Human Reproduction vol.15 no.11 pp.2360–2368, 2000

Intra- and inter-individual variability in human sperm concentration, motility and vitality assessment during a workshop involving ten laboratories

J.Auger[1,12], F.Eustache[1], B.Ducot[2], T.Blandin[3], M.Daudin[4], I.Diaz[5], S.El Matribi[1], B.Gony[6], L.Keskes[7], M.Kolbezen[8], A.Lamarte[3], J.Lornage[9], N.Nomal[10], G.Pitaval[9], O.Simon[11], I.Virant-Klun[8], A.Spira[2] and P.Jouannet[1]

[1]Service de Biologie de la Reproduction-CECOS, Hôpital Cochin, 75014 Paris, France, [2]INSERM U292, CHU de Bicêtre, 94276 Le Kremlin Bicêtre, France, [3]CECOS Aquitaine (Dr J.J.Berjon), Hôpital Pellegrin, 33076 Bordeaux, France, [4]Centre de Stérilité Masculine, CECOS Midi-Pyrénées (Dr L.Bujan), Hôpital La Grave, 31052 Toulouse, France, [5]Unidad de Fertilidad, Procreation Medicamente Asistida, Clinica Marly, Santafe de Bogota, Colombia, [6]CECOS Basse Normandie (Dr A.Sauvalle), CHU Côte de Nacre, 14033 Caen, France, [7]Laboratoire d'Histologie, Faculté de Médecine, Sfax, Tunisia, [8]Centere za andrologico (Dr B.Zorn), Gynekološka klinika, Univerzitetni Klinični Center, Ljubljana, Slovenija, [9]Laboratoire de Biologie de la Reproduction, CECOS Lyon (Pr J.F.Guérin), Hôpital Edouard Herriot, 69373 Lyon, France, [10]Service d'Anatomie Pathologique, Antenne CECOS (Dr M.Roudier), CHU de Pointre à Pître, 97159 Guadeloupe, France and [11]Laboratoire de Biologie de la Reproduction, CECOS Ouest (Pr D.Le Lannou), Hôpital de l'Hôtel-Dieu, 35000 Rennes, France

[12]To whom correspondence should be addressed at: Laboratoire d'Histologie, Biologie de la Reproduction, CECOS, and Groupe d'Etude de la Fertilité Humaine (GREFH), Université Paris V, Pavillon Cassini, Hôpital Cochin, 123 Bd de Port-Royal, 75014 Paris, France. E-mail: jacques.auger@cch.ap-hop-paris.fr

The aim of the present study was to assess variability in the evaluation of human sperm concentration, motility and vitality. Technicians and biologists from 10 teams involved in multicentre studies on semen quality attended the same laboratory, each team using its own methods and equipment to analyse the same semen samples. Inter-individual variability was assessed from 17 fresh semen samples of varying quality. Intra-individual variability was assessed from pools of frozen samples for sperm concentration and motility and stained smears for vitality with three blind evaluations by sample and smear. The mean inter-individual coefficients of variation were 22.9, 21.8 and 17.5% for sperm concentration, motility and vitality respectively. There was no statistical difference among participants for sperm concentration assessment, but significant differences for both motility and vitality (both $P < 0.05$). The mean intra-individual coefficients of variation were 15.8, 26.2 and 13.1% for sperm concentration, motility and vitality respectively, with marked differences between expert and novice participants: concentration 9.8% versus 28.0%; motility 22.8% versus 33.0%; and vitality 10.0% versus 19.3%. The present data confirm the need for external quality control schemes for diagnostic purposes, and indicate their utmost importance in multicentre studies on semen quality.

Key words: quality control/routine semen analysis

Introduction

The routine evaluation of human semen quality is subjective by nature. A high variability in the assessment of sperm characteristics is generally considered unacceptable, though very little is done to correct it. The laboratory methods used are not fully standardized, despite the World Health Organization publishing guidelines in successive editions (1980 to 1999) of its *Manual for Routine Semen Analysis* (WHO, 1999). Moreover, it is known that relatively important differences in methodology exist between laboratories which claim that they follow the WHO recommendations. In order to evaluate the source(s) and extent of errors in routine semen analysis, internal quality control (IQC) has been recently introduced in the reproductive biology laboratory (Mortimer *et al.*, 1986; Knuth *et al.*, 1989; Cooper *et al.*, 1992; Mortimer 1994; Clements *et al.*, 1995). Such IQC is essential to maintain accuracy, precision and competence, and it is a prerequisite for correct appreciation and interpretation in the diagnostic process of human infertility. External quality assessment (EQA) is the evaluation of results for the same samples in several laboratories. Using such an approach, a broad disagreement between the routine assessment of semen by different laboratories has been recently reported (Neuwinger *et al.*, 1990; Matson, 1995). A great inter-laboratory variability of the results of semen analysis is not without consequence for the patient: based on the semen sample, he might be classified as normal by one laboratory, yet infertile by another. Therefore, EQA is also required to ensure that different laboratories produce comparable results which in turn allow similar assessment of the probability of conception, or advice on appropriate treatment for couples under investigation, whichever centre is managing the infertility problem.

Recently published retrospective studies indicate secular and geographical differences in semen quality (Carlsen *et al.*, 1992; Auger *et al.*, 1995; Fédération CECOS *et al.*, 1997; Swan *et al.*, 1997). However, there are many possible methodological biases which prevent the drawing of final conclusions. Among them, laboratory skews such as variability in semen analysis procedures and assessments have been mentioned (Brake and Krause, 1992; Tumon and Mortimer, 1992). Considering forthcoming studies in this area, it is of special importance to evaluate whether differences in semen quality are real, or reflect differences in measuring methods. Therefore, EQA

and IQC—which are complementary processes—should be performed in the time course of this type of investigation on semen quality.

In the few EQA schemes reported previously (Neuwinger *et al.*, 1990; Matson, 1995; Cooper *et al.*, 1999), samples of prepared semen were sent to the participating laboratories. Such essential practice provides the opportunity for individual laboratories to evaluate grossly their own methods against those of others. However, one limitation is the fact that the assessment of semen cannot be performed on native samples, or under the usual conditions of semen analysis. For example, assessment of sperm motility requires either frozen material to be diluted in a cryoprotectant, or is made by video recordings.

On the initiative of the Paris group, technicians and biologists involved in prospective multicentre studies on sperm production and quality were invited to join for one week at the Reproductive Biology Laboratory, Hospital Cochin in Paris, in order to analyse native or prepared semen samples of various quality. This offered the possibility of an in-depth assessment of their intra- and inter-individual variability in monitoring of sperm concentration, motility and vitality.

Materials and methods

Study participants

The 13 participants working in 10 centres were laboratory technicians (t) and biologists or physicians (b), mostly from sperm banks of the French network CECOS (Centre d'Etude et de Conservation des Œufs et du Sperme humains) [T.B. (t), B.G. (t), O.S. (t), A.L. (t) and M.D. (b)], but also from andrology or Assisted Reproductive Technology laboratories in metropolitan France [G.P. (t), J.L. (b) and S.E.M. (t)] and abroad [N.N. (t), I.D. (b), M.K. (b), I.V.-K. (b) and L.K. (b)]. For data analysis, an identification number was assigned to each participant. However, two participants from one centre made all the semen analyses alternately or jointly at the same microscope. Since they had very similar results for the three sperm characteristics studied, they were assimilated to a single participant for data analysis. Consequently, in the results section, data are reported for only 12 participants.

Experimental design

The assessment of intra- and inter-individual variability in routine semen analysis was made from semen samples obtained from healthy donors and infertile patients who gave informed consent for participation. All semen samples were collected by masturbation in the laboratory after 3–5 days of sexual abstinence. The semen characteristics evaluated in the present study were sperm concentration, the percentage of motile spermatozoa, and the percentage of living spermatozoa. Only the overall motility (grades a + b + c; World Health Organization, 1992) was considered in data analysis because only five teams had a separate evaluation of the four WHO grades. No participant followed rigorously the WHO guidelines for routine semen analysis, and there were some differences in procedure among centres, as summarized in Table I. Except for the microscopes, each participant used their own equipment, e.g. counting chamber, diluents, pipettes and tips, dyes, and followed their usual working method.

The assessment of inter-individual variability was made from the analysis of 17 fresh semen samples collected by the healthy donors and infertile patients during the time course of the workshop. This allowed coverage of a wide range of values for the three characteristics studied. Once collected, the samples were kept at 37°C for 1 h before assessment by the different participants of the workshop. Twelve aliquots of equal volume (150 µl) of the semen samples were taken, and distributed to the participants according to a pre-established order of distribution in order to avoid bias related to time in the evaluation of sperm motility.

Since the volume of the semen sample was not large enough, intra-individual variability could not be assessed from evaluations made on native material. For a blind evaluation, the samples for intra-individual assessment were coded and distributed at even intervals of time during the entire week. Sperm concentration was assessed from pools of five frozen samples kept at –20°C without cryoprotectant. Each participant made three evaluations per sample. Each participant evaluated at random the percentage of motile spermatozoa three times in five pools of frozen straws kept at –196°C with a cryoprotectant added. All straws for the motility assessment were thawed for 10 min at 37°C before the analysis. The percentage of live spermatozoa was assessed on five slides from patients previously used for IQC, with three evaluations per sample for each participant. Eosin–nigrosin-stained smears were prepared according to WHO procedures (WHO, 1992).

Data analysis

Inter-participant variability

Inter-participant variability in the assessment of sperm concentration and the percentages of motile and live spermatozoa was expressed as the coefficient of variation: CV (%) = 100×SD/mean value. A random effect model (SAS mixed model software; SAS Institute Inc. Cary, NC, USA) was used to compare the values found by each participant for the three sperm characteristics studied. Correlation (Spearman's rank correlation test) was used to assess whether the inter-participant variability in the evaluation of sperm concentration and the percentages of motile and live spermatozoa were related to the average values of these characteristics. Bland–Altman plots (Bland and Altman, 1986) were used to illustrate the differences to the mean (%) for each participant and the 17 semen samples studied.

Intra-participant variability

For the three sperm characteristics, intra-participant variability was expressed as the coefficient of variation: CV (%) = 100×SD/mean value.

Influence of training

The participants were allocated to two groups according to their level of practice in order to assess the possible role of training. The first group included eight participants who had a daily practice of semen analysis, and at least 3 years experience. The second group included four participants with recent training and/or episodic semen analysis practice. According to these two groups, the differences in inter-individual variability were assessed by classifying the participants into three categories: (i) exact and accurate; (ii) exact and inaccurate or inexact and accurate; and (iii) inexact and inaccurate. The thresholds chosen for exactness were an average difference from the mean (%) for the 17 samples studied ⩽15% for sperm concentration and the percentage of motile spermatozoa, and ⩽10% for the percentage of live spermatozoa. The thresholds chosen for accuracy were an average SD of the difference to the mean for the 17 samples studied ⩽10% for the three sperm characteristics. The differences in intra-individual variability were expressed as the mean of the intra-individual CV in both groups. After data analysis, an individual detailed report with recommendations was sent to each participant; this allowed them to

Table I. Semen analysis procedures used by the 10 participating laboratories

Characteristic	Equipment or procedure	Number of teams
Sperm concentration	Type of counting chamber	
	Neubauer[a]	1
	Malassez	5
	Thoma	3
	Makler	1
	Type of dilution	
	1/2 to 1/100	1
	1/2 to 1/20	1
	1/10–1/20	2
	1/20	5
	No dilution	1
	Positive displacement pipette for dilution	
	Yes[a]	2
	No	8
	Cells counted	
	Mature germinal cells with tail[a]	1
	Mature germinal cells with tail + tail-less heads	9
% Motile spermatozoa	Depth of sample	
	20–30 μm[a]	3
	<20 μm	2
	>30μm	4
	Variable	1
	Final magnification	
	×200 and ×400[a]	4
	×200	1
	×100 and ×400	1
	×400	4
	Temperature	
	37°C[a]	8
	Room temperature	2
	Grading	
	a/b/c/d[a]	5
	a+b/c/d	2
	a/b+c/d	2
	a+b+c/d	1
% Live spermatozoa	Staining method	
	Eosin–nigrosin[a]	10
	Counting on smear	
	Yes[a]	6
	No[b]	4
	Final magnification	
	×400	5
	×1000	5
	Number of spermatozoa counted	
	100[a]	8
	200	2

[a]Equipment or procedure recommended in the WHO Manual (World Health Organization, 1992).
[b]From a 10 μl drop of the mixture between slide and coverslip.

evaluate their own results in comparison with the mean values obtained by the group.

Results

Inter-participant variability

The mean values and means of percentage of variation between each participant and the mean of the 12 individuals for the three sperm characteristics are summarized in Table II. The overall mean coefficients of variation for the 12 participants and the 17 samples studied were 22.9% for sperm concentration, 21.8% for the percentage of motile spermatozoa, and 17.5% for the percentage of live spermatozoa. There was no significant difference among participants for sperm concentration assessment, while the evaluations of the percentage of motile spermatozoa and the percentage of live spermatozoa were found to differ significantly, using the random model effect ($P < 0.05$ for both). There was a positive significant correlation between the mean value of sperm concentration obtained by the 12 participants and the SD ($r = 0.97$, $P < 0.001$; Figure 1a), indicating the lowest homogeneity in the assessment for the highest values, and also the converse. A negative, non-significant correlation was found between the mean value of the percentage of motile spermatozoa obtained by the 12 participants and the SD ($r = –0.38$, $P = 0.14$; Figure 2a). A significant negative correlation was found between the mean value of the percentage of live spermatozoa and SD ($r = –0.76$, $P < 0.001$; Figure 3a), indicating greatest

Table II. Inter-individual variability in sperm concentration, motility and vitality for 17 fresh semen samples

Participant	Sperm concentration		% Motile spermatozoa		% Live spermatozoa	
	Mean value ($\times 10^6$/ml)[a]	Mean variation[b]	Mean value (%)[a]	Mean variation[b]	Mean value (%)[a]	Mean variation[b]
1	72.2	−6.9	48.8	−5.9	73.8	9.8
2	77.8	2.7	46.8	−10.1	73.3	9.6
3	51.0	−30.0	63.7	20.5	60.8	−11.5
4	85.0	12.3	53.8	3.4	65.9	−2.6
5	82.4	11.4	52.1	−3.3	59.9	−13.6
6	72.0	−9.1	49.1	−8.2	76.5	13.3
7	67.6	−12.4	46.5	−14.1	54.6	−20.3
8	73.8	15.4	45.8	−15.9	55.4	−19.2
9	78.3	−0.8	47.4	−11.0	76.5	14.2
10	70.4	1.0	54.5	2.1	78.9	17.8
11	82.4	3.1	64.2	22.8	65.9	−2.2
12	81.5	4.7	61.2	17.6	69.7	4.0
Mean	74.6		52.9		67.6	
CV (%)[c]		22.9		21.8		17.5

[a]Mean value of the 17 samples analysed.
[b]Mean for the 17 samples of the percentage of variation between the estimation of each sample by the participant and the mean calculated from the values measured by all 12 participants.
[c]Mean of the 17 coefficients of variation (CV = 100×SD/mean value) corresponding to the evaluation by all 12 participants.

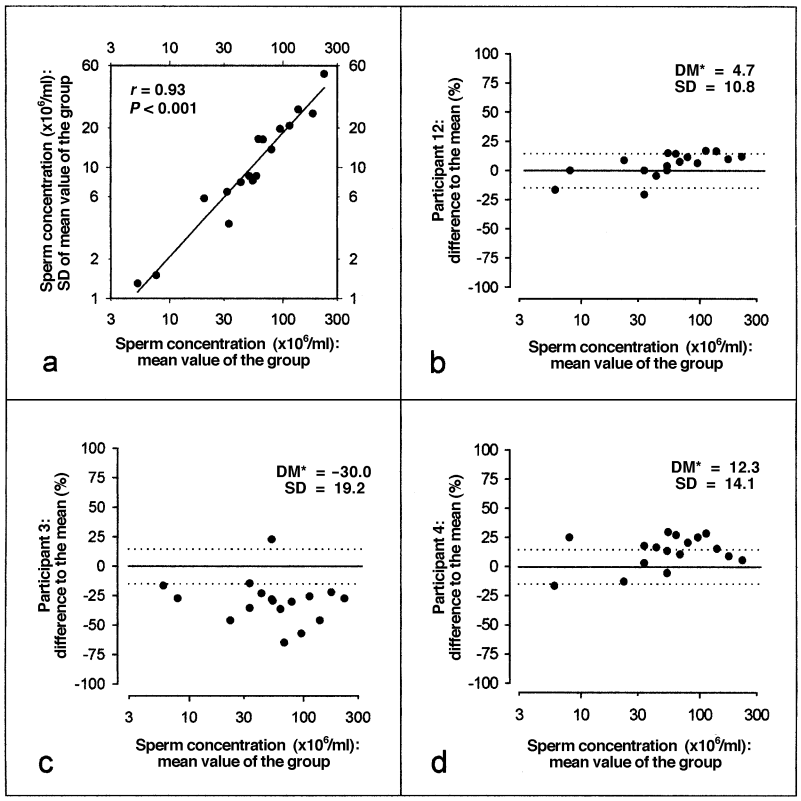

Figure 1. Inter-individual variability in sperm concentration. (**a**) Relationship between SD and mean value of sperm concentration calculated by the 12 participants for the 17 semen samples analysed; (**b**, **c**, **d**) Bland–Altman plots of three typical profiles of sperm count assessment by three different participants according to the mean value for the group. (**b**) Participant 12 is counting in the 15% interval around the mean value; (**c**) participant 3 is counting low; (**d**) participant 4 is counting high.

homogeneity in the assessment for the highest values, and also the converse. Among participants, and with regard to the mean values of the group, some participants provided superimposable results (Figures 1b, 2b and 3b), while some had a tendency to evaluate systematically low (Figures 1c, 2c and 3c) or systematically high (Figures 1d, 2d and 3d). In the present

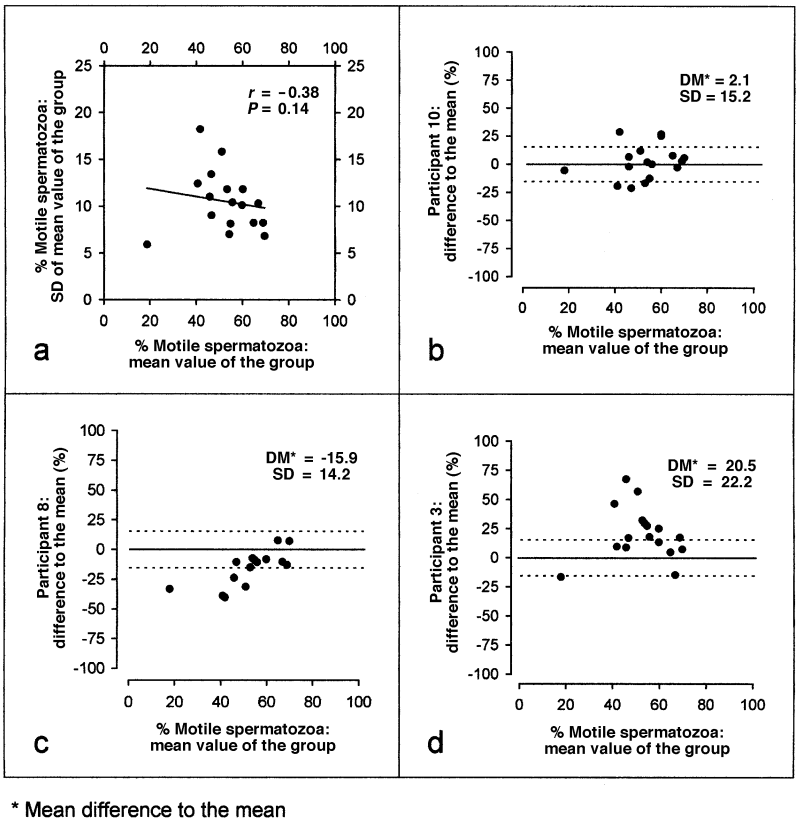

Figure 2. Inter-individual variability in the percentage of motile spermatozoa. (**a**) Relationship between SD and mean value of percent motility calculated by the 12 participants for the 17 semen samples analysed. (**b**, **c**, **d**) Bland–Altman plots of three typical profiles of motility assessment according to the mean value for the group. (**b**) Participant 10 is evaluating in the 15% interval around the mean value for a majority of samples; (**c**) participant 8 is evaluating low; (**d**) participant 3 is evaluating high.

study, no deviations from the mean values or wider intra-individual variations were found which could be related to the equipment or procedure used.

Intra-participant variability

For each participant, a mean coefficient of variation was calculated which measured the extent of variation among the three evaluations for the five frozen–thawed semen samples studied blindly for sperm concentration and motility, and also among the three evaluations for the three smears studied for vitality (Figure 4). The average values of CV for the 12 participants were 15.8% for the evaluation of sperm concentration (despite noticeable variation among participants), 26.2% for the percentage of motile spermatozoa (despite homogeneous variation among participants), and 13.1% for the evaluation of the percentage of live spermatozoa.

Influence of training

Inter- and intra-participant variability in sperm concentration and percentages of motile and living spermatozoa evaluations categorized according to the training of the participants are shown in Table III. There were marked differences in the inter- and intra-individual variability (although not significant due to the low sizes of the groups) between both groups of participants according to their level of experience and training. The intra-participant mean CV for the group of participants with episodic practice or low training were 2.9-fold higher for sperm concentration, 1.4-fold higher for the percentage of motile spermatozoa, and 1.9-fold higher for the percentage of live spermatozoa.

Discussion

Semen analysis is important in the diagnosis of male infertility (Rowe *et al.*, 1993) and for measuring the influence of xenogenic factors on male genital tract function (Wyrobek, 1983). Even when standardized according to recommendations of national or international organizations (WHO, 1999), the methods used by most laboratories remain very subjective. Therefore it has been recommended that internal and external quality controls should be developed, and the variations observed between and within the persons performing semen analysis be quantified (Mortimer, 1994; Michelmann, 1997; De Jonge, 1998; WHO, 1999). Both are necessary to find the causes of the differences and to assess the influences of those differences on sperm evaluation which may act as confounding factors in the evaluation of temporal and geographical variations in semen quality.

A mean inter-individual CV for sperm concentration of 22.9% was found for the 12 participants and the 17 semen samples studied. There was no significant difference in the values of sperm concentration obtained by the different participants using different dilution methods and counting chambers. In an external quality control study (Neuwinger

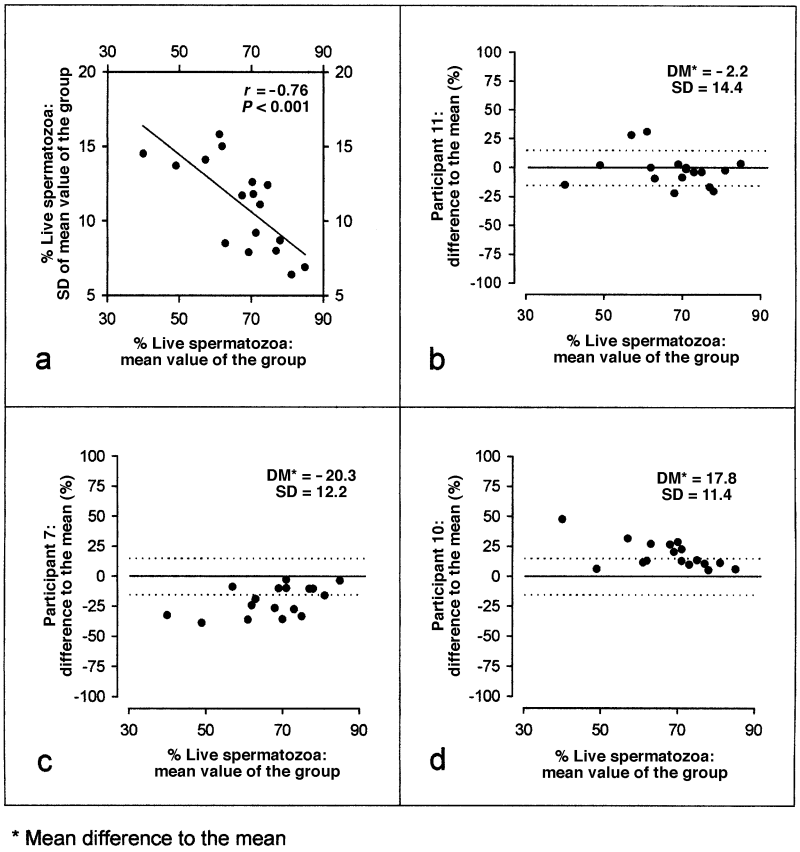

Figure 3. Inter-individual variability in the percentage of living spermatozoa. (**a**) Relationship between SD and mean value of percent live spermatozoa calculated by the 12 participants for the 17 semen samples analysed. (**b**, **c**, **d**) Bland–Altman plots of three typical profiles of sperm vitality assessment according to the mean value for the group. (**b**) Participant 11 is counting in the 15% interval around the mean value for a majority of samples; (**c**) participant 7 is counting low; (**d**) participant 10 is counting high.

et al., 1990), which included 10 experienced German laboratories for the evaluation of eight sperm samples, the mean CV was 37.5%. This result was obtained despite the study being carried out on clean preparations of spermatozoa selected by swim-up—a condition that is not normally applied for routine semen analysis. From the data of the EQA made under the auspices of the British Andrology Society reported previously (Matson, 1995), the mean inter-individual coefficient of variation for sperm concentration assessment was calculated to be 64.7% for the technicians from the 20 laboratories which were supposed to be trained for routine semen analysis and who evaluated 24 semen samples. In the current study, and in the German and British studies, the samples studied covered a wide range of sperm concentrations. The observed differences in CV might reflect a more important disparity in the equipment and procedure steps used for sperm concentration measurement in the British and German studies, which unfortunately were not reported in the publications. The differences in the British and German studies could also have resulted from additional factors of variation related to the mailing of the samples. The time between the collection and preparation of samples and their analysis might lead to the biological material being damaged. From the current study, it could be postulated that when the sperm concentration assessments are made on fresh samples, the inter-individual CV is lower than previously reported. Two studies have reported results of workshops organized on a similar principle to the present study. In the first (Jequier and Ukombe, 1983), 26 technicians and pathologists from medical laboratories participated, and a mean inter-individual CV of 44.3% for sperm concentration was found. However, only a single semen sample was studied (mean value 46.7×10^6 spermatozoa/ml; range: $10–98\times10^6$ spermatozoa/ml). In the second study (Jorgensen *et al.*, 1997), technicians from four experienced teams involved in research on geographical variations of semen quality joined for 1 week to analyse 26 semen samples. Despite the use of different equipment and procedures, and that the mean inter-individual CV for sperm concentration was not provided, the authors concluded that there was a remarkable consistency between teams for the vast majority of samples studied. From the present study, it could be said that deviations from the mean values or the intra-individual variations were not dependent on the equipment used or the procedure followed (data not shown), and that daily practice and training are important modulators of the variations observed between laboratories. However, the unexpected result of greater inter-participant variations for high concentrations (Figure 1a) rather than lower variations (WHO, 1999) despite a greater number of spermatozoa being counted by most participants, suggested that the different counting chambers used, as well as the different dilutions applied for high concentrations or the different pipettes used for dilution, contributed to this

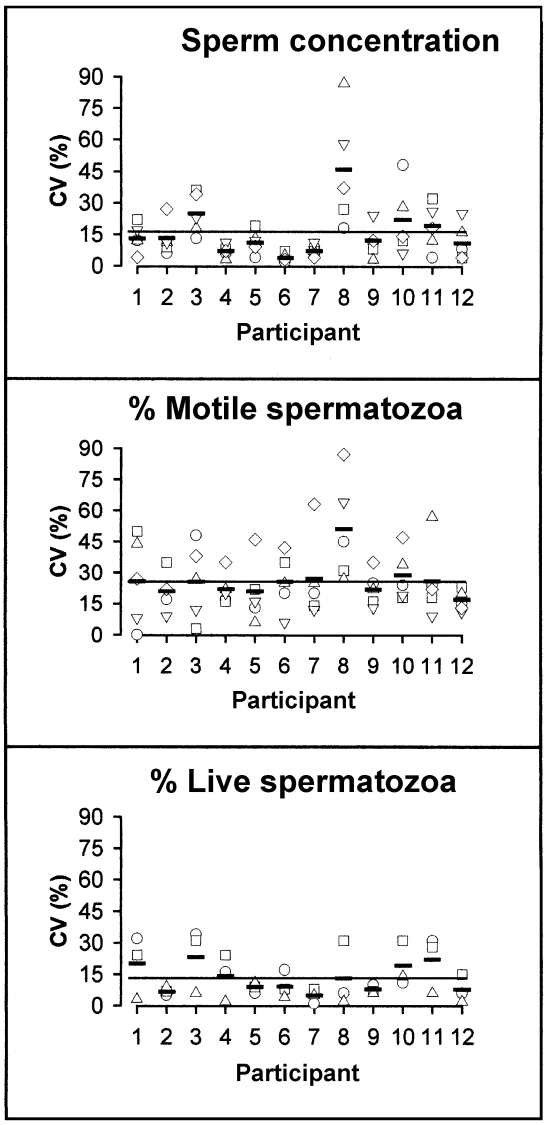

Figure 4. Intra-individual variability [coefficient of variation, CV (%)] in the evaluations of sperm concentration, motility and vitality. Five samples (□, ○, ◇, ▽, △) were analysed blindly three times for sperm concentration and motility by each of the 12 participants. Three slides (□, ○, △) were analysed blindly three times for sperm vitality by each of the 12 participants. Results are expressed as the CV for each sample, the mean CV for the five samples for each participant (–) and for all participants (continuous line).

higher variation. This result illustrated the utmost need for standardized methods to minimize variations in sperm counting among laboratories. The present study also suggested that EQA using the same semen samples evaluated by various people at the same time lowers variation compared with EQA using biological materials sent to various laboratories.

The inter-individual CV of sperm motility assessment was 21.8%, and therefore very similar to the CV found in an earlier study (Neuwinger *et al.*, 1990). Since the assessments of overall motility in this last study were made from material frozen with a cryoprotectant (which makes the evaluation more difficult), it might be supposed that the methodologies used were more homogeneous and/or the participants more trained. Very wide variations in the evaluations of motility were found in a more recent study (Jorgensen *et al.*, 1997), where the methodologies for sperm motility assessment were heterogeneous. In the present study, there was also an important disparity in the methodology for the assessment of sperm motility. Sperm motility assessment is clearly influenced by the temperature or the depth of the chamber used (Le Lannou *et al.*, 1992; Kraemer *et al.*, 1998). However, there is no *a priori* reason that this could influence markedly the estimation of the overall motility (a + b + c WHO grades). Therefore, the major factors of variation are probably related to the amount of training of the observer: the results of the present trial for intra-individual variability revealed that experienced participants had a CV of 22.8% compared with 33.0% for participants recently trained and/or with episodic practice. However, it should be pointed out that these values expressed the overall within-participant variation for all participants: in Figure 4, it can be seen that there were quite important differences in intra-individual variation among participants, and from one sample to another. Nevertheless, intra-observer variability in assessing sperm motility appeared to be related to the amount of training of the observer. Low intra-individual variation in the evaluation of sperm motility (CV $\leqslant$15%) was reported for highly trained technicians from the same laboratory (Neuwinger *et al.*, 1990; D.Mortimer, personal communication). However, better reproducibility in the assessment of sperm motility could also depend on the natural ability of the observer for this subjective task, as was suggested in an earlier study (Dunphy *et al.*, 1989).

The current study appears to be the first to report results of quality control in the assessment of the percentage of live spermatozoa. Due to the principle of the test of vitality (immobilized spermatozoa, with or without staining) and its quantitative nature, a low variability was expected. The lowest inter- and intra-individual CV were found for this characteristic (17.5% and 13.1% respectively) in comparison with CV found for the two other sperm characteristics studied. This result was obtained despite there being small variations in procedures among participants, or that in the intra-individual trial some participants evaluated this characteristic on smears, despite normally performing the test with a fresh drop of the stained semen deposited on a slide (see Table I). Because of the remarkable homogeneity found for the percentage of living spermatozoa (which can be further improved), percentage of living spermatozoa should be incorporated in repeated EQA (and of course IQC) schemes. Moreover, it could be useful to report percentage of living spermatozoa in studies on secular and geographical variation in semen quality because of the probably low confounding effect of its measure, and that this characteristic reflects the maturation of spermatozoa in the male genital tract, which in turn influences their survival in the female genital tract and their fertilizing ability.

Past and present EQA and IQC raised an unsolved question in the absence of highly reproducible methods to assess semen quality, namely, what is the target value? As has been proposed in the UK NEQAS (United Kingdom National Quality Control Assessment Schemes, Sheffield, UK) in Andrology, it can be decided that the mean value obtained from highly experienced laboratories is the reference value (Cooper *et al.*, 1999).

Table III. Inter- and intra-individual variability in sperm concentration, motility and vitality for participants re-grouped according to their level of experience and practice

	Inter-individual variability Number of participants:			Intra-individual variability Mean CV (%)
	Exact and accurate[a]	Exact and inaccurate or inexact and accurate[a]	Inexact and inaccurate[a]	
Sperm concentration				
Group A[b]	4	4	0	9.8
Group B[b]	0	2	2	28.0
% Motile spermatozoa				
Group A	1	6	1	22.8
Group B	1	1	2	33.0
% Live spermatozoa				
Group A	0	5	3	10.0
Group B	0	2	2	19.3

[a]The thresholds chosen for exactness were an average difference to the mean (%) for the 17 samples studied ≤15% for sperm concentration and the percentage of motile spermatozoa and ≤10% for the percentage of live spermatozoa; the thresholds chosen for accuracy were an average SD of the difference to the mean for the 17 samples studied ≤10% for the three sperm characteristics.
[b]Group A: technicians or biologists with daily practice of semen analysis and ≥3 years of training (n = 8); Group B: technicians or biologists with recent training and/or episodic semen analysis practice (n = 4).

However, previously reported studies (Neuwinger *et al.*, 1990; Jorgensen *et al.*, 1997) and the present study indicate that even experienced groups have a noticeable amount of disagreement for some characteristics. It has not been demonstrated that the mean value obtained by these teams provides the best reference point. Therefore, efforts should be made to develop reproducible objective methods in order to provide reliable target values, particularly for quality control schemes. There are some perspectives with the use of flow cytometry applied to sperm concentration assessment (Neuwinger *et al.*, 1990). Unfortunately, there is no current objective method which allows reproducible assessment of the percentage of motile spermatozoa. In particular, computer-assisted semen analysis (CASA), which is the sole technology offering the possibility to analyse sperm motion reliably (provided that there is rigorous control of all stages of the analysis; Kraemer *et al.*, 1998), has not proved to be superior to visual estimation in terms of reproducibility of results. Expert groups in andrology do not recommend the use of CASA to assess percent motility of spermatozoa (Mortimer *et al.*, 1995; ESHRE Andrology Special Interest Group, 1998). However, it should be pointed out that CASA might be very useful in quality control schemes to discriminate between the relative amounts of WHO grades a and b motile spermatozoa (Yeung *et al.*, 1997), since assessment of such spermatozoa is a major source of variability among individuals and laboratories (Dunphy *et al.*, 1989). Finally, no reproducible objective method has been proposed for the assessment of sperm vitality by microscopy, and the methods to distinguish between viable and non-viable cells using fluorescent dyes (e.g. propidium iodide) and flow cytometry applied to the evaluation of mammalian sperm viability (Garner *et al.*, 1986; Auger *et al.*, 1989) are not adapted for routine semen analysis. Furthermore, it was shown recently that, by using microscopy, vital staining with propidium iodide gave different results than staining with eosin–nigrosin (Pintado *et al.*, 2000).

Basic semen analysis courses and training are prerequisites for novices in the field to minimize their basal variability in assessment (Mortimer, 1994). Subsequently, regular IQC and EQA are needed to reduce the variability inherent to semen analysis practice, and therefore the differences between evaluations made by different laboratories. Discussion of the results with the biologists in charge of the laboratories is essential for motivating the participants and defining corrective measures if necessary. The positive effects of these measures have been reported previously (Björndahl and Kvist, 1998; Punjabi and Spiessens, 1998). The significant improvements in the evaluation of semen characteristics resulting from these strategies are particularly important in order to harmonize results between laboratories, and ultimately for the management of infertile couples.

IQC is also required for intra-centre studies on temporal trends in semen quality to provide evidence that the observed variations are real, and that a better agreement in semen assessment made by various laboratories is also the basis for validating conclusions of multicentre studies on differences in semen quality. Therefore, any future prospective study in this field should be based on standardized methods and should include internal and/or external quality assessments, depending on the type of study. For planned multicentre studies, a prestudy EQA should be performed, followed by corrective measures if necessary, as in a recent study of geographical variation in semen quality in Europe (Jorgensen *et al.*, 1997). This is very useful when the same characteristic cannot be analysed centrally, as may be done for sperm morphology. Moreover, the initial EQA should be followed by repeated quality control in the complete time course of the study in order to identify any possible deviation in assessment. Such approaches offer the opportunity to adjust data in the statistical analysis for taking into account the variations related to methodological factors.

Acknowledgements

We would like to thank Françoise Roques and Michel Doudeau for their excellent technical assistance during the workshop. The study

was supported by contract EN96C1 from Ministère de l'Environnement and INSERM (Institut National de la Santé et de la Recherche Médicale) and a research grant (1752) from Direction des Recherches, Etudes, et Technologies, Ministère de l'Education Nationale.

References

Auger, J., Ronot, X. and Dadoune, J.P. (1989) Human sperm mitochondrial function related to motility: a flow and image cytometric assessment. *J. Androl.*, **10**, 439–448.

Auger, J., Kunstmann, J.M., Czyglik, F. and Jouannet, P. (1995) Decline in semen quality among fertile men in Paris during the last 20 years. *N. Engl. J. Med.*, **332**, 281–285.

Björndahl, L. and Kvist, U. (1998) Basic semen analysis courses: experience in Scandinavia. In Ombelet, W., Bosmans, E., Vandeput, H. *et al*. (eds), *Modern ART in the 2000s. Andrology in the Nineties*. Parthenon Publishing Group, New York, pp. 91–101.

Bland, J.M. and Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **i**, 307–310.

Brake, A. and Krause, W. (1992) Decreasing quality of semen (letter). *Br. Med. J.*, **305**, 1498.

Carlsen, E., Giwercman, A., Keiding, N. and Skakkebaek, N.E. (1992) Evidence for decreasing quality of semen during past 50 years. *Br. Med. J.*, **305**, 609–613.

Clements, S., Cooke, I.D. and Barratt, C.L. (1995) Implementing comprehensive quality control in the andrology laboratory. *Hum. Reprod.*, **10**, 2096–2106.

Cooper, T.G., Neuwinger, J., Bahrs, S. and Nieschlag, E. (1992) Internal quality control of semen analysis. *Fertil. Steril.*, **58**, 172–178.

Cooper, T.G., Atkinson A.D. and Nieschlag E. (1999) Experience with external quality control in spermatology. *Hum. Reprod.*, **14**, 765–769.

De Jonge, C. (1998) Total quality management and the clinical andrology laboratory: essential partners. In Ombelet, W., Bosmans, E., Vandeput, H. *et al*. (eds), *Modern ART in the 2000s. Andrology in the Nineties*. Parthenon Publishing Group, New York, pp. 55–60.

Dunphy, B.C., Kay, R., Barratt, C.L.R. and Cooke, I.D. (1989) Quality control during the conventional analysis of semen, an essential exercise. *J. Androl.*, **10**, 378–385.

ESHRE Andrology Special Interest Group (1998) Guidelines on the application of CASA technology in the analysis of spermatozoa. *Hum. Reprod.*, **13**, 142–145.

Fédération CECOS, Auger, J. and Jouannet, P. (1997) Evidence for regional differences of semen quality among fertile French men. *Hum. Reprod.*, **12**, 740–745.

Garner, D.L., Pinkel, D., Johnson, L.A. and Pace, M.M. (1986) Assessment of spermatozoal function using dual fluorescent staining and flow cytometric analyses. *Biol. Reprod.*, **34**, 127–138.

Jequier, A.M. and Ukombe, E.B. (1983) Errors inherent in the performance of a routine semen analysis. *Br. J. Urol.*, **55**, 434–436.

Jorgensen, N., Auger, J., Giwercman, A. *et al.* (1997) Semen analysis performed by different laboratory teams: an intervariation study. *Int. J. Androl.*, **20**, 201–207.

Knuth, U.A., Neuwinger, J. and Nieschlag, E. (1989) Bias to routine semen analysis by uncontrolled changes in laboratory environment – detection by long-term sampling of monthly means for quality control. *Int. J. Androl.*, **12**, 375–383.

Kraemer, M., Fillion, C., Martin-Pont, B. and Auger, J. (1998) Factors influencing human sperm kinematic measurements by the Celltrak computer-assisted sperm analysis system. *Hum. Reprod.*, **13**, 611–619.

Le Lannou, D., Griveau, J.F., Le Pichon, J.P. and Quero, J.C. (1992) Effects of chamber depth on the motion pattern of human spermatozoa in semen or in capacitating medium. *Hum. Reprod.*, **7**, 1417–1421.

Matson, P.L. (1995) External quality assessment for semen analysis and sperm antibody detection: results of a pilot scheme. *Hum. Reprod.*, **10**, 620–625.

Michelmann, H.W. (1997) Quality management in the andrology laboratory. *Int. J. Androl.*, **20** (Suppl. 3), 50–54.

Mortimer, D. (1994) Technician training and quality control aspects. *Practical Laboratory Andrology.* Oxford University Press, Oxford, pp. 337–347.

Mortimer, D., Shu, M.A. and Tan, R. (1986) Standardization and quality control of sperm concentration and sperm motility counts in semen analysis. *Hum. Reprod.*, **1**, 299–303.

Mortimer, D., Aitken, R.J., Mortimer, S.T. and Pacey, A.A. (1995) Workshop report: clinical CASA – the quest for consensus. *Reprod. Fertil. Dev.*, **7**, 951–959.

Neuwinger, J., Behre, H.M. and Nieschlag, E.(1990) External quality control in the andrology laboratory: an experimental multicenter trial. *Fertil. Steril.*, **54**, 308–314.

Pintado, B., de la Fuente, J. and Roldan, E.R.S. (2000) Permeability of boar and bull spermatozoa to the nucleic acid stains propidium iodide or Hoechst 33258, or to eosin: accuracy in the assessment of cell viability. *J. Reprod. Fertil.*, **118**, 145–152.

Punjabi, U. and Spiessens, C. (1998) Basic semen analysis courses: experience in Belgium. In Ombelet, W., Bosmans, E., Vandeput, H. *et al*. (eds), *Modern ART in the 2000s. Andrology in the Nineties*. Parthenon Publishing Group, New York, pp. 107–113.

Rowe, P.J., Comhaire, F.H., Hargreave, T.B. and Mellows, H.J. (1993) *WHO Manual for the Standardized Investigation and Diagnosis of the Infertile Couple*. Cambridge University Press, Cambridge, pp. 33–39.

Swan, S.H., Elkin, E.P. and Fenster, L. (1997) Have sperm densities declined? A reanalysis of global trend data. *Environ. Health Perspect.*, **105**, 1228–1232.

Tumon, A. and Mortimer, D. (1992) Decreasing quality of semen (letter). *Br. Med. J.*, **305**, 1228–1229.

World Health Organization (1992) *WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction*, 3rd edn, Cambridge University Press, Cambridge.

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

Wyrobek, A.J. (1983) Methods for evaluating the effects of environmental chemicals on human sperm production. *Environ. Health Perspect.,* **48**, 53–59.

Yeung, C.H., Cooper, T.G. and Nieschlag, E. (1997) A technique for standardization and quality control of subjective sperm motility assessments in semen analysis. *Fertil. Steril.*, **67**, 1156–1158.

Received on March 31, 2000; accepted on July 18, 2000