

## OPINION

# Are sperm counts declining? Or did we just change our spectacles?

Allan A Pacey

*Asian Journal of Andrology* (2013) 15, 187–190; doi:10.1038/aja.2012.165; published online 28 January 2013

**T**he popular and scientific debates about a possible decline in semen quality over the past decades are largely based on retrospective analyses of semen analysis data performed in the past. This article will argue that the conclusions from such analyses are significantly weakened because the methods of laboratory andrology have changed considerably since the 1950s. In the last 20–30 years, there have been significant developments in training and competence, increased emphasis on Quality Assurance (QA) and Quality Control (QC) as well as a major attempt at standardisation of technique through five successive editions of the World Health Organization (WHO) laboratory manuals. Interestingly, the only large prospective study carried out to date shows no change in sperm concentration over 15 years, being consistent with the idea that when laboratory methods are adequately controlled, no secular change in sperm counts are observed.

In 1954, a High Court judge in the Court of Appeal of England and Wales made a landmark ruling that has become an important legal principle in English Tort law. It related to a case brought against the Minister of Health by two men who, 7 years earlier, had suffered paraplegia as a consequence of being given anaesthetic by lumbar puncture before a surgical procedure. While the precise details of the case are not really relevant to the topic of this paper, it is important to know that, unbeknown to the anaesthetists, the vials of anaesthetic had been contaminated in a manner that had not previously occurred, and could

not be detected, before it was administered. Following the incident, clinical practice was immediately changed to prevent further occurrences. But, perhaps understandably, the two men sued for negligence and the matter was resolved in the courts. In his judgement, the Judge Lord Denning dismissed the men's claim saying, 'We must not look at the 1947 incident with 1954 spectacles'.<sup>1</sup> In effect, he was pointing out that we could not always view things that happened in the past with modern eyes. While this is now an important principle for lawyers, this paper will argue why it is also an important principle for interpreting whether or not we can establish if semen quality has been declining in recent years.

When Elizabeth Carlsen and colleagues<sup>2</sup> published their review of data from 61 papers published between 1938 and 1991, their regression analysis relied on the assumption that, during the 53 years in question, andrology laboratory methods had been wholly comparable and had remained unchanged. While they understandably excluded papers where sperm counting had been performed by computer-assisted methods or by flow cytometry (these methods were not widely used and were introduced only in later years), they recognized the apparent imprecision of counting sperm by other methods. However, they concluded that 'there is no reason to believe that this test in itself has been subject to secular trend' and in support of this argument, they cited that 'the same types of counting chambers have been used for the past 50 years by haematologists, who have not reported a similar secular trend in blood cell counts.' The issue of how accurate haematological methods may or may not have been in the past will be dealt with later in this paper. But it is fair to conclude that these arguments did not completely resonate with many at the andrology lab coalface. Their concern was

highlighted only weeks later in the *British Medical Journal*,<sup>3</sup> where Carlsen *et al.* were urged to 'establish that their comparison of historical data is free of methodological bias'. To this day, this is a criticism that remains unresolved, both for the Carlsen *et al.*'s paper<sup>2</sup> and those which have followed it (see Fisch<sup>4</sup> for review). This opinion paper will argue that even with today's relatively well-standardized laboratory methods to assess semen quality, with well-established training programmes for laboratory staff, increased emphasis on laboratory accreditation and comprehensive internal and external quality assurance programmes in place, we are still far from generating consistent error-free data for semen analysis. As a consequence, it is very hard to look back into the archives with any sense of confidence about the precision and reliability of measurements made in the past. In effect, we are wearing the wrong spectacles.

To set the scene, there are three areas of current and historical laboratory practice that need to be examined before the main arguments can be set out: (i) the development of semen analysis methodology; (ii) the selection and implementation of laboratory tests; and (iii) Quality Assurance (QA) and Quality Control (QC). Each will now be discussed in turn.

### DEVELOPMENT OF SEMEN ANALYSIS METHODOLOGY

Briefly, modern semen analysis can be traced back to a 1956 paper published in *Fertility and Sterility* by John MacLeod<sup>5</sup> (although by modern standards, the details are surprisingly vague). Therefore, in the context of the Carlsen *et al.*'s analysis,<sup>2</sup> it is clear that the first 18 years of data (10 studies out of the 61) were potentially obtained using laboratory methodology that was, at best, poorly described.

Academic Unit of Reproductive and Developmental Medicine, Department of Human Metabolism, University of Sheffield, Sheffield S10 2SF, UK

Correspondence: Dr AA Pacey (A.Pacey@Sheffield.ac.uk)  
Received: 12 December 2012; Revised: 18 December 2012; Accepted: 18 December 2012; Published online: 28 January 2013

Even then, it took a further 24 years for any level of international standardisation to be agreed with the publication of the first World Health Organization (WHO) laboratory manual for the examination of human semen and sperm–cervical mucus interaction in 1980.<sup>6</sup> While the publication of the WHO manual was obviously a major step forward, it was revised again in 1987<sup>7</sup> and then again in 1992.<sup>8</sup> While the latter revision obviously fell outside the period of data collection for the Carlsen *et al.* 1992 paper,<sup>2</sup> these developments may have had an impact on some of the studies published later. Moreover, it illustrates that the international andrology community was not yet satisfied that the methods of semen analysis were sufficiently well established. Indeed, this is further exemplified by the fact that more revisions were published in 1999<sup>9</sup> and 2010<sup>10</sup> and methodology has continued to be refined.

#### IMPLEMENTATION OF SEMEN ANALYSIS METHODOLOGY

If it were possible to accept that relatively standardized sperm counting methods were used throughout the period of time covered by Carlsen *et al.*,<sup>2</sup> and subsequent studies in which archival data for semen analysis was examined, then a second question to be asked would be how effectively these laboratory procedures were implemented?

First, it is now well recognized that semen, as a fluid, is not as straightforward to deal with as blood or urine and there are difficulties encountered by accurately dispensing aliquots of a viscous non-homogeneous fluid for dilution and counting. As far back as 1984, it was established that using a haemocytometer with a white blood cell pipette (used by haematologists) was less precise than a Makler chamber which in turn was not as accurate as a haemocytometer used with a tuberculin syringe.<sup>11</sup> This is interesting given Carlsen's confidence about haematological methods being robust (see above), but also has implications for the assumption (of them and others) that all counting chambers are equal. It has long been established that the Makler and Horwell chambers gave results that were 1.5 and 2.7 times higher than those obtained by using the haemocytometer.<sup>12</sup> Furthermore, that the Makler Chamber can give relatively poor precision in comparison to the haemocytometer when both were compared to data obtained from flow cytometry.<sup>13</sup> Therefore, if over time laboratories slowly moved toward using the haemocytometer (as the successive editions of the WHO manuals<sup>6–10</sup> have encouraged), then

the only inevitable consequence of such secular changes in laboratory method is that the apparent mean values for sperm concentration would have declined—entirely consistent with what has been observed.

Even if all laboratories had only ever used the haemocytometer to determine sperm concentration, an overestimation of sperm concentration could also occur because of counting error or misidentification of sperm through inexperience or lack of training. Data from training programmes<sup>14,15</sup> show that the results from trainees become less variable and closer to the target value over the duration of the course. It is noteworthy that such training courses are a relatively recent development in the context of the sperm count debate. For example, the courses established by the European Society for Human Reproduction and Embryology<sup>14</sup> first took place in 1994 and those established by the Tygerberg hospital in South Africa 2001.<sup>15</sup> Before this time, it is unclear how training was provided in many parts of the world, if at all.

#### QA AND QC

While we now recognize that QA and QC are critical elements of any laboratory testing process,<sup>16</sup> it is fairly clear that laboratory andrology was somewhat late to embrace the concept in comparison to other disciplines.<sup>17,18</sup> For example, in the United Kingdom, External Quality Assurance Schemes were established in Clinical Chemistry, Cytology, Endocrinology, Haematology, Histopathology, Microbiology, Pharmacology and Toxicology by the late 1970s,<sup>19</sup> whereas a similar scheme in laboratory andrology was not established until 1994.<sup>20</sup> So when Carlsen *et al.*<sup>2</sup> assured us that sperm counting using haematological methods were robust because no secular change in blood cell counts had been seen, they failed to acknowledge that haematology measurements were already well controlled by their own QA and QC programmes at a time when andrology laboratories were not. Interestingly, the first three versions of the WHO laboratory manuals<sup>6–8</sup> made almost no reference at all to the need for QA and QC of semen analysis.

The need for QA and QC programmes in laboratory andrology was brought into sharp focus in the early 1990s by a series of publications,<sup>20–22</sup> demonstrating that significant disagreement could occur between the results generated from analytical laboratories, often when analysing the same sample. Therefore, it was a significant step forward for the fourth edition of the WHO manual published in 1999<sup>9</sup> to provide detailed guidelines for QA and QC for laboratory andrology for the first

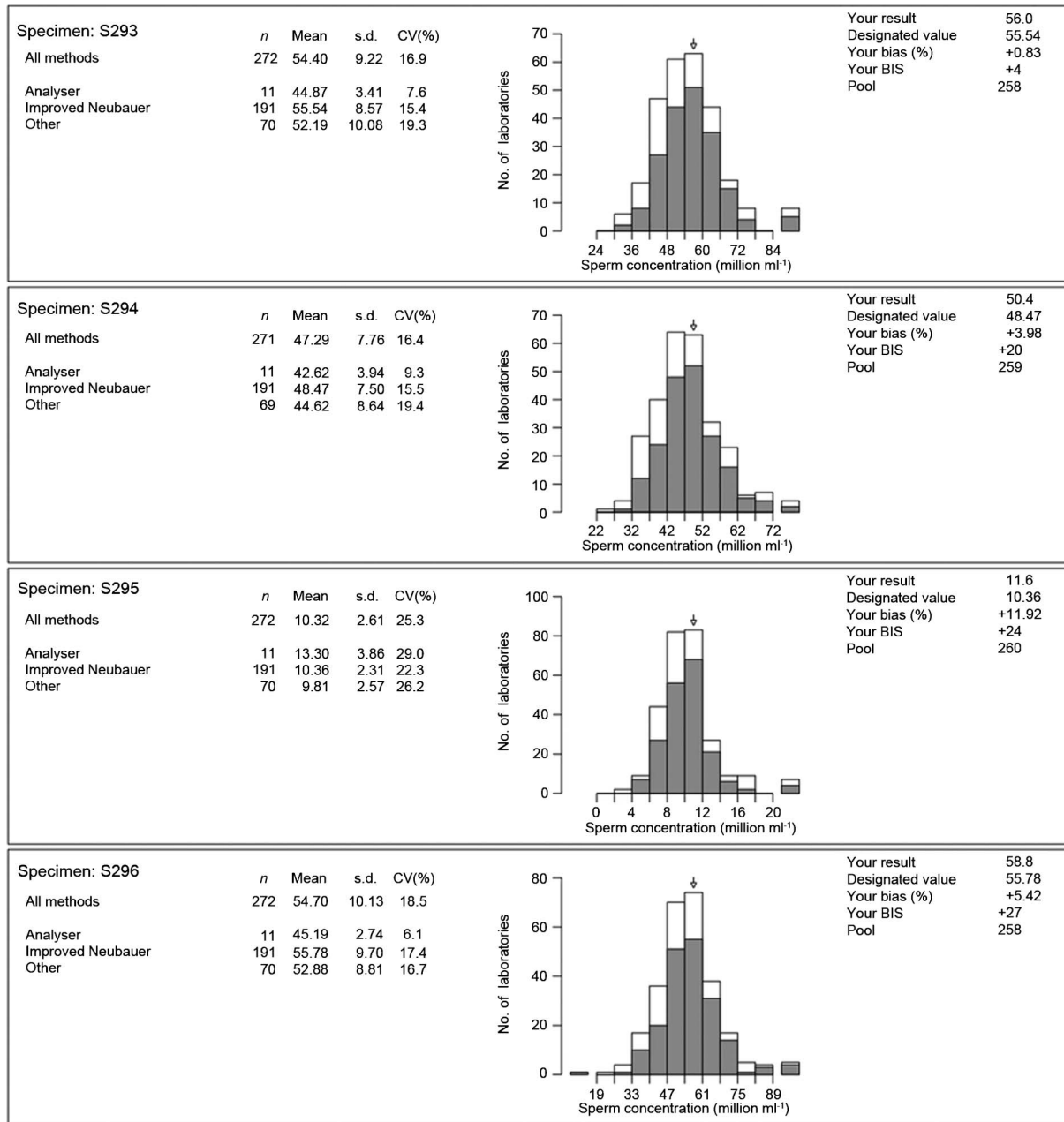
time. Critically, for the present discussion, this means that probably none of the studies included by Carlsen *et al.*'s analysis were conducted with what we would now consider to be acceptable QA and QC. Interestingly, the increased emphasis placed on QA and QC procedures by the fourth edition of the WHO manual<sup>9</sup> was not welcomed by everyone and in 2005 and 2006, this was fiercely debated in the pages of *Human Reproduction*.<sup>17,23–25</sup> However, QA and QC are now central pillars of International Organization for Standards-based Quality Systems<sup>26</sup> as well as national and regional legislation such as the European Commission on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells.<sup>27</sup> In comparison to even 10 years ago, let alone 50 years ago, andrology laboratories are more likely to have such quality systems in place, or are actively working towards them.

#### LABORATORY ANDROLOGY IN 2012

In 2012, a randomly selected andrology laboratory anywhere in the world is more likely to have appropriately trained staff, be following WHO methods, and have effective QA and QC procedures in place compared to an equivalent (or even the same) laboratory only 10 years ago. It is not possible to say what the relative impact of each will have been on the accuracy of the results generated, but it is useful to consider the relative performance of modern andrology laboratories with all of these factors in place.

**Figure 1** shows the results obtained for four specimens distributed by the UK External Quality Assurance Scheme for Andrology (UK NEQAS, St Mary's Hospital, Manchester, UK) in the autumn of 2012. What is interesting about these data is that: (i) of the 272 enrolled laboratories only 70% ( $n=191$ ) were counting spermatozoa by using a haemocytometer (i.e., following WHO methods); and (ii) there remains considerable variation in the results obtained from the participating laboratories. For example, specimen S296 shows a range of results from below 19 to above 90 million per ml! The inevitable conclusion has to be that even in 2012, accurate and precise estimates of sperm concentration are difficult to achieve reliably.

While such data have been shown before,<sup>17,28</sup> it is of concern that few outside the andrology laboratory see the significance for the interpretation of retrospective datasets that have contributed to the decline in sperm



**Figure 1** The results obtained from UK NEQAS for four External Quality Assurance specimens for sperm concentration distributed to 271 laboratories around the world. The arrow on each histogram shows the result obtained by the author's laboratory alongside the results for all other participants using the improved Neubauer haemocytometer (grey bars) as well as those using other methods (white bars). UK NEQAS, UK External Quality Assurance Scheme for Andrology. BIS, Bias Index Score; CV, Coefficient of variation.

counts debate. In the recent paper describing an apparent decline in sperm concentration in France from 1989 to 2005,<sup>29</sup> the authors state in the discussion that, regarding the methods used to measure sperm concentration, 'experts have confirmed that the methods have not changed noticeably during the study period', yet they do not elaborate on what methods were used. Furthermore, with regard to QC, they also state that 'there is no reason to think that they did not follow national and WHO recommendations', yet

as was discussed earlier in this paper, the very emphasis on QC by the various WHO manuals has developed significantly since 1999 (10 years after the beginning of the French dataset).

Our only hope in untangling the debate is to establish prospective studies that have standardized methodology and adequate QA and QC procedures in place from the very start. Interestingly, the first of such studies was published controversially<sup>30</sup> in late 2011 and showed that across 5000 military draftees

who provided semen samples between 1996 and 2010,<sup>31</sup> there had been no obvious change in sperm counts. While one explanation may be that the decline had stopped before this study commenced, to those of us working in laboratory andrology, it feels too much of a coincidence that when a prospective study was performed, no change was found. It is clear that more studies of this type are needed, and interestingly, guidelines have now been published to assist in the design and interpretation of studies involving semen quality.<sup>32</sup>

## CONCLUSION

It may never be known to what extent changes in laboratory methods have had on semen analyses performed in the past, but the evidence is sufficiently clear to suggest that we should be cautious about drawing conclusions from historical data. To go back to Lord Denning's 1954 statement,<sup>1</sup> it is a mistake even to try and view the past with modern spectacles as the necessary detail will always be out of focus. As an alternative, perhaps laser eye surgery is needed, so we can move forward without spectacles and design adequate prospective studies that might answer the question once and for all.

## COMPETING FINANCIAL INTERESTS

The author declares that there are no competing financial interests.

## ACKNOWLEDGMENTS

The author would like to thank Deborah Saxton and Katrina Williams for comments on the draft article.

- 1 Roe v Minister of Health and Anr. *Court of Appeal*. 1954; 2QB66.
- 2 Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. *BMJ* 1992; **305**: 609–13.
- 3 Tummon IS, Mortimer D. Decreasing quality of semen. *BMJ* 1992; **305**: 1228–9.
- 4 Fisch H. Declining worldwide sperm counts: disproving a myth. *Urol Clin North Am* 2008; **35**: 137–46.
- 5 MacLeod J. Human semen. *Fertil Steril* 1956; **7**: 368–86.
- 6 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction. 1st ed. Singapore: Press Concern; 1980.
- 7 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction 2nd ed. Cambridge: Cambridge University Press; 1987. p67.
- 8 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction. 3rd ed. Cambridge: Cambridge University Press; 1992. p107.
- 9 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction 4th ed. Cambridge: Cambridge University Press; 1999. p128.
- 10 World Health Organization. WHO Laboratory Manual for the Examination and Processing of Human Semen. 5th ed. Geneva: World Health Organization; 2010. p271.
- 11 Menkveld R, Van Zyl JA, Kotze TJ. A statistical comparison of three methods for the counting of human spermatozoa. *Andrologia* 1984; **16**: 554–8.
- 12 Imade GE, Towobola OA, Sagay AS, Otubu JA. Discrepancies in sperm count using improved Neubauer, Makler, and Horwells counting chambers. *Arch Androl* 1993; **31**: 17–22.
- 13 Christensen P, Stryhn H, Hansen C. Discrepancies in the determination of sperm concentration using Bürker–Türk, Thoma and Makler counting chambers. *Theriogenology* 2005; **63**: 992–1003.
- 14 Björndahl L, Barratt CL, Fraser LR, Kvist U, Mortimer D. ESHRE basic semen analysis courses 1995–1999: immediate beneficial of standardized training. *Hum Reprod* 2002; **17**: 1299–305.
- 15 Franken DR, Aneck-Hahn N, Lombaard C, Kruger TF. Semenology training programs: 8 years' experience. *Fertil Steril* 2010; **94**: 2615–9.
- 16 Burnett D. Understanding Accreditation in Laboratory Medicine. London: ACB Venture Publications; 1996. p312.
- 17 Pacey AA. Is quality assurance in semen analysis still really necessary? A view from the Andrology Laboratory. *Hum Reprod* 2006; **21**: 1105–9.
- 18 Pacey AA. Quality assurance and quality control in the andrology laboratory. *Asian J Androl* 2010; **12**: 21–5.
- 19 Whitehead TP, Woodford FP. External quality assessment of clinical laboratories in the United Kingdom. *J Clin Pathol* 1981; **34**: 947–57.
- 20 Cooper TG, Atkinson A, Nieschlag E. Experience with external quality control in spermatology. *Hum Reprod* 1999; **14**: 765–9.
- 21 Neuwinger J, Beijre H, Nieschlag E. External Quality Control in the andrology laboratory: an experimental multicenter trial. *Fertil Steril* 1990; **54**: 308–14.
- 22 Matson PL. External quality assessment for semen analysis and sperm antibody detection: results of a pilot scheme. *Hum Reprod* 1995; **10**: 620–5.
- 23 Jequier AM. Is quality assurance in semen analysis still really necessary? A clinician's viewpoint. *Hum Reprod* 2005; **20**: 2039–42.
- 24 Holt WV. Is quality assurance in semen analysis still really necessary? A spermatologist's viewpoint. *Hum Reprod* 2005; **20**: 2983–6.
- 25 Castilla JA, Alvarez C, Aguilar J, González-Varea C, Gonzalvo MC *et al*. Influence of analytical and biological variation on the clinical interpretation of seminal parameters. *Hum Reprod* 2005; **21**: 847–51.
- 26 Clinical Pathology Accreditation Standards for the Medical Laboratory. Sheffield: Clinical Pathology Accreditation Ltd; 2007. p57.
- 27 Directive 2004/23/EC of the European Parliament and of the Council of 31 March 2004 on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells. *Off J Eur Union* 2004; **47**: L102/48–58.
- 28 Tomlinson M. Is your andrology service up to scratch? *Hum Fertil (Camb)* 2010; **13**: 194–200.
- 29 Rolland M, Le Moal J, Wagner V, Royère D, de Mouzon J. Decline in semen concentration and morphology in a sample of 26,609 men close to general population between 1989 and 2005 in France. *Hum Reprod* 2012 Dec 4. [Epub ahead of print].
- 30 Wilcox AJ. On sperm counts and data responsibility. *Epidemiology* 2011; **22**: 615–6.
- 31 Bonde JP, Ramulau-Hansen CH, Olsen J. Trends in sperm counts. The saga continues. *Epidemiology* 2011; **22**: 1–3.
- 32 Sánchez-Pozo MC, Mendiola J, Serrano M, Mozas J, Björndahl L *et al*. On behalf of the Special Interest Group in Andrology (SIGA) of the European Society of Human Reproduction and Embryology. Proposal of guidelines for the appraisal of SEMen QUALity studies (SEMQUA). *Hum Reprod* 2013; **28**: 10–21.