Asian Journal of Andrology (2010) 12: 14–20

www.nature.com/aja

Perspectives

Practical semen analysis: from A to Z

Charlene Brazil

Center for Health and the Environment, University of California, Davis, CA 95616, USA

Abstract

Accurate semen analysis is critical for decisions about patient care, as well as for studies addressing overall changes in semen quality, contraceptive efficacy and effects of toxicant exposure. The standardization of semen analysis is very difficult for many reasons, including the use of subjective techniques with no standards for comparison, poor technician training, problems with proficiency testing and a reluctance to change techniques. The World Health Organization (WHO) Semen handbook (2010) offers a vastly improved set of standardized procedures, all at a level of detail that will preclude most misinterpretations. However, there is a limit to what can be learned from words and pictures alone. A WHO-produced DVD that offers complete demonstrations of each technique along with quality assurance standards for motility, morphology and concentration assessments would enhance the effectiveness of the manual. However, neither the manual nor a DVD will help unless there is general acknowledgement of the critical need to standardize techniques and rigorously pursue quality control to ensure that laboratories actually perform techniques 'according to WHO' instead of merely reporting that they have done so. Unless improvements are made, patient results will continue to be compromised and comparison between studies and laboratories will have limited merit.

Asian Journal of Andrology (2010) 12: 14–20. doi: 10.1038/aja.2008.51.

Keywords: andrology laboratory, quality assurance, semen analysis, sperm concentration, sperm morphology, sperm motility

1 My perspective

Twenty-eight years ago, I learned how to perform a semen evaluation. 'You could teach a chimp to do this', quipped my instructor. I wonder how he would feel if he knew that so many years later, basic semen evaluation is steeped in controversy. Since those days spent at the bench, I have seen the andrology laboratory from many angles. I supervised a full-service clinical andrology laboratory, taking it through several years of inspections and accreditation. I have coordinated andrology centre activities for several multicentre studies [1–3], which involved standardizing procedures, training technicians, and preparing and distributing proficiency-testing (PT) materials. I have visited andrology laboratories on three continents and have been privileged to work one-on-one with over 100 US and international technicians. I have seen how different laboratories are managed and have learned that some of the most respected facilities in the United States still have problems with standardization. I have come to appreciate that most laboratory directors and technicians are already using the techniques that they think are best, and therefore they may be reluctant to change methods. I have discovered that what the director thinks is going on at the bench is often different from what I see or hear from technicians. These experiences have helped me understand why standardizing semen evaluation will likely continue to pose challenges, even after the publication of the World Health Organization (WHO) 2010 guidelines [4]. Given the problems I have seen and the difficulties inherent in standardizing semen evaluation, I am not surprised that the literature is filled with conflicting data about whether sperm counts are decreasing [5, 6], whether the percentage of spermatozoa with normal morphology predicts *in vitro* fertilization (IVF) success [7, 8], whether concentration, motility or morphology is most important

Correspondence to: Ms. Charlene Brazil, Center for Health and the Environment, University of California, Davis, CA 95616, USA.
Fax: +1-530-752-5300 E-mail: cbrazil@ucdavis.edu
Received: 2 November 2008 Revised: 4 November 2008
Accepted: 11 November 2008

for assessing male fertility [1, 9, 10] or, more generally, whether semen evaluation is of any use at all [11].

2 Will the 2010 WHO guidelines help or not?

Implementation of the suggestions in the WHO manual will pose a substantial challenge: it is extremely difficult to learn a technique as subjective as semen evaluation from words and pictures alone. Another obstacle will be educating scientists about this vastly improved edition and convincing them to use it. The latest edition has the potential to help standardize semen evaluation in a way that previous editions did not. Along with many more instructive photographs, it contains very detailed and clear explanations of all the basic techniques. One of the most important lessons I have learned about technician training is that what sounds clear and straightforward to one person may well be interpreted in very different ways by others. Words and phrases such as 'mix well', 'slow', 'rapid', 'large', 'oval', 'smooth edges' and 'on the line' are almost useless without additional detail. One major improvement in this version is that it contains simplified techniques for concentration determinations, as well as many examples of the equations used to determine the final concentration, which are essential for those technicians with limited mathematical skills. The new manual provides more specific guidance for performing the Strict morphology analysis as well as micrographs of hundreds of Papanicolaou (Pap)-stained spermatozoa with corresponding Strict analysis classification. The section on motility assessment offers a more detailed level of guidance as well.

The manual has been written to serve many different types of end users, including those in state-of-the-art full-service andrology clinics, large and small hospital laboratories, clinical service facilities and independent physician offices. The manual also recognizes that these facilities are located in both developed and developing countries, where availability of resources may differ vastly. In addition, the manual must serve clinicians and research scientists as well as medical technicians. Although each type of laboratory and end user brings unique challenges, a universal problem is that semen evaluation is very subjective and difficult to standardize, in large part because there is no absolute quantitative standard for comparison of values for concentration, percentage motility or percentage normal morphology. Although WHO 2010 [4] offers a globally accessible written standard for how techniques should be performed, in the absence of a quantitative comparison standard, its potential for effecting large-scale change is limited. Technicians wanting to learn the procedures can either follow the techniques in the manual or learn from someone else, but in the end, they have to simply hope that they are getting accurate results. Those who have other technicians to compare with have an advantage, but there is potential for problems there as well, for there is no guarantee that the other technicians or other centres are obtaining accurate results either [12]. Clearly, a technique initially learned poorly can be passed on through multiple laboratories, accumulating new errors along the way.

There are numerous reports on the lack of standardization among technicians performing semen evaluations, but I think the severity of the problem is best exemplified by the errors seen on patient laboratory reports around the United States, for example, a semen evaluation report that showed 20 million spermatozoa per mL, 'all heads without tails' (The man was later diagnosed with congenital absence of the vas), a laboratory that often reports 90% motility, a report with a flagged abnormal motility value—there were too few 'slow' spermatozoa (> 50% rapidly moving spermatozoa), a laboratory whose reference values for concentration marks anything below 60 million spermatozoa per mL as abnormal and another reporting 68% normal morphology by the Tygerberg Strict method. These are obvious problems, as the erroneous results are easily recognized. Even more worrisome are the laboratory reports with errors that are not as glaring, that is, reports that contain values that are within expected ranges but that may be highly inaccurate owing to poor technician training. Probably the most problematic are the reports of < 4% normal morphology, as this is typically a point at which a clinical decision might be made to direct a man down an expensive treatment path, perhaps unnecessarily if the result is inaccurate.

3 Each semen analysis technique poses its own set of challenges

3.1 Semen volume

The method used for volume measurement—whether by a pipette or a cylinder or by weight—has a detectable influence on the obtained value. Many may still be skeptical of the large difference between weighed and pipetted volume measurements [13]. Even I did not expect to see such profound differences before we collected the data for hundreds of semen evaluations [2, 14]. Given that the clinical emphasis seems more appropriately placed on total sperm number rather than solely on sperm concentration, accurate volume assessment is critical. The mean difference between values obtained by pipetting and weighing from 509 subjects in the Study for Future Families was 0.5 mL [2, 14]. For a specimen of volume ~3 mL, that discrepancy yields a 15% difference or error on total sperm numbers from volume measurement alone. Volume measurement is one parameter that could be easily standardized globally.

3.2 Sperm concentration

Without having worked closely with trainees, it is nearly impossible to imagine how much variability there can be in the performance of this technique. The variability comes from several sources: different techniques used for mixing, measuring and aliquotting, as well as simple disagreement regarding what constitutes a spermatozoon. Uncertainties and challenges arise in recognizing spermatozoa with tightly coiled tails or very amorphous heads and odd tails, as well as in learning to ignore cell fragments and seminal debris. Steps as simple as scoring spermatozoa on the grid line become controversial, that is, how one should score a spermatozoon whose head is on the outside line and whose tail points away from the grid *vs*. a spermatozoon with its head on the line but with its tail pointing towards the inside of the grid. During our semen evaluation training sessions, we spend many hours over several days standardizing the sperm-counting technique, typically with one trainer per one or two trainees. One of the activities involves having each trainee count the spermatozoa in one small area of the grid, unaware of the other's results. Even after several days of training, there is sometimes inconsistency in these counts, and occasionally even between trainers. The inconsistency may result from counting 9 or 11 spermatozoa instead of 10, but this can translate into a 10% error in concentration determination. Add this to the previously mentioned potential volume error of 15%; error from those two sources alone can cause a 25% difference in total reported sperm number.

In addition to challenges in enumerating spermatozoa, different counting chambers can cause large differences in concentration measurements [14, 15]. Total sperm number for 509 fertile men was determined in two ways. Volume was assessed by pipette and weight for all specimens and concentration was determined in a 100-μm hemacytometer and a 20-μm MicroCell chamber. The mean total sperm number was 42% higher when using hemacytometer-derived concentrations and weight than it was for the same specimens using MicroCell-derived concentrations and pipetted volumes (299.7×10^6 *vs*. 210.9×10^6 total spermatozoa). This result illustrates how carefully one must interpret study data that suggest there are changes in concentration or total sperm number when different methods for assessing volume and concentration have been used. One must even be careful when comparing studies from different laboratories reporting the same techniques [14].

The WHO manual recommends an easy, improved method for the hemacytometer technique and it is hoped that laboratories will comply with its suggestions. Our laboratory has used fixed-depth chambers for multicentre studies because it was easier to train technicians to use them than to use the more complicated technique outlined in WHO 1992 [16] and WHO 1999 [17] for determining concentration. With the new simplified hemacytometer technique described in WHO 2010 [4], along with the knowledge that disposable chambers yield different results from the hemacytometer, I would now recommend hemacytometers for all future multicentre studies.

3.3 Sperm motility

In my experience, almost without fail, technicians at the beginning of a training session overestimate motility by a large margin. Overestimation seems to be more of a problem with high-concentration specimens. The new manual gives much better guidance on how to avoid over- or underestimating motility, specifically by using the ocular grid, which facilitates the scoring of small sections at a time, allowing for more accurate counting and easier standardization.

3.4 Sperm morphology

I see morphology assessment as the biggest technical challenge for the andrology laboratory. The ease with which one can become proficient at performing Strict method morphology analysis is greatly overstated, as is the confidence that each centre performing this morphology assessment is actually using the same technique [18]. Although the statement 'All the borderline spermatozoa are abnormal' sounds straightforward enough, the reality is that spermatozoa come in a continuous spectrum of sizes and shapes. In my mind, it is no easier to draw the line between perfect and borderline spermatozoa than it is to do so between borderline and abnormal spermatozoa. If scrutinized closely enough, no spermatozoon truly appears 'perfect', a concept expressed indirectly by Mortimer and Menkveld [19] in their thorough morphology review article: 'All people assessing human sperm morphology must be aware of the risk of becoming too strict'.

The reason that this vast morphology problem has not received the attention it deserves is twofold. When a publication or a laboratory report states that morphology is assessed by the Strict method, it is easy to assume that there is a single well-established, validated method for strict morphology, but this is simply not true. The desired length for a Pap-stained normal sperm has changed: 3–5 μm [20], 4–5.5 μm [21] or 4–5 μm [19]. Similar discrepancies are reported for head width, length-to-width ratio, and the allowable size of the cytoplasmic droplet. This is noteworthy because there has been an emphasis on how important it is to measure the borderline sperm to distinguish normal from abnormal [21, 22]. It also seems that the expected range of the percentage of normal spermatozoa when applying the technique has changed; that is, it appears that Strict has become 'really strict'. The original article typically cited for the Strict procedure,

by Kruger *et al.* [7], provided the first suggestion of 15% normal forms as a point below which fertility may be decreased. The authors described four groups of men whose wives were undergoing IVF and whose sperm morphologies ranged between 0% and 60% normal forms. The other commonly referenced article on the technique, by Menkveld *et al*. [20], showed a range of 3%–41% normal morphology (mean 16.7%) in samples from a group of 106 men. In studies performed 10–15 years later, in which Kruger (TK) [23] and Menkveld (RM) [24] personally analysed slides for fertile male studies, the use of the same technique generated a range of percentage normal spermatozoa nearly 50% lower. Smears from 61 fertile males, analysed by TK [23], contained a range of only 2%–30% (mean 14.9%) normal forms, and slides analysed by RM from a group of 107 fertile men [24] showed a range of only 1%–19% (mean 6.5%). These percentage normal ranges from fertile men, along with the instruction and guidance presented during Kruger-led training sessions I attended in 1998 and 2003, suggest that there was a change in the technique, and that it became stricter over time. It is quite possible that a technician who learned the technique in the 1990s (or learned it from someone who had learned it at that time) may well be applying the Strict technique on the basis of a different set of guidelines and a different expected percentage-normal range than one who learned the technique much later.

Also masking the morphology analysis problem is the fact that the proficiency testing (PT) programmes for morphology assessment are so forgiving. From the PT schemes with which I am familiar, it is extremely difficult to fail a Strict morphology challenge. In the United States, the American Association of Bioanalysts (AAB)—the major supplier of PT challenges for andrology laboratories—follows federal government guidelines and the recommendation of a panel of scientific advisors for the grading of PT results: every result falling in the range of the mean ± 3SD is graded as acceptable [25]. For a recent AAB testing challenge, technicians using the Strict technique and Spermac stain passed if they scored the test smear anywhere between 0% normal and 28% normal [25].

In Menkveld's description of the Strict method, he refers to the 'holistic' approach of Strict analysis, emphasizing proper smear preparation and Pap staining as being important for appropriate visualization of the sperm head for analysis [19]. The experience of our unit, as a central morphology assessor for several large multicentre studies, confirms this concern. Our laboratory has analysed more than 50 000 morphology smears from over 50 research laboratories [1–3, 26], and we have found that both smear preparation and Pap staining "according to WHO guidelines" are very poorly controlled. Because of the poorly stained slides we received in earlier studies, for subsequent projects we switched to having all centres ship unstained smears to us at regular intervals; these smears were then fixed, stained and read centrally. I would recommend this for all multicentre studies. The new manual contains very detailed and explicit instructions on how to prepare and stain a semen smear properly in order to optimize the sperm images for analysis.

4 Quality control (QC)/quality assurance (QA)

There is controversy over whether QA in the andrology laboratory is really necessary. Jequier [27] notes, 'semen analysis is for the most part carried out satisfactorily all over the world'. I can see how someone associated with a state-of-the-art andrology facility with competent technicians might come to that conclusion, but I do not think it is accurate. I think frustration over QA also arises because many of the government regulations seem to motivate more attention and focus on passing inspections than on improving the actual technical output. It is very possible for a laboratory to pass all US federal inspections easily, yet still be providing poor-quality patient semen evaluation results. I was recently consulted about the use of a counting chamber by a large accredited assisted reproduction laboratory with a PhD director. On reviewing the counting procedure in use, I found that the calculations had been off by a factor of two, and for the previous year, all patient results for sperm concentrations had been erroneously reported at half of their correct values.

I think if one looks carefully at PT results, one will find evidence that QA is indeed necessary and semen analysis is not carried out satisfactorily in many centres. In 2008, approximately 573 US laboratories participated in the AAB PT. The majority of the participating centres were reproductive or andrology laboratories. The results are not reassuring. Over the last 1.5 years, five diluted samples were sent out to participating centres for PT of the determination of sperm concentration. The results from each of these five samples showed such large variation that the participants were not graded, meaning that no matter what concentration the technicians reported, they did not fail. Morphology results were equally disturbing. The mean results were very different according to whether the technician reported that he or she was using WHO 1999 [17] or the Strict method, which is interesting given that the Strict method is the morphology technique described in WHO 1999 [17].

Consider the results from the AAB PT distribution for a test slide (first shipment, 2008) in which 143 laboratories reported that they used the Strict technique and the Diff Quik stain. The mean value for all participants was 8.9% normal forms and the passing range, by definition,

was the mean ± 3SD, which in this case was 0%–19% normal. That means that the technician who evaluated the morphology smear as 0% normal passed, as did the one who evaluated the slide as 19% normal. Other morphology analysis methods were no better. For laboratories using WHO 1992 [16], the passing range was 5%–70% on the same slide. Yet, somehow all this continues to satisfy federal requirements and, apparently, some scientists, clinicians and laboratory directors as well.

In the absence of a global standard, PT is potentially a very useful tool. One advantage of a PT programme is that each of the participants receives a report of the data, and even in the absence of grading and participant consensus, they can still see where their results fall with respect to other centres tested. Although PT may be a requirement and may offer useful guidance to a laboratory, the PT programmes are in need of global standardization, as shown in a study by Cooper *et al.* [12], in which three external PT providers seemed to use conflicting standards.

5 Strategies for laboratories to optimize the predictive value of semen analysis

5.1 Technician selection

I think the single most important thing a laboratory can do to ensure the quality of its semen analysis is to choose an appropriate technician. The necessity for choosing bright, motivated technicians with natural abilities as observers, an idea suggested nearly 20 years ago [28], is crucial and underestimated. I have trained many technicians who had little or no relevant experience and have also standardized techniques for some who had many years of experience. My conclusion is that experience does not matter nearly as much as having a motivated individual with an analytical mind, manual dexterity and a good grasp of simple mathematical concepts (percentages, means, *etc*.).

5.2 Technician training

Once hired, technicians should undergo a thorough, extensive training over many days using a wide range of semen samples. Training workshops can be very useful [29, 30], but unfortunately their availability is limited and most laboratories do not have access to, or the resources to attend, such workshops. If workshops are unavailable, technicians should read very carefully and follow WHO 2010 [4] for learning techniques. After initial training, technicians should practise many semen evaluations before evaluating those from patients. Ideally, these practice evaluations would be performed on semen previously evaluated by a well-trained technician, to allow for comparisons. WHO 2010 [4] offers guidance as to what technicians can do to monitor their own performance and the stability of their results over time.

5.3 Quality control

Good technicians will double-check everything they do, catching most of their own errors, which is the best QA a laboratory can have. Units with more than one technician should set up their own blind comparisons; that is, the technicians should be unaware of each other's values during scoring concentration, motility and morphology assessment, and then they should work together to understand any discrepancies identified. This type of comparison is very helpful; in my experience during training, technicians nearly always match closely when they are openly comparing values side by side at the bench. Discrepancies are much more likely to be identified during blind comparisons, which then allows the trainer to work with the technician to understand the differences. Laboratories with only a single technician, and hospitals or clinical centres, especially those evaluating semen infrequently, face the biggest challenge. Given the minimal requirement for semen analysis training in some medical technology programmes, it is possible that neither a technician nor the director of a clinical laboratory would be aware of inaccurate semen evaluation results, especially as the technicians would almost certainly pass their PT programmes as discussed above. The directors of these centres must be educated about the problems associated with semen evaluation and encouraged to seek further training for their technicians as necessary.

6 A training DVD is needed

Despite vast overall improvements, the WHO manual clearly has its limitations. I am convinced that the quality of semen analysis on a global level could be improved if the WHO were to produce a training DVD to accompany the manual. Ideally, a person learning a new technique should have a written procedure, a visual demonstration, hands-on practice and eventually the opportunity to compare their values with a standard value as a final guarantee of mastery of the technique. It is very difficult, and perhaps impossible, to give enough unambiguous detail in a written protocol to allow mastery of procedures as subjective as the assessment of concentration, motility or morphology in particular. Therefore, I think the DVD should demonstrate in detail all of the basic techniques of a semen evaluation and, in addition, should include a QA section with video clips for assessing motility, morphology and even concentration, as this could be used as guidance for both training purposes and for ongoing QA.

Ideally, this training DVD should be available online as well and it should contain multiple fields from several video-recorded semen samples, representing a range of

motilities. The 'motility target value' could be a consensus value, set from a few well-controlled laboratories throughout the world. The DVD should also contain a large number of images of stained sperm from different semen smears together with a key for Strict assessment. Setting the key for Strict has been controversial, as some would argue it should be set by the originators of the technique (as in WHO 2010) and others would suggest it would be more appropriate to set the reference values through a consensus of those practising the technique.

Images of sections of hemacytometer grids containing spermatozoa for counting would allow technicians to compare results with the key, which again could be set by consensus of a small group. The key could indicate which objects or spermatozoa should or should not be counted, as the new manual has now done, but with many more examples, as well as demonstrate how these values are used to determine the final concentration. Ideally, new standard images would be added annually, allowing technicians to check themselves against the standards on a regular basis.

7 Summary

Although WHO 2010 is vastly improved over previous editions, how much the new manual will help to standardize semen evaluation remains to be seen. Clearly, there is a lack of motivation to standardize to WHO-recommended methods, and in some cases there may be a trend away from them. In the United States, for example, the AAB PT program of 2008 showed that only 127 of 553 laboratories (23%) reported using a hemacytometer chamber for sperm counting, down from 53% as reported by Keel *et al*. [31]. In 2008, only 10/426 (2%) of laboratories reporting morphology results had used the combination of Pap morphology stain and the Strict analysis technique as recommended by WHO 1999 [17], although more than 50% reported using the Strict with other stains. A lack of use of WHO methods is also widespread in the United Kingdom [32].

It is essential that all physicians involved in patient care relating to fertility evaluations be informed of and understand the problems associated with semen analysis and reporting. They must ensure that the laboratories that they use for semen evaluation are well-controlled and use appropriate WHO-recommended methods. They need to put pressure on laboratory directors to change or update procedures if necessary, or switch to laboratories that have higher standards. Directors have the responsibility to be sure that the semen evaluations from their units are as accurate as possible. Hospital and clinical service laboratory directors should be educated such that they are aware that additional training of technicians for semen evaluation may be necessary.

Even if the WHO eventually offers a training/standardization DVD, simply making it available will not solve the problems of standardizing semen evaluation. Laboratory directors, supervisors and technicians must be motivated to use it, and be willing to change their methods if necessary. Change requires time and effort. It is complicated for referring physicians when a clinical laboratory makes changes, as the clinician would have become familiar with the usual ranges and uses them for clinical judgements. It is also difficult to make changes in research settings, as there is always a desire to compare current studies with older studies and to maintain continuity of techniques. However, when properly motivated, it is indeed possible to make substantial changes [33]. One possible scenario would be for the WHO to invest in the education of health-care professionals regarding the problems with semen evaluation. A target date could be set by which time all laboratories would be encouraged to adopt WHO techniques and comparison standards in a concerted effort towards global standardization.

References

1 Guzick DS, Overstreet JW, Factor-Litvak P, Brazil CK, Nakajima ST, *et al.* Sperm morphology, motility, and concentration in fertile and infertile men. N Engl J Med 2001; 345: 1388–93.
2 Swan SH, Brazil C, Drobnis EZ, Liu F, Kruse RL, *et al*. Geographic differences in semen quality of fertile U.S. males. Environ Health Perspect 2003; 111: 414–20.
3 Overstreet JW, Fuh VL, Gould J, Howards SS, Lieber MM, *et al*. Chronic treatment with finasteride daily does not affect spermatogenesis or semen production in young men. J Urol 1999; 162: 1295–300.
4 World Health Organization. WHO Laboratory Manual for the Examination and Processing of Human Semen, 5th ed. Geneva: World Health Organization; 2010.
5 Fisch H. Declining worldwide sperm counts: disproving a myth. Urol Clin North Am 2008; 35: 137–46.
6 Andersson AM, Jørgensen N, Main KM, Toppari J, Rajpert-De Meyts E, *et al.* Adverse trends in male reproductive health: we may have reached a crucial 'tipping point'. Int J Androl 2008; 31: 74–80.
7 Kruger TF, Menkveld R, Stander FS, Lombard CJ, Van der Merwe JP, *et al*. Sperm morphologic features as a prognostic factor in *in vitro* fertilization. Fertil Steril 1986; 46: 1118–23.
8 Morgentaler A, Fung MY, Harris DH, Powers RD, Alper MM. Sperm morphology and in vitro fertilization outcome: a direct comparison of World Health Organization and strict criteria methodologies. Fertil Steril 1995; 64: 1177–82.
9 Nallella KP, Sharma RK, Aziz N, Agarwal A. Significance of sperm characteristics in the evaluation of male infertility. Fertil Steril 2006; 85: 629–34.
10 Bonde JP, Ernst E, Jensen TK, Hjollund NH, Kolstad H, *et al*. Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. Lancet 1998; 352:

1172–7.
11 Lewis S. Is sperm evaluation useful in predicting human fertility? Reproduction 2007; 134: 31–40.
12 Cooper TG, Björndahl L, Vreeburg J, Nieschlag E. Semen analysis and external quality control schemes for semen analysis need global standardization. Int J Androl 2002; 25: 306–11.
13 Cooper TG, Brazil C, Swan SH, Overstreet JW. Ejaculate volume is seriously underestimated when semen is pipetted or decanted into cylinders from the collection vessel. J Androl 2007; 28: 1–4.
14 Brazil C, Swan SH, Drobnis EZ, Liu F, Wang C, *et al*; Study for Future Families Research Group. Standardized methods for semen evaluation in a multicenter research study. J Androl 2004; 25: 635–44.
15 Tomlinson M, Turner J, Powell G, Sakkas D. One-step disposable chambers for sperm concentration and motility assessment: how do they compare with the World Health Organization's recommended methods? Hum Reprod 2001; 16: 121–4.
16 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction, 3rd edn. Cambridge: Cambridge University Press; 1992.
17 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction, 4th edn. Cambridge: Cambridge University Press; 1999.
18 Ombelet W, Bosmans E, Janssen M, Cox A, Maes M, *et al*. Multicenter study on reproducibility of sperm morphology assessments. Arch Androl 1998; 41: 103–14.
19 Mortimer D, Menkveld R. Sperm morphology assessment-historical perspectives and current opinions. J Androl 2001; 22: 192–205.
20 Menkveld R, Stander FS, Kotze TJ, Kruger TF, van Zyl JA. The evaluation of morphological characteristics of human spermatozoa according to stricter criteria. Hum Reprod 1990; 5: 586–92.
21 Kruger TF. Strict criteria for sperm morphology. In: Lipshultz L, Howards S, editors. Infertility in the Male. St. Louis: Mosby; 1997. p487–90.
22 Kruger TF, Acosta AA, Simmons KF, Swanson RJ, Matta JF, *et al*. Predictive value of abnormal sperm morphology in *in vitro* fertilization. Fertil Steril 1988; 49: 112–7.
23 Gunalp S, Onculoglu C, Gurgan T, Kruger T, Lombard CJ. A study of semen parameters with emphasis on sperm morphology in a fertile population: an attempt to develop clinical thresholds. Hum Reprod 2001; 16: 110–4.
24 Menkveld R, Wong WY, Lombard CJ, Wetzels AM, Thomas CM, *et al*. Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort towards standardization of *in-vivo* thresholds. Hum Reprod 2001; 16: 1165–71.
25 American Association of Bioanalysts. Available from: http://www.aab.org.
26 Hellstrom WJ, Overstreet JW, Yu A, Saikali K, Shen W, *et al*. Tadalafil has no detrimental effect on human spermatogenesis or reproductive hormones. J Urol 2003; 170: 887–91.
27 Jequier AM. Is quality assurance in semen analysis still really necessary? A clinician's viewpoint. Hum Reprod 2005; 20: 2039–42.
28 Dunphy BC, Kay R, Barratt CL, Cooke ID. Quality control during the conventional analysis of semen, an essential exercise. J Androl 1989; 10: 378–85.
29 Björndahl L, Barratt C, Fraser LR, Kvist U, Mortimer D. ESHRE basic semen analysis courses 1995–1999: immediate beneficial effects of standardized training. Hum Reprod 2002; 17: 1299–305.
30 Brazil C, Swan SH, Tollner CR, Treece C, Drobnis EZ, *et al*. Quality control of laboratory methods for semen evaluation in a multicenter research study. J Androl 2004; 25: 645–56.
31 Keel BA, Quinn P, Schmidt CF Jr, Serafy NT Jr, Serafy NT Sr, *et al*. Results of the American Association of Bioanalysts national proficiency testing programme in andrology. Hum Reprod 2000; 15: 680–6.
32 Riddell D, Pacey A, Whittington K. Lack of compliance by UK andrology laboratories with World Health Organization; recommendations for sperm morphology assessment. Hum Reprod 2005; 12: 3441–5.
33 Cooper TG, Atkinson AD, Nieschlag E. Experience with external quality control in spermatology. Hum Reprod 1999; 14: 765–9.