile and subfertile men

The part of WHO5 that has caught most attention in the field of reproductive medicine is the lowered reference limits calculated from results on semen provided by recent fathers and men in a general population. It appears that there is a common belief that the biology of subfertility has changed as a result of the lowering of the ‘normal/fertile’ reference limits or ranges. There are, however, a number of problems related to the establishment of reference ranges based only on individuals without the disorder, i.e. men who are not subfertile (Björndahl, 2011). Furthermore, since the data were collected during a long period of time, and external quality control had not been implemented in all contributing laboratories (Cooper et al., 2010) the validity of the suggested reference limits can be questioned. Due to the considerable overlap of results from fertile and subfertile men a valid approach would be to identify three zones: (i) ‘normal results’, i.e. a low probability of subfertility and high probability of fertility; (ii) ‘abnormal results’, i.e. a high probability of subfertility and low probability of fertility and (iii) ‘borderline results’, i.e. no clear discrimination between subfertility and fertility (Björndahl, 2010; Björndahl et al., 2010).

Dividing the range of results into these three zones is well-established in andrology (e.g. Eliasson, 1977; Mortimer, 1994), and the material presented in WHO5 provides no evidence that might contradict the validity of this principle.

A further concern regarding the origin of the WHO5 reference values is that the data came from studies on semen samples obtained after 2–7 days of abstinence, as has been advocated in all five editions of the WHO manual. This persistently ignores the fact that MacLeod and Gold (1952) clearly demonstrated that ejaculate volume, and sperm concentration in particular, increase considerably with each day of increasing abstinence: e.g. sperm concentration more than doubled when the abstinence increased from 3 to 10 days. Similar results have been reported by others (e.g. Mortimer et al., 1982). For the purpose of standardization, and especially comparisons between groups, it is therefore of the utmost importance that the prescribed period of abstinence before a semen analysis should be from 3 to 4 days (Menkveld, 2007b; Björndahl et al., 2010).

That the abstinence periods in the source studies for the WHO5 reference values were not so standardized creates further doubt as to the usefulness of the derived reference values.

Usefulness of the ESHRE BSA courses

More than 50 courses have been run since the initiation in Sheffield, 1994. Most courses have been given within Europe (Sweden, Denmark, Finland, Norway, the Netherlands, Belgium, Spain, Portugal, Greece, Poland and Ukraine) but also in other parts of the world (South Africa, Canada). The experiences from Scandinavia, Belgium and the Netherlands have been published (Björndahl and Kvist, 1998; Punjabi et al., 1998; Vreeburg et al., 1998), as well as a comprehensive account for the immediate effects of assessment quality in several different courses (Björndahl et al., 2002). Though courses given in Spanish and Portuguese, propagation of the training course is under way to Latin America. Courses in English are given in Stockholm, Sweden, to train possible course organizers for more regions to establish courses in the local language.

As a follow-up and guidance for laboratories developing and maintaining the quality in the laboratory work, the ESHRE SIGA External Quality Assessment Programme was set up in 1999. There are now more than 50 laboratories regularly taking part in the programme in Europe (Sweden, Norway, Denmark, Finland, the Netherlands, Belgium, Spain, Italy, Greece, Switzerland, Portugal, Czech Republic and Poland) and in the rest of the world (Australia, Israel, South Africa, Colombia, Canada and USA).

Conclusions

In general the authors welcome any improved reference work that can lead to increased standardization, improved accuracy and precision and reduced uncertainty of measurement in diagnostic laboratory andrology. The WHO has certainly significantly contributed to these goals. However, for all the abovementioned reasons, the ESHRE BSA course will in future maintain standards of procedures that are not fully concordant (‘WHO compliant’), with those recommended in the most recent WHO laboratory manual (WHO5). These include:

• Sperm motility: keeping the four class differential motility count allowing distinction between rapid and slow progressive sperm to provide better prognostic information for ART.
• Sperm morphology: maintaining the four-category assessment with calculation of the TZI to improve the diagnostic and prognostic information of the morphology assessment.
• Interpretation: continuing to advocate the use of three categories of results: ‘normal’, ‘borderline’ and ‘abnormal’. In addition, using nomenclature terms to describe semen analysis results (e.g. oligozoospermia) will continue to be eschewed.
• Abstinence: continuing to advocate a prescribed abstinence period of 3–4 days.
• Sperm concentration: simpler and less-prone-to-error method for low sperm concentration assessment.
• Sperm vitality: using correct criteria for live and dead spermatozoa combined with lower total numbers counted since duplicate counting of high numbers of spermatozoa will not improve the precision of the results.
• Sperm preparation: using an optimized discontinuous density gradient method with a lower layer of density 1.1 g/ml in order to separate the fully mature spermatozoa.

The ESHRE BSA course recommendations will provide results with the same or better quality than those recommended in WHO5, often with less demand on time and efforts in the laboratory. Therefore, the handbook A Practical Guide to Basic Laboratory Andrology (Björndahl et al., 2010) will be the reference text for the ESHRE BSA course from here on.

Authors’ roles

All the authors were involved in the conception, analysis, discussion and writing of the manuscript. The authors are listed alphabetically and form the sub-committee for training of SIGA.

Conflict of interest

C.L.R.B. and L.B. were members of the WHO editorial committee for WHO5.
**Funding**

ESHRE provided funding for a meeting of the authors to update the BSA course content.

**References**


Menkveld R, Stander FSH, Kruger TF. Comparison between acrosome index and teratozoospermia index as additional criteria to sperm morphology in the prediction of expected in-vitro fertilisation outcome. Hum Reprod 1998;13:52.


