

## DEBATE—CONTINUED

### Influence of analytical and biological variation on the clinical interpretation of seminal parameters

J.A.Castilla<sup>1,5</sup>, C.Álvarez<sup>2</sup>, J.Aguilar<sup>3</sup>, C.González-Varea<sup>1</sup>, M.C.Gonzalvo<sup>4</sup> and L.Martínez<sup>1</sup>

<sup>1</sup>Unidad de Reproducción, Hospital Virgen de las Nieves, Granada, <sup>2</sup>Departamento de Ciencias de la Salud, Universidad de Jaén, Jaén, <sup>3</sup>Unidad de Reproducción, Clínica Avicena, Jaén and <sup>4</sup>Laboratorio CEIFER, Granada, Spain

<sup>5</sup>To whom correspondence should be addressed at: Unidad de Reproducción, Hospital Virgen de las Nieves, E-18014, Granada, Spain.  
E-mail: josea.castilla.sspa@juntadeandalucia.es

Quality assurance in semen analysis has been questioned recently in this journal. Based on the limited capacity of seminal parameter in the determination of fertility, the authors advocated abandoning methods of quality assurance in semen analysis for clinical situations. In this article, we explore arguments as to why quality assurance in semen analysis for clinical use is not 'a waste of time'. Imprecision and within-subject biological variations are the two major components involved in the dispersion of seminal parameter results obtained by analysis of a semen sample from an individual. As within-subject biological variation is constant across geography, time and population, imprecision is a very important factor in the quality of laboratory test results. We analyse this influence on various seminal parameters and observe that there is an amount of error that can be tolerated without invalidating the medical usefulness of seminal parameter determination. However, there is a maximum allowable analytical error above which the medical usefulness of seminal parameter results is invalidated. The level of performance required to facilitate clinical decision-making is termed quality specification. We comment on different strategies to define the maximum allowable analytical error.

*Key words:* analytical variation/biological variation/quality specification/semen analysis

#### Introduction

For the correct interpretation of laboratory test results concerning an individual, it is necessary to consider analytical and biological variability. Understanding of this variability is an integral part of accurate clinical judgment in test interpretation. All too often the magnitude of a change is confused with the 'sensitivity' of that change. Jequier (2005) has stated that only extreme numerical changes can be used to make a reasonably accurate prognosis (i.e. semen analysis in the determination of fertility), the influence of analytical variability on the results is small; this overlooks the fact that such an influence depends on the inherent biological variation around the homeostatic setting point (within-subject biological variation) of the biological parameter. Andrology laboratories must provide the information required by clinicians for the correct interpretation and use of semen analysis results.

#### Imprecision

Random analytical error is a very important factor in the quality of test results. The quantitative estimate of random analytical error is termed imprecision, this being defined as the closeness of agreement between independent results of measurements obtained under stipulated conditions. Laboratories

can determine imprecision by conducting replicate experiments in which the same material is analysed a given minimum of times according to National Committee for Clinical Laboratory Standards (1985) recommendations. The imprecision can then be quantified by calculating the standard deviation (SD) and the coefficient of variation [ $CV = 100 \times (SD/\text{mean})$ ]. The percentage of dispersion of a result obtained by one analysis of a single sample from an individual is calculated from the formula:

$$\text{Dispersion (\%)} = Z * CV$$

where  $Z$  is the number of standard deviations appropriate to the probability selected: 1.96 for 95% ( $P < 0.05$ ) for example, is the most commonly used  $Z$ -score.

Imprecision often varies according to the level of the analyte. In our experience, using World Health Organization (1999) guidelines for routine seminal analysis, analytical  $CV$  ( $CV_a$ ) with different control materials and different seminal quality levels for seminal parameters ranged between 8 and 12% for concentration, 7–14% for progressive motility and 7–18% for morphology (Álvarez *et al.*, 2003). These results are similar to those reported by other authors (Neuwinger *et al.*, 1990; Cooper *et al.*, 1992).

Let us assume that a result for sperm concentration is reported as  $35 \times 10^6/\text{ml}$ . Supposing zero analytical bias, if  $CV_a$  at a concentration of  $35 \times 10^6/\text{ml}$  is 10% (similar to the value described previously by us; Álvarez *et al.*, 2003), then we are 95% confident that the sperm concentration of that specimen lies between 28 and  $42 \times 10^6/\text{ml}$  ( $\pm 19.6\%$ ). If the laboratory has a  $CV_a$  of 20%, the 95% confidence interval at a concentration of  $35 \times 10^6/\text{ml}$  is 21 and  $49 \times 10^6/\text{ml}$  ( $\pm 39.2\%$ ).

In the above example, given an individual with a true homeostatic setting point of  $35 \times 10^6/\text{ml}$ , for the clinician's understanding of the patient's fertility potential it is the same to have a result of  $21 \times 10^6/\text{ml}$  (low limit of 95% confidence interval with a  $CV_a$  of 20%) as one of  $49 \times 10^6/\text{ml}$  (high limit of 95% confidence interval with a  $CV_a$  of 20%). We can accept a dispersion of results approximately less than  $\pm 40\%$ . Therefore, Jequier (2005) is correct when she states: 'the accuracy of a sperm count does little to enhance the clinician's understanding of its fertility potential'. However, the statement by Jequier (2005) does not take into account that biological variation is a factor which must be kept in mind when a semen result is interpreted.

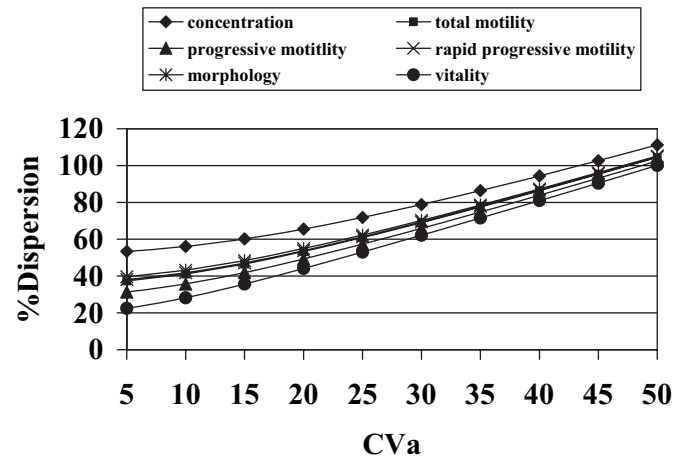
**Biological variation**

In the simplest and most widely used model, each analyte in an individual varies around the homeostatic setting point of that individual. This variation is assumed to be random in nature and therefore, like imprecision, can be expressed quantitatively as a statistic such as SD or CV. The generation and application of quantitative data on the components of biological variation (within- and between-subject biological coefficient of variation,  $CV_{Bw}$  and  $CV_{Bb}$ , respectively) has been addressed in detail by Fraser (2001). We have reported (Álvarez *et al.*, 2003) a study of components of biological variation of seminal parameters following the above-mentioned model. The model is based on healthy subjects, sperm donor candidate and strict protocol-controlled conditions (i.e. 3–4 days' abstinence, same period of study, same analytical procedure and same frequency per sample).

According to Fraser and Harris (1989), the total random variation of a laboratory test result ( $CV_t$ ) is the square root of the sum of the squares of the component variations ( $CV_t = [CV_a^2 + CV_{Bw}^2]^{1/2}$ ). Therefore, the percentage of dispersion of a result obtained by the analysis of a single sample from an individual is calculated from the formula:

$$\text{Dispersion (\%)} = Z \left( CV_a^2 + CV_{Bw}^2 \right)^{1/2}$$

Recalculating the above example, using the  $CV_{Bw}$  for sperm concentration described by us (26.8%; Álvarez *et al.*, 2003), and a  $CV_a$  of 10%, we obtain that 95% dispersion of the result is  $\pm 56.1\%$ , equivalent to an interval of 15–55 for  $35 \times 10^6/\text{ml}$ . It is unlikely, but possible, that a clinician might consider the same clinical decision with a result of  $15 \times 10^6/\text{ml}$  as with  $55 \times 10^6/\text{ml}$ . If we suppose a  $CV_a$  of 20%, 95% dispersion of the result is  $\pm 65.5\%$ , equivalent to an interval of 12–58 for  $35 \times 10^6/\text{ml}$ . We believe that in this case, a clinician could take a different decision in an individual with a reported sperm concentration value of  $12 \times 10^6/\text{ml}$  (low limit of 95% confidence interval with a  $CV_a$  of 20%) than with a reported value of



**Figure 1.** Effects of imprecision expressed as analytical coefficient of variation (x-axis) at 95% confidence on the dispersion within a certain percentage (y-axis) of the true individual homeostatic setting point of seminal parameters.

$58 \times 10^6/\text{ml}$  (high limit of 95% confidence interval with a  $CV_a$  of 20%). For example, in a young couple who, in the absence of relevant pathology (tubal, andrological, endocrine, infectious or other seminal parameters), do not conceive within 3 years, a clinician could recommend directly ICSI or IVF (if the value reported is  $12 \times 10^6/\text{ml}$ ) or intrauterine insemination (if the value reported is  $58 \times 10^6/\text{ml}$ ). Using the Internet calculator of Fraser (2005), it is shown in Figure 1 that the dispersion of seminal parameter results increases with the imprecision. Therefore, the effect of imprecision in the dispersion of the result is important, on the contrary to what was reported by Jequier (2005).

**Quality specification**

In the above example, it was stated that there is a dispersion of results ( $< 40\%$ ), which does not invalidate the medical usefulness of the sperm concentration result, and therefore that the  $CV_a$  is acceptable. What amount of analytical error can be tolerated without invalidating the medical usefulness of sperm concentration determination? The level of performance required to facilitate clinical decision-making is termed the quality specification, and it is a prerequisite for instituting quality management. The maximum allowable analytical error can be defined by various strategies (Fraser *et al.*, 1999) and these have been ordered hierarchically (Kenny *et al.*, 1999). Ideally, quality specifications should be derived objectively from an analysis of medical needs. Unfortunately, this is very difficult and the necessary calculations have been made for only a few analytes in a limited number of different clinical settings (Petersen *et al.*, 1999). Other strategies that have been recommended for determining quality specifications include professional recommendations (guidelines by national or international expert groups or by expert individuals or institutional groups), those established by regulation or by external quality assessment schemes (EQAS) or those derived from data on the state of the art.

Quality specifications based on components of biological variation, within and between subjects, have been proposed by various professional groups (Fraser, 2001). Cotlove *et al.* (1970) proposed that a desirable quality specification expressed as the analytical coefficient of variation for assays should be equal to or less than half of the within-subject biological coefficient of variation ( $<0.5CV_{Bw}$ ). However, for assays that, with currently available techniques, could not easily meet this analytical goal, Fraser *et al.* (1997) suggested a minimum analytical goal expressed as the analytical coefficient of variation of  $<0.75CV_{Bw}$ . For assays for which it is easy to meet desirable standards, the same authors suggested an optimum quality specification expressed as the analytical coefficient of variation of  $<0.25CV_{Bw}$ .

We have reported (Álvarez *et al.*, 2003) a study of analytical goals for semen parameters using the components of biological variation, following the above-mentioned recommendations of Fraser *et al.* (1997) ( $<0.75CV_{Bw}$ ). However, the lack of a standardized methodology used by those seeking to obtain the values of the components of biological variability, together with the fact that it is unclear whether biological variation components derived from healthy subjects can be extrapolated to pathological situations (Ricós *et al.*, 1999) limit their use.

For all these reasons, it is necessary to obtain analytical goals from another model. The use of the state of the art has been proposed by many bodies, including the French Society of Clinical Biology (Vassault *et al.*, 1999) and the Spanish Association of Analytic Pharmaceutics (Calafell *et al.*, 2002). Comparison of analytical quality can be accomplished through reference to the performance achieved by the best laboratories participating in EQAS. This method was used by us to calculate quality specifications for the seminal parameter (Castilla *et al.*, 2005). In summary, there exist strategies for setting the amount of error, which will allow one to establish the quality of seminal parameter determination in andrology laboratories to assist clinicians in practising good medicine.

From our results (Castilla *et al.*, 2005), and from other authors (Keel, 2004), we conclude that a significant number of laboratories need to improve the analytical quality achieved in semen analysis. Various approaches have been shown to have a positive impact on the quality of semen analysis: improvement of training (Franken *et al.*, 2000; Björdahl *et al.*, 2002; Franken *et al.*, 2003), implementing effective internal quality control (Cooper *et al.*, 1992) and quality assurance programmes (Byrd, 1992) and participating in and evaluating the results of external quality control programmes (Cooper *et al.*, 1999).

### How we can reduce the dispersion of results of a laboratory test?

As commented above, the dispersion of results depends on analytical and biological variations. Variation (as SD or CV) is reduced by the square root of the number of replicates. A general formula is  $CV_t = [CV_a^2/n_a + CV_{Bw}^2/n_s]^{1/2}$ , where  $n_a$  is the number of replicate analyses performed on each sample and  $n_s$  is the number of patient samples (Fraser, 2001). From

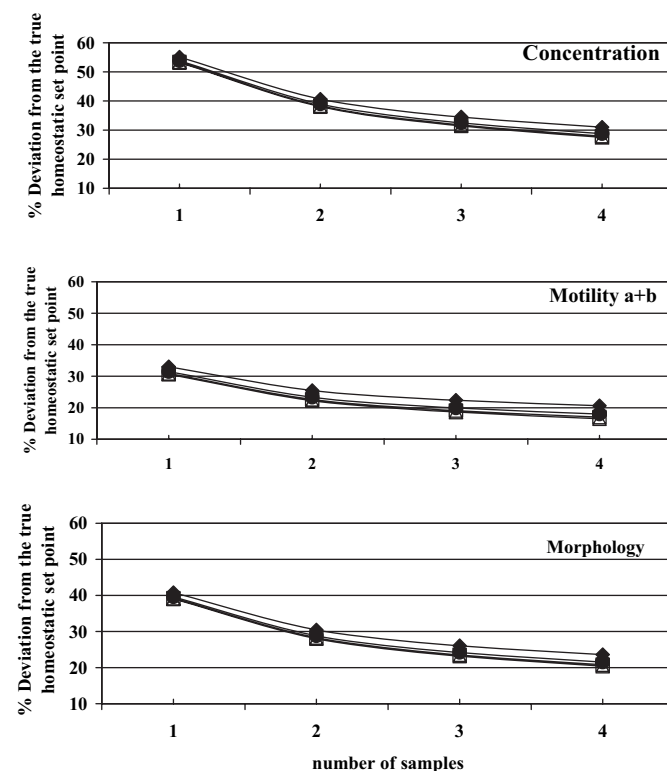
this, the number of samples required to obtain an estimate within a certain percentage of the true individual homeostatic setting point of the individual can be determined by the Internet calculator of Fraser (2005), using the formula given by Fraser (2001):

$$n = \left( \frac{Z [CV_a^2 + CV_{Bw}^2]^{1/2}}{D} \right)^2$$

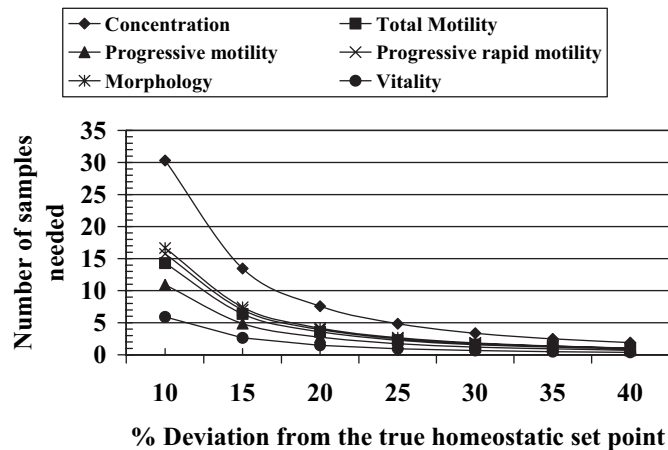
where  $D$  is the deviation in per cent from the true homeostatic set point that one would consider acceptable.

Figure 2 shows that the percentage of dispersion from the true homeostatic setting point of seminal parameters is progressively less when we increase the number of samples analysed or the replicates of a sample tested. Reducing  $CV_a$  even further through duplicate analyses has very little effect on the percentage of dispersion. However, since biological variation is larger than imprecision, taking replicate samples does not decrease the dispersion of the test result.

The number of semen samples required to obtain an estimate of the true individual homeostatic setting point of seminal parameters [concentration, total motility (grades a + b + c), progressive motility (grades a + b), rapid progressive motility (grade a), sperm morphology and vitality], according to the probability selected, is shown in Figure 3. To obtain more precise estimations ( $D < 15\%$ ) from the homeostatic setting



**Figure 2.** Percentage of dispersion from the true homeostatic setting point of seminal parameters (y-axis) according to the number of replicates of semen samples analysed (x-axis) and the number [one (◆), two (●), three (□) and four (Δ)] of replicate analyses of the same semen sample according to the probability selected ( $P < 0.05$ ).



**Figure 3.** Number of semen samples (y-axis) needed to obtain a percentage of deviation (x-axis) of the true individual homeostatic setting point of seminal parameters according to the probability selected ( $P < 0.05$ ).

point with a 95% probability, it is necessary to analyse 14 samples for seminal concentration; six for total motility, seven for morphology and three for vitality. To obtain an estimation with  $D < 40\%$  from the true homeostatic setting point with a 95% probability, it is sufficient to analyse two samples for all seminal parameters except for concentration, for which we must analyse three. Using this formula, Sergerie *et al.* (2005) recently calculated the number of tests needed to estimate the homeostatic setting point for sperm DNA fragmentation and reported that at 80% confidence three semen samples are required to obtain an estimate within  $\pm 20\%$  of the true setting point.

Traditionally, it is recommended in clinical practice that at least two semen samples should be analysed to estimate an individual's potential fertility (World Health Organization, 1999). We consider this practice correct for sperm motility, morphology and vitality. However, for sperm concentration, three semen samples should be provided to estimate an individual's profile. A similar conclusion has been reached by other authors (Poland *et al.*, 1985; Carlsen *et al.*, 2004). Jeyendran (2000) suggests that in order to obtain objective data, up to four ejaculates should be analysed.

In summary, an amount of analytical error in the seminal parameter determination is allowable without the clinical or investigating usefulness of the results being invalidated. As the biological variation of the seminal parameter is high, a very low  $CV_a$  is unnecessary. However, there is a maximum allowable analytical error (quality specification), above which the medical usefulness of the seminal parameter result is invalidated. Since the biological variation of seminal parameters is high, to reduce the dispersion of the result it is more important to obtain multiple semen samples than to replicate the analysis of a given semen sample. Andrology laboratories should communicate to the clinician the effect of the imprecision and the biological variation in the results of the seminal parameter in order to facilitate an improved clinical interpretation of semen analysis as a fundamental step in the management of male

infertility. We believe, like Fraser (2004), that the calculations outlined here should be undertaken to practise evidence-based medicine.

## References

- Álvarez C, Castilla JA, Molina R, Ramírez JP, Vergara F, Yoldi A and Gaforio JJ (2003) Biological variation in seminal parameters in healthy subjects. *Hum Reprod* 18,2082–2088.
- Björdahl L, Barratt CL, Fraser LR, Kuist U and Mortimer D (2002) ESHRE basic semen analysis courses 1995–1999: immediate beneficial effects of standardized training. *Hum Reprod* 17,1299–1305.
- Byrd W (1992) Quality assurance in the reproductive biology laboratories. *Arch Pathol Lab Med* 116,418–422.
- Calafell R, Domingo J, Fernández C, Goya R, Guardiola J, Martínez S, Morancho J, Prats R and Salve ML (2002) Especificaciones de calidad analítica. *An Clin* 27,95–100.
- Carlsen E, Petersen JH, Andersson AM and Skakkebaek NE (2004) Effects of ejaculatory frequency and season on variations in semen quality. *Fertil Steril* 82,358–366.
- Castilla JA, Morancho-Zaragoza J, Aguilar J, Prats-Gimenez R, Gonzalvo MC, Fernández-Pardo E, Álvarez C, Calafell R and Martínez L (2005) Quality specifications for seminal parameters based on state of the art. *Hum Reprod* 20,2573–2578.
- Cooper TG, Neuwinger J, Bahrs S and Nieschlag E (1992) Internal quality control of semen analysis. *Fertil Steril* 58,172–178.
- Cooper TG, Atkinson AD and Nieschlag E (1999) Experience with external quality control in spermatology. *Hum Reprod* 14,765–769.
- Cotlove E, Harris EK and Williams GZ (1970) Biological and analytic components of variation in long term studies of serum constituents in normal subjects. III. Physiological and medical implications. *Clin Chem* 16,1028–1032.
- Franken DR, Smith M, Menkveld R, Kruger TF, Sekadde-Kigundu C, Mbizvo M and Akande EO (2000) The development of a continuous quality control (CQS) program for strict sperm morphology among sub-Saharan African laboratories. *Hum Reprod* 5,667–671.
- Franken DR, Menkveld R, Kruger TF, Sekadde-Kigundu C and Lombard C (2003) Monitoring technologist reading skills in a sperm morphology quality control program. *Fertil Steril* 79,1637–1643.
- Fraser CG (2001) *Biological Variation: From Principles to Practice*, 1st edn. AACC Press, Washington, DC, USA.
- Fraser CG (2004) Test result variation and the quality of evidence-based clinical guidelines. *Clin Chim Acta* 346,19–24.
- Fraser CG (2005) Are 'Scientific Statements' the Scientific Truth? Guest editorial (<http://www.westgard.com/guest23.htm>).
- Fraser CG and Harris EK (1989) Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 27,409–437.
- Fraser CG, Petersen PH, Libeer JC and Ricós C (1997) Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 34,8–12.
- Fraser CG, Kallner A, Kenny D and Peterson PH (1999) Introduction: strategies to set global quality specifications in laboratory medicine. *Scand J Clin Lab Invest* 59,477–478.
- Jequier AM (2005) Is quality assurance in semen analysis still really necessary? A clinician's viewpoint. *Hum Reprod* 20,2039–2042.
- Jeyendran RS (2000) *Interpretation of Semen Analysis Results*, 1st edn. Cambridge University Press, Cambridge.
- Keel BA (2004) How reliable are results from the semen analysis? *Fertil Steril* 82,41–44.
- Kenny D, Fraser CG, Petersen PH and Kallner A (1999) Consensus agreement. *Scand J Clin Lab Invest* 59,585.
- National Committee for Clinical Laboratory Standards (1985) Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline. NCCLS Publication EP5-A.
- Neuwinger J, Behre HM and Nieschlag E (1990) External quality control in the andrology laboratory: an experimental multicenter trial. *Fertil Steril* 54,308–314.
- Petersen PH, Fraser CG, Kallner A and Kenny D (1999) Strategies to set global analytical quality specifications in laboratory medicine. *Scand J Clin Lab Invest* 59,1–164.
- Poland ML, Moghissi KS, Giblin PT, Ager JW and Olson JM (1985) Variation of semen measures within normal men. *Fertil Steril* 44,396–400.

Ricós C, Álvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, Minchinela J, Perich C and Simón M (1999) Current database on biologic variation: pros, cons and progress. *Scand J Clin Lab Invest* 59,491–500.

Sergerie M, Laforest G, Boulanger K, Bissonnette F and Bleau G (2005) Longitudinal study of sperm DNA fragmentation as measured by terminal uridine nick end-labelling assay. *Hum Reprod* 20,1921–1927.

Vassault A, Grafmeyer D, de Graeve J, Cohen R, Beaudonnet A and Bienvenu J (1999) *Analyses de biologie médicale: spécification et normes*

d'acceptabilité à l'usage de la validation de techniques. *Ann Biol Clin* 57,685–695.

World Health Organization (1999) *WHO Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, 4th edn. Cambridge University Press, Cambridge.

*Submitted on September 28, 2005; resubmitted on November 2, 2005; accepted on November 9, 2005*



# QUALITY, INNOVATION, AND SERVICE—IT'S ALL AT THE CENTER OF EVERYTHING WE DO.

From developing assisted reproductive technologies that maximize performance, like the first ART media and cultures, to expertise that streamlines productivity, FUJIFILM Irvine Scientific brings together decades of industry expertise with a powerhouse of innovation, turning opportunities into realities. Together, we're working to support healthy futures—from retrieval to realization.

**ALL IN** FOR LIFE.  
[irvinesci.com/ALLIN](http://irvinesci.com/ALLIN)

**FUJIFILM**  
Value from Innovation

 **IrvineScientific**