Human Reproduction, Vol.27, No.2 pp. 314–322, 2012
Advanced Access publication on December 5, 2011 doi:10.1093/humrep/der413

human reproduction

ORIGINAL ARTICLE *Andrology*

Acceptable variability in external quality assessment programmes for basic semen analysis

E.R. Palacios[1], A. Clavero[1], M.C. Gonzalvo[1,2], A. Rosales[1], J. Mozas[1], L. Martínez[1], J.P. Ramírez[3,4], L. Björndahl[5,6], J. Morancho-Zaragoza[7], E. Fernández-Pardo[7], and J.A. Castilla[1,2,3,4,*]

[1]Unidad de Reproducción, Hospital Universitario Virgen de las Nieves, Granada E-18014, Spain [2]Programa de Control de Calidad Externo de Análisis de Semen, Asociación para el Estudio de la Biología de la Reproducción (ASEBIR), Granada, Spain [3]Banco de Semen Ceifer, Granada E-18005, Spain [4]Más Vida Reproducción, Sevilla E-41012, Spain [5]Centre for Andrology and Sexual Medicine, Clinic of Endocrinology, Karolinska University Hospital, Huddinge, Stockholm, Sweden [6]Department of Medicine, Karolinska University Hospital, Huddinge, Stockholm, Sweden [7]Programa de Supervisión Externa de la Calidad, Asociación Española de Farmacéuticos Analistas, AEFA, Madrid, Spain

*Correspondence address: Fax: +34-958020226; E-mail: josea.castilla.sspa@juntadeandalucia.es

Submitted on May 16, 2011; resubmitted on July 22, 2011; accepted on October 17, 2011

BACKGROUND: External quality assessment is essential in modern andrology laboratories. To assess the proficiency of laboratories participating in an external quality assessment programme (EQAP), limits for acceptable variability must be determined. Limits currently specified largely depend on criteria set by the organizers of individual EQAP schemes. The objective of this study was to evaluate the different criteria described in ISO 13528:2005 for calculating acceptable variability in EQAP when applied to basic semen analysis parameters.

METHODS AND RESULTS: The data used in this study were the means and standard deviations obtained for independent samples from two EQAPs, one national (Spanish) and one international (European). The acceptable variability according to ISO 13528:2005 was calculated using four types of criteria: (i) ± 3 standard deviations of the results of all participating laboratories; (ii) ± 3 standard deviations of the results of expert laboratories; (iii) quality specifications based on biological variability, state-of-the-art and clinicians' opinions and (iv) the same quality specifications adjusted for the uncertainty of the assigned value. The first two strategies resulted in very wide ranges of acceptable variability. Conversely, the strategy based only on quality specifications resulted in very narrow ranges. For the fourth strategy, acceptable ranges were intermediate between the results produced with the other strategies. The third and fourth strategies did not produce observable differences in acceptable ranges when the model used for calculating the specifications of analytical quality was changed.

CONCLUSIONS: It is essential that EQAPs for semen parameters should determine the ranges for acceptable variability in results. Moreover, these ranges must be clinically useful, i.e. the variability should have a minimal negative impact on clinical decisions. The exact definition of 'expert laboratory' is more important than the model chosen for estimating analytical quality specifications in an EQAP for semen parameters in basic semen analysis.

Key words: quality control / semen / basic semen analysis / acceptable variability

Introduction

Semen analysis provides useful information in the clinical management of infertility, in the evaluation of masculine contraceptive methods and in epidemiological studies of reproductive health or toxicology (Hancock *et al*., 2002). This analysis should be performed in accordance with international recommendations aimed at reducing biological and analytical variability (Björndahl *et al*., 2010; WHO, 2010). These two factors are crucial to obtaining a correct interpretation of the results produced (Castilla *et al*., 2006).

To reduce analytical variability in and between andrology laboratories, it is essential to participate in external quality assessment programmes (EQAPs) (McCulloh, 2004), which are currently implemented in many countries. Participating laboratories evaluate samples from the same semen pool and receive feedback on the proficiency of their evaluations. EQAPs monitor all laboratory procedures relating to collecting and reporting data to ensure that laboratory processes are under control. An EQAP allows a laboratory to compare its results with those of others. It permits different methods to be evaluated and compared on a scale not possible in a single laboratory.

EQAP and internal quality control (IQA) are complementary processes. An EQAP may reveal problems with accuracy that may not be apparent from IQA if control samples are not adequately masked or selected. An EQAP has the advantage that it allows a laboratory to monitor the accuracy and stability of its methods (Plaut and Westgard, 2002).

The competence of laboratories participating in EQAPs should be evaluated following one of the methods described by the International Organisation for Standardization. According to ISO 13528:2005a, performance is satisfactory when the laboratory result is within established upper and lower acceptance limits (i.e. the acceptance interval). This range is calculated by adding and subtracting, respectively, the margin of error from the assigned value. The range of acceptable variability for laboratory results in an EQAP is of crucial importance: if it is too wide, poorly performing laboratories will not be identified, while if it is too narrow, good laboratories run a substantial risk of being labelled as poor performers. In calculating acceptance limits, crucial aspects include the selection of quality specifications (ISO/TR 15196:2001) and calculation of the assigned target value by the EQAP (ISO 13528:2005a).

At present, the variability considered acceptable usually depends on the criteria of EQAP organizers. Thus, for the same set of results, the performance of a laboratory might be considered acceptable in one programme and insufficient in others (Cooper *et al*., 2002; Arnaud *et al*., 2008). This observation is well known and has been described previously (Christensen and Olsen, 1996; Taylor *et al*., 2002). The discrepancy arises for two main reasons: firstly, the difficulty in selecting quality specifications that fit the criteria of the Stockholm Consensus (Kenny *et al*., 1999); and secondly, the uncertainty contributed by the organizers themselves in the measures of central tendency (mean values) assigned, which are subsequently used for comparison purposes (ISO 13528:2005a). This lack of standardization among EQAPs means that their usefulness in human semen analysis is not always fully realized by clinicians (Jequier, 2005).

The aim of this study was to assess the different criteria described in ISO 13528:2005a to calculate acceptable variability in a basic semen analysis EQAP. In addition, the influence on acceptable variability of different criteria under the Stockholm Consensus for determining analytical quality specifications is evaluated.

Materials and Methods

Materials

The data used in this study are the values for mean and standard deviation obtained for independent control samples provided in two semen analysis EQAPs. Both included the determination of sperm concentration, motility, morphology and vitality. The first EQAP was the Special Interest Group in Andrology (SIGA) of the European Society for Human Reproduction and Embryology (ESHRE). This involved a total of 80 laboratories of which 14–47 participated in particular deliveries. The programme was implemented from 2001 to 2009, and two deliveries per year were made. Data on three deliveries were lost due to computer failure, and so data from only 15 deliveries were available for analysis. Each distribution comprised the following elements: for the determination of sperm concentration, two samples of undiluted and two of diluted semen; for motility, videotape/DVD recordings of four sperm samples; for morphology, two unstained and two stained semen smears and for vitality, two eosin Y-Nigrosin semen smears. The second EQAP was that of the Spanish Association of Pharmaceutical Analysts (AEFA) and the Spanish Association of Medical Biopathology (AEBM), organized in collaboration with the Centre for Studies and Research into Fertility (CEIFER) and sponsored by the Association for the Study of Reproductive Biology (ASEBIR). This EQAP sampled a total of 206 laboratories from 1999 to 2009 with 31–144 laboratories participating at a given time. This scheme also made two deliveries per year, and so 22 results are available for analysis. Each distribution contained two sperm suspensions for the analysis of sperm concentration, at least two samples on videotape/DVD for motility, two unstained and two stained semen smears for morphology and two eosin Y-Nigrosin semen smears for vitality (Castilla *et al*., 2005).

The samples used for both semen analysis EQAPs were obtained from donor candidates or patients, all of whom had given their informed consent for their ejaculates to be used in the investigation. Serum studies were performed for HIV, hepatitis B, hepatitis C and syphilis. All were negative. All samples were treated with preservatives to eliminate harmful micro-organisms.

Methods

Four different strategies were examined for calculating the acceptable variability according to ISO 13528:2005a (Palacios *et al*., 2011) (Table I):

(i) The target values were determined from mean values calculated from the results of all participating laboratories and the acceptable variability was calculated as the SD multiplied by three (ISO 13528:2005a 5.6 and 6.6.1).

(ii) The assigned target value was the mean values of the results from the expert laboratories. In this case the range of acceptable variability was calculated in the same way as in (1) but using the SD of the data from reference laboratories (ISO 13528:2005a 5.5 and 6.6). Laboratories considered 'expert laboratories' are those with special interest and competence in andrology (research and clinical laboratory andrology work) including training laboratory staff in basic semen analysis methods as described in the Manual on Basic Semen Analysis (EHSRE Monographs, 2002) and its predecessors. Furthermore, they should be involved in arranging ESHRE-SIGA basic semen analysis courses and should have implemented IQC and be participating in EQA.

(iii) Using desirable quality specifications for the total analytical error calculated according to different models proposed at the Stockholm Consensus Conference. Target values were assigned as the mean values for the expert laboratories. The acceptable variability was determined by multiplying the target value by a coefficient derived from: (i) biological variability (Álvarez *et al*., 2003); (ii) state-of-the-art (Castilla *et al*., 2005) and (iii) clinicians' opinion (Aguilar *et al*., 2008) (ISO 13528:2005a 5.5.1 and 6.3) (Supplementary data 1).

(iv) Adjusting for the uncertainty of the assigned value. First, the target value was calculated from the mean of the expert laboratories' results. Then, the acceptable variability was calculated from the SD derived from the quality specifications for total analytical error, for each of the models described in Strategy 3, and the uncertainty of the assigned target value (ISO 13528:2005a 5.5.2 and 6.3).

In order to determine whether the accuracy of sperm concentration measurements made on diluted and undiluted samples was similar in each case, we compared linear regression lines fitted to plots of the mean versus 3 SD of the sperm concentration for the 30 samples of each type in the European EQAP.

Concerning sperm morphology, no distinction was made between the data for stained and unstained semen smears, as Álvarez *et al*. (2005) showed, there are no differences between these results. The SD in

Table I Strategies employed in calculating acceptance limits, according to ISO 13528 (9).

Strategy	Assigned value, by consensus	Dispersion value, by consensus	Margin of error	Allowable margin of error
1. Based on participating laboratories	Mean ($\bar{X}_1$) of European participating laboratories (SIGA)	Standard deviation (SD_1) of European participating laboratories (SIGA)	$\pm 3\ SD_1$	$\bar{X}_1 \pm 3\ SD_1$
	Mean ($\bar{X}_2$) of Spanish participating laboratories	Standard deviation (SD_2) of Spanish participating laboratories	$\pm 3\ SD_2$	$\bar{X}_2 \pm 3\ SD_2$
2. Based on expert laboratories	Mean ($\bar{X}_3$) of European expert laboratories (SIGA)	Standard deviation (SD_3) of European expert laboratories (SIGA)	$\pm 3\ SD_3$	$\bar{X}_3 \pm 3\ SD_3$
3. Using quality specifications	Mean ($\bar{X}_3$) of European expert laboratories (SIGA)	Quality specification based on biological variability (bv)	$\pm(e_{bv}\ \bar{X}_3)$	$\bar{X}_3 \pm (e_{bv}\ \bar{X}_3)$
		Quality specification based on the state of the art (sa)	$\pm(e_{sa}\ \bar{X}_3)$	$\bar{X}_3 \pm (e_{sa}\ \bar{X}_3)$
		Quality specification based on clinicians' opinion (co)	$\pm(e_{co}\ \bar{X}_3)$	$\bar{X}_3 \pm (e_{co}\ \bar{X}_3)$
4. Adjusting for the uncertainty of the assigned value	Mean ($\bar{X}_3$) of European expert laboratories (SIGA)	Quality specification based on biological variability (bv)	$\pm 3 \cdot \sqrt{\left(\frac{e_{bv}\ \bar{X}_3}{3}\right)^2 + \left(\frac{1,25\ s^*}{\sqrt{n}}\right)^2}$	$\bar{X}_3 \pm 3 \cdot \sqrt{\left(\frac{e_{bv}\ \bar{X}_3}{3}\right)^2 + \left(\frac{1,25\ s^*}{\sqrt{n}}\right)^2}$
		Quality specification based on the state of the art (sa)	$\pm 3 \cdot \sqrt{\left(\frac{e_{sa}\ \bar{X}_3}{3}\right)^2 + \left(\frac{1,25\ s^*}{\sqrt{n}}\right)^2}$	$\bar{X}_3 \pm 3 \cdot \sqrt{\left(\frac{e_{sa}\ \bar{X}_3}{3}\right)^2 + \left(\frac{1,25\ s^*}{\sqrt{n}}\right)^2}$
		Quality specification based on clinicians' opinion (co)	$\pm 3 \cdot \sqrt{\left(\frac{e_{co}\ \bar{X}_3}{3}\right)^2 + \left(\frac{1,25\ s^*}{\sqrt{n}}\right)^2}$	$\bar{X}_3 \pm 3 \cdot \sqrt{\left(\frac{e_{co}\ \bar{X}_3}{3}\right)^2 + \left(\frac{1,25\ s^*}{\sqrt{n}}\right)^2}$

s^*, standard deviation obtained from the linear regression of the mean for the expert laboratories against the corresponding standard deviation.
n, the number of laboratories participating in this delivery.
e, quality specification for total analytical error.

Table II Upper (UL) and lower (LL) limits of acceptance according to the strategies analysed in the European EQAP for target values equal to the lower reference limits for different sperm parameters specified in the WHO Handbook (5th edition).

Strategies	Concentration (15×10^6/ml)		Total motility (40%)		Progressive motility (32%)		Vitality (58%)		Morphology (4%)	
	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
Based on the results of expert laboratories	37	0	58	22	59	5	73	43	10	0
Using quality specifications										
Biological variability	21	9	49	31	49	25	67	49	5	3
State of the art	21	9	48	32	41	23	78	38	7	1
Clinicians' opinions	23	7	58	22	45	19	74	42	7	1
Adjusting for the uncertainty of the assigned value										
Biological variability	28	2	54	26	48	16	70	46	8	0
State of the art	28	2	54	26	50	14	80	36	9	0
Clinicians' opinions	29	1	62	18	52	12	76	40	8	0

Strategy 4 was calculated from the regression lines between the mean values and the SDs for the expert laboratories. In calculating all these regression lines, outlier values were excluded, using the Studentized residual method (Freund and Littell, 1991).

To facilitate the calculation of acceptable variability by the organizers of the basic semen analysis EQAP, a spreadsheet was developed, on which, from a given target value by the laboratories, the corresponding allowable margins of error were obtained. The SD for proficiency assessment used was calculated from the regression curves between the mean values and the SD for the expert laboratories. Table II provides an example with the lower reference limits of WHO (2010), using all the above-described strategies.

Figure 1 Assigned target values and ranges of acceptable variability for the European EQAP analyses of progressive sperm motility according to Strategy 1 based on participating laboratories.

Table III Highest value of the different semen parameters for which 0 is an acceptable result, according to the various strategies analysed.

Strategies	Concentration (10^6/ml)	Total motility (%)	Progressive motility (%)	Rapid motility (%)	Vitality (%)	Morphology (%)
Based on the results of participating laboratories	121	27	38	51	36	19
Based on the results of expert laboratories	25	17	29	48	28	13
Using quality specifications						
Biological variability	0	0	0	0	0	0
State of the art	0	0	0	1	0	3
Clinicians' opinions	1	0	0	0	0	1
Adjusting for the uncertainty of the assigned value						
Biological variability	13	10	17	13	9	4
State of the art	13	10	17	20	10	22
Clinicians' opinions	14	11	18	15	10	18

To analyse the inter-laboratory variability between the European and Spanish EQAPs, we compared the coefficients of variation (CV) obtained by the European EQAP (participating and expert laboratories) and the Spanish EQAP (participating laboratories), using the Kruskal–Wallis test.

All data were analysed using SPSS statistical 15.0 software (IBM Company Headquarters, IL, USA).

Results

Comparison of the regression lines of the mean and 3SD for sperm concentration in the diluted and undiluted samples presented no statistically significant differences. This was true for both the participating and the expert laboratories (Supplementary data 2). Accordingly, in the rest of the analyses, this factor was not taken into account and the data from diluted and undiluted samples were analysed together.

On applying Strategies 1 and 2 to the basic semen parameters, we observed that similar target values often had widely varying margins of error (Fig. 1 and Supplementary data 3, Figs S1–5a).

Based on the results from centres participating in the European EQAP for concentration, morphology and progressive rapid motility, the analysis of Strategy 1 revealed a high number of analyses in which a result of 0 was included in the range of acceptable values. With respect to concentration, this was the case in 39 samples (88.63%), with assigned target values ranging from 10 to 200 $\times$ 10^6/ml. For morphology, the same occurred in 59 analyses (98.33%) with assigned target values ranging from 1 to 19% of normal forms. For progressive rapid motility, 0 was included in the acceptable range in 45 cases (75%), with assigned target values ranging from 3 to 54% of rapidly progressive forms (Table III).

Similar results were observed for the data from centres participating in the Spanish EQAP for the different semen parameters. In the case of morphology, all 60 analyses (100%) had 0% of normal forms as an acceptable result, with assigned values ranging from 8 to 34%. With respect to spermatozoa with progressive rapid motility, this occurred in 57 analyses (95%) with target values ranging from 5 to 55% of mobile forms. For concentration, similar results were observed, although with narrower margins of acceptance. In this case, in 17 analyses (38.6%) 0×10^6/ml was considered an acceptable result, despite the fact that the assigned target value in these 17 analyses ranged from 2 to 18×10^6/ml (Fig. 2 and Supplementary data 3, Figs S1–5b).

In the analysis of Strategy 2, based on the results from expert laboratories in the European EQAP, narrower ranges of acceptable variability were observed. For concentration, there were fewer analyses in which 0×10^6/ml was included in the range of acceptable values: only 10 analyses (22.72%) with assigned values of $5–70 \times 10^6$/ml. This also occurred with the percentage of spermatozoa with progressive motility, in which for 37 analyses (61.66%), with assigned target values ranging from 2 to 48% of progressive sperm, 0% of progressive sperm was considered to be an acceptable result. Nevertheless, considerable differences remained in the ranges of acceptable variability for samples with similar target values, for all the semen parameters analysed (Fig. 3 and Supplementary data 3, Figs S1–5c).

In the case of Strategy 3, based on specifications of analytical quality estimated according to criteria of biological variability or state-of-the-art knowledge, the range of acceptable variation was much narrower for all semen analysis parameters, particularly for the lower target values, compared with ranges obtained with the other strategies. Thus, not a single parameter with a value of 0 was

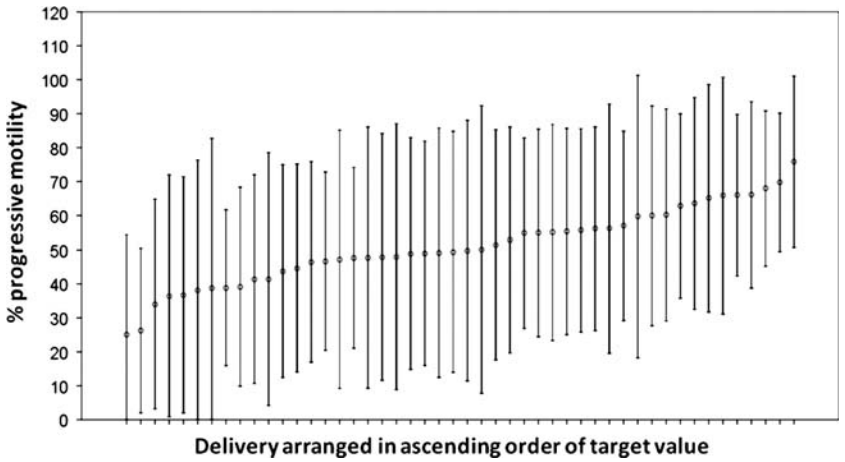

Figure 2 Assigned target values and ranges of acceptable variability for the Spanish EQAP analyses of progressive sperm motility according to Strategy 1 based on participating laboratories.

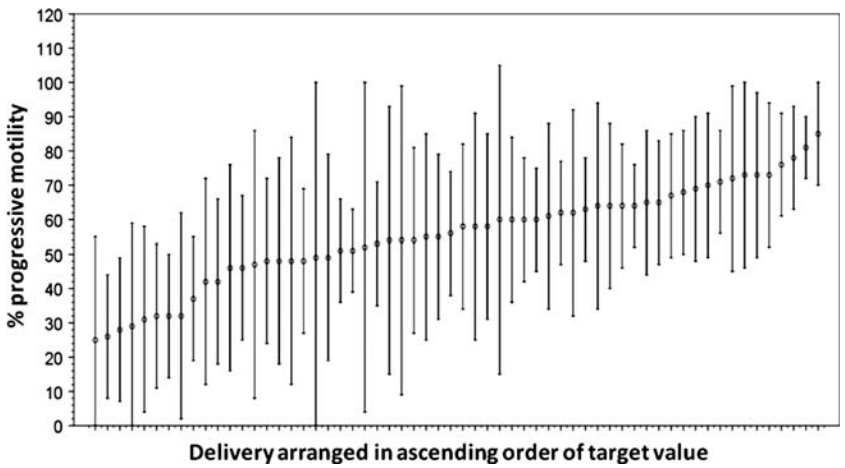

Figure 3 Assigned target values and ranges of acceptable variability for the European EQAP analyses of progressive sperm motility according to Strategy 2 based on expert laboratories.

Figure 4 Assigned target values and ranges of acceptable variability for the European EQAP analyses of progressive sperm motility according to Strategy 3 based on specifications of analytical quality according to the model of biological variability.

Figure 5 Assigned target values and ranges of acceptable variability for the European EQAP analyses of progressive sperm motility according to Strategy 4 adjusted for the uncertainty of the assigned target value, in which the quality specification is biological variability.

considered acceptable, even among the lowest assigned values (Fig. 4 and Supplementary data 3, Figs S1–5d). When the same calculations were performed using the criterion of clinicians' opinions to estimate the quality specifications, the results varied very little, with the resulting intervals being slightly wider, but still very similar.

Strategy 4 included an adjustment for the uncertainty of the assigned target value, and the resulting ranges of acceptable variability were between the extremes of the other strategies (Fig. 5 and Supplementary data 3, Figs S1–5e). As with Strategy 3, the variations between these ranges, under the different criteria for estimating quality specifications, were very slight.

Table IV shows the mean values for the CVs between the participating and the expert laboratories with respect to each analysis, for the two EQAPs analysed. For all semen parameters, the CVs were lower among the experts than among the other participating laboratories.

An Internet application can be found (http://www.ceifer.es/ceifer/descargas/acceptable_variability.xls) for calculating acceptable variability on the basis of an assigned target value by expert laboratories. Table II shows these acceptance limits for the lower reference limits of the various semen parameters in accordance with the manual published by the WHO (2010), applying the different strategies studied, based on results from expert laboratories.

Discussion

Participation in EQAPs or proficiency testing schemes allows laboratories to evaluate the quality of their results. It is also necessary in

Table IV Between laboratories CV % of the Spanish and European EQAPs (participant and expert laboratories) for all sperm parameters.

	CV Spain (44 samples)	CV European participants (60 samples)	CV European expert laboratories (60 samples)
Concentration (10^6/ml)	30.0 27.3–33.6	34.7 31.5–37.5	21.3[a,b] 17.5–23.6
Morphology (% normal forms)	65.0 60.3–68	74.2[c] 66.7–85.7	50.0[b,d] 50–60
Total motility (% motile forms)	15.7 13–17	13.6 12.5–15	9.9[a,b] 8.6–10.9
Progressive motility (% motile forms)	10.4 10–11.4	15.3[c] 13.2–18.1	14.7[a] 11–18.1
Rapid motility (% motile forms)	55.5 50.1–59.6	50.0 39.4–58.2	41.4[d] 30.3–50
Vitality (% live forms)	15.3 12.9–17.8	14.9 12.1–17.8	10.3[a,e,g] 6.7–14

Values are the median (95% confidence interval) for the number of samples stated.
[a]$P < 0.001$ CV European experts versus CV Spanish participants.
[b]$P < 0.001$ CV European experts versus CV European participants.
[c]$P < 0.001$ CV European participants versus CV Spanish participants.
[d]$P < 0.01$ CV European experts versus CV Spanish participants.
[e]$P < 0.01$ CV European experts versus CV European participants.
[f]$P < 0.01$ CV European participants versus CV Spanish participants.
[g]30 samples.

order to fulfil the requirements for accreditation (ISO 15189:2003, ISO/IEC 17025:2005b) and, in some jurisdictions, it may be mandated by law (Killeen, 2009; Howerton *et al.*, 2010). Hitherto, the definitions of acceptable performance have largely depended on the judgment of scheme organizers. Therefore, from the same set of results, the performance of a laboratory may be considered as adequate in one scheme and inadequate in another, as clearly demonstrated by our results. With the aim of standardizing the monitoring of participants' performance and the associated assessment of competence in EQAP for semen analysis, we have considered the four strategies proposed in ISO 13528:2005a for calculating acceptance limits.

On analysing Strategy 1, based on the participating laboratories in the European EQAP, we found very wide ranges for acceptable variability for the assigned target values to basic semen analysis parameters. Thus, laboratories would be accepted despite the considerable variation in results—results that would lead to different clinical decisions. For example, in the case of a couple without signs of negative female factors for a true result of 80×10^6/ml both 2×10^6/ml and 160×10^6/ml would be regarded as acceptable results. However, for results at the lower end ICSI would be chosen in most centres, while other results would commonly lead to IVF or the couple could even be recommended to continue attempting to conceive by themselves. Thus, the existence of a large range of acceptable variation allows significant possibilities of an inadequate option being adopted.

The variability among the laboratories participating in the Spanish EQAP is similar to that found for those participating in European EQAPs for most of the basic semen analysis parameters analysed, as shown by the comparison of the inter-laboratory CVs obtained. On the other hand, expert laboratories tend to present lower CVs among laboratories than do participating laboratories, which suggest that the data to be used for calculating allowable margins of error should be those provided by expert laboratories, as is in fact suggested in the new WHO handbook on Semen Analysis (WHO, 2010).

The lower level of variability among expert laboratories means that when their results are used for calculating the ranges of acceptable variability, these ranges are narrower. Thus, Strategy 2 based on expert laboratories, produces narrower ranges of acceptable variability, although they still remain relatively broad (Table IV).

Using quality specifications (Strategy 3), the ranges of acceptable variability are too strict and it is too difficult to achieve results within the range. This strategy is not feasible with current technologies and procedures, not only for the participating laboratories, but also for the experts. This conclusion is corroborated by the previous findings of our group (Castilla *et al.*, 2005). Quality specifications based on biological variability, the state-of-the-art or clinicians' opinions should be considered a goal and not constitute inflexible criteria of acceptability (Kenny *et al.*, 1999).

Adjusting for the uncertainty of the assigned target value (Strategy 4), using any of the three methods described for calculating quality specifications, provides clinically useful acceptance intervals, as the variability will have minimal influence on clinical decisions based on the laboratory results. In our opinion, therefore, this selection method should be taken into account for establishing allowable errors of margin in an EQAP of basic semen analysis parameters, with the exception of sperm morphology.

In the latter case, when Strategy 4 is used, with clinically significant assigned values (e.g. 4%), a 0% value would continue to be 'acceptable'. This situation arises because analytical inaccuracy increases rapidly when assigned values are very low. To overcome this problem, other EQAPs have used two criteria for establishing acceptance limits (Fraser, 1999; Taylor *et al.*, 2002).

The SD values obtained from both the participating and the expert laboratories, for samples with similar assigned target values, sometimes differed considerably. To avoid inconsistencies between analyses, we have developed a software application where, for a given target value by expert laboratories, ranges for acceptable variability are obtained. By this procedure, the organizers should obtain more consistent results.

None of the analyses made, in either the European or the Spanish programmes, presented an assigned target value of 0 for any of the sperm parameters. Various authors have addressed the problem of having an assigned target value of 0 in quality control programmes (Cooper *et al*., 2002). For these cases, margins of error might be calculated using estimation of the theoretical counting error, although these margins of error would be very strict (WHO, 2010). We believe an international consensus among experts is necessary regarding this question, because unlike other biological magnitudes, the value 0 in basic semen analysis is of great clinical importance (for example, post vasectomy). Indeed, for the safe determination of very low sperm concentrations, special operating procedure standards are required (Kvist and Björndahl, 2002; Björndahl *et al*., 2010).

Another issue in EQAPs for basic semen analysis is that of standardizing the preparation of distributed samples. The fact that no differences were observed in the European EQAP, in the linear regression between the mean and 3 SDs in assessing the concentration in diluted and undiluted samples, suggests that inter-laboratory variability is due more to the counting procedure than to dilution methods. This in turn suggests that in an EQAP aiming only at proficiency testing, it is not necessary to deliver both types of samples, as is currently done in the programmes run by the American Association of Bioanalysts (2010) in the USA and in the Spanish programme (Álvarez *et al*., 2005). However, for an EQAP also aiming at education and the development of participating laboratories' operating procedures, the comparison of diluted and undiluted samples for sperm concentration assessment can be of value in evaluating causes of poor performance. As the dilution methods used by each participating laboratory were not specified, it was not possible to investigate this point further.

On the other hand, the period of validity of the result obtained by a laboratory in a single round of a proficiency testing schemes is limited to the time that the laboratory performed the test. Thus, if a laboratory achieves a satisfactory result in a single round, the results should not be used to support a claim that the laboratory obtained reliable data on any other occasion. A laboratory that operates a quality system and achieves a history of satisfactory results in many rounds of a proficiency testing scheme entitled to use the results as evidence that it is able to obtain consistently reliable data.

In summary, the organizers of a semen analysis EQAP should establish ranges for acceptable variability using strategies that take into account the uncertainty of the assigned target value, and calculate the latter on the basis of the results provided by expert laboratories. Therefore, the definition of what is an expert laboratory constitutes a fundamental element in the semen analysis EQAP in order to establish allowable margins of error. An expert laboratory, in our opinion, must apply standardized, internationally accepted methods (Kvist and Björndahl, 2002; Björndahl *et al*., 2010; WHO, 2010), run an internal quality control programme and participate in external quality control programmes.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors' roles

E.R.P. was active in study conception and design, data analysis, drafting the manuscript and final approval. A.C. was involved in conception and design and critical revision of the article. M.C.G. took part in interpretation of data and drafting the manuscript. A.R. was involved in data analysis and drafting the manuscript. J.M. and L.M. contributed to the interpretation of the data and final approval of article. J.P.R. drafted the manuscript and took part in the acquisition of the data. L.B., J.M.-Z. and E.F.-P. were involved in the acquisition of the data and data analysis. In addition, L.B. revised the article critically. J.A.C. took part in the conception and design and final approval of article.

Funding

The funding organizations played no role in the design of study, review and interpretation of data, or preparation or approval of manuscript.

Conflict of interest

There are no conflicts of interest.

References

Aguilar J, Álvarez C, Morancho-Zaragoza J, Prats-Gimenez R, Ramírez JP, Fernández-Pardo E, Martínez L, Calafell R, Duran I, Castilla JA. Quality specifications for seminal parameters based on clinicians' opinions. *Scand J Clin Lab Invest* 2008;**68**:68–76.

Álvarez C, Castilla JA, Martínez L, Ramírez JP, Vergara F, Gaforio JJ. Biological variation of seminal parameters in healthy subjects. *Hum Reprod* 2003;**10**:1–7.

Álvarez C, Castilla JA, Ramírez JP, Vergara F, Yoldi A, Fernández A, Gafforio JJ. External quality control program for semen analysis: Spanish experience. *J Assist Reprod Genet* 2005;**22**:379–387.

American Association of Bioanalysts. Proficiency testing service. http://www.aab-pts.org/ (19 September 2010, date last accessed) 2010.

Arnaud J, Weber JP, Weykamp CW, Parsons PJ, Angerer J, Mairiaux E, Mazzarasa O, Valkonen S, Menditto A, Patriarca M *et al*. Quality specifications for the determination of copper, zinc, and selenium in human serum or plasma: evaluation of an approach based on biological and analytical variation. *Clin Chem* 2008;**54**:1892–1899.

Björndahl L, Mortimer D, Barratt CLR, Castilla JA, Menkveld R, Kvist U, Alvarez JG, Haugen TB. *A Practical Guide to Basic Laboratory Andrology*, 1st edn. Cambridge, UK: Cambridge University Press, 2010.

Castilla JA, Morancho-Zaragoza J, Aguilar J, Prats-Gimenez R, Gonzalvo MC, Fernández-Pardo E. Quality specifications for seminal parameters based on the state of the art. *Hum Reprod* 2005;**20**:2573–2578.

Castilla JA, Álvarez C, Aguilar J, González-Varea C, Gonzalvo MC, Martínez L. Influence of analytical and biological variation on the clinical interpretation of seminal parameters. *Hum Reprod* 2006;**21**:847–851.

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**:306–311.

Christensen JM, Olsen E. Comparison of some European external quality assessment schemes in the field of occupational and environmental medicine. *Ann Ist Super Sanita* 1996;**32**:285–294.

Fraser CG. General strategies to set quality specifications for reliability performance characteristics. *Scand J Clin Lab Invest* 1999;**59**:487–490.

Freund RJ, Littell RC. *SAS System for Regression*, 2nd edn. Cary, NC, USA: SAS Institute Inc, 1991.

Hancock P, McLaughlin E, The British Andrology Society. British Andrology Society guidelines for the assessment of post vasectomy semen samples. *J Clin Pathol* 2002;**55**:812–816.

Howerton D, Krolak JM, Manasterski A, Handsfield JH. Proficiency testing performance in US laboratories: results reported to the Centre for Medicare & Medicaid Services, 1994 through 2006. *Arch Pathol Lab Med* 2010;**134**:751–758.

International Organization for Standardization. Identification and determination of analytical and clinical performance goals for laboratory methodologies. ISO/TR 15196. Geneva: ISO, 2001.

International Standards Organization. Medical laboratories: particular requirements for quality and competence. ISO-15189. Geneva: International Organization for Standardization, 2003.

International Organization for Standardization. Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO 13528. Geneva: ISO, 2005a.

International Standards Organization. General requirements for the competence of testing and calibration laboratories. ISO/IEC-17025. Geneva: International Organization for Standardization, 2005b.

International Organization for Standardization. Conformity assessment—general requirements for proficiency testing. ISO/IEC 17043. Geneva: ISO, 2010.

Jequier AM. Is quality assurance in semen analysis still necessary? A cliniciańs viewpoint. *Hum Reprod* 2005;**20**:2039–2042.

Kenny DJ, Conway RM, Johnston DH. The development of ISO 9002 quality management standards for Canadian dental practices. *J Can Dent Assoc* 1999;**65**:105–108.

Killeen AA. Laboratory sanctions for proficiency testing sample referral and result communication: a review of actions from 1993-2006. *Arch Pathol Lab Med* 2009;**133**:979–982.

Kvist U, Björndahl L. Manual on basic semen analysis. In: Heineman MJ (ed). *ESHRE Monographs*. Oxford: Oxford University Press, 2002.

McCulloh DH. Quality control: maintaining stability in the laboratory. In: Gardner DK, Weissman A, Howles CM, Shoham Z (eds). *Textbook of Assisted Reproductive Techniques*. London, UK: Taylor & Francis, 2004, 25–39.

Palacios ER, Gonzalvo MC, Clavero A, Ramírez JP, Rosales A, Mozas J, Björndahl L, Martínez L, Castilla JA. Aplicación de las especificaciones de calidad analíticas al cálculo de los límites admisibles de error en Programas de Evaluación Externa de la Calidad para concentración espermática. *Rev Lab Clin* 2011;**4**:134–144.

Plaut DA, Westgard JOW. QC external quality assessment. In: Westgard JO (ed). *Basic QC Practices: Training in Statistical Quality Control for Healthcare Laboratories*. Madison: WI QC Publishing, 2002, 125–163.

Taylor A, Angerer J, Claeys F, Kristiansen J, Mazarrasa O, Menditto A, Patriarca M, Pineau A, Schoeters I, Sykes C *et al*. Comparison of procedures for evaluating laboratory performance in external quality assessment schemes for lead in blood and aluminum in serum demonstrates the need for common quality specifications. *Clin Chem* 2002;**48**:2000–2007.

World Health Organization. *Laboratory Manual for the Examination of Human Semen and Semen-Cervical Mucus Interaction*, 5th edn. Cambridge, UK: Cambridge University Press, 2010.

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*

FUJIFILM
Value from Innovation

IrvineScientific