High variability in results of semen analysis in Andrology Laboratories in Tuscany (Italy): the experience of an external quality

ANDROLOGY

ISSN: 2047-2919

ANDROLOGY

ORIGINAL ARTICLE

Correspondence:
Mario Maggi, Elisabetta Baldi, Department of Clinical Physiopathology, Sexual Medicine and Andrology Unit, Florence, Italy. E-mails: m.maggi@dfc.unifi.it; elisabetta.baldi@unifi.it

Keywords:
external quality control, semen analysis, sperm concentration, sperm morphology, sperm motility

Received: 20-Jul-2012
Revised: 17-Oct-2012
Accepted: 18-Oct-2012

doi: 10.1111/j.2047-2927.2012.00042.x

High variability in results of semen analysis in Andrology Laboratories in Tuscany (Italy): the experience of an external quality control (EQC) programme

E. Filimberti,* S. Degl'Innocenti,* M. Borsotti,† M. Quercioli,† P. Piomboni,‡ I. Natali,§ M. G. Fino,* C. Caglieresi,¶ L. Criscuoli,** L. Gandini,†† A. Biggeri,‡‡ M Maggi*,1 and E. Baldi*

*Department of Clinical Physiopathology, Sexual Medicine and Andrology Unit, Florence, Italy, †Quality Laboratory Control Center, Azienda Ospedaliera-Universitaria Careggi, Florence, Italy, ‡Department of Biomedical Sciences, Applied Biology Section, University of Siena, Siena, Italy, §Sterility Center, UO Gynecology and Obstetrics, Hospital of Pescia, Pescia, Italy, ¶Medical School and Hospital, University of Pisa, Pisa, Italy, **Fertility Center of the Department of Sciences for Woman and Child's Health, University of Florence, Florence, Italy, ††Laboratory of Seminology-Sperm Bank, Department of Experimental Medicine, University of Rome "La Sapienza", Rome, Italy, and ‡‡Department of Statistics, University of Florence, Florence, Italy

SUMMARY

We report the results of the first three trials of an external quality control (EQC) programme performed in 71 laboratories executing semen analysis in Tuscany Region (Italy). At the end of the second trial, participants were invited to attend a teaching course illustrating and inviting to adhere to procedures recommended by WHO (V edition). Results of the first three trials of the EQC documented a huge variability in the procedures and the results. The highest variability was found for morphology (CV above 80% for all the trials), followed by count (CV of about 60% for all the trials) and motility (CV below 30% for all the trials). When results of sperm count and morphology were divided according to the used method, mean CV values did not show significant differences. CV for morphology dropped significantly at the third trial for most methods, indicating the usefulness of the teaching course for morphology assessment. Conversely, no differences were observed after the course for motility and for most methods to evaluate count, although CV values were lower at the second and third trial for the laboratories using the Burker cytometer. When results were divided according to tertiles of activity, the lowest mean bias values (difference between each laboratory result and the median value of the results) for count and morphology were observed for laboratories in the third tertile (performing over 200 semen analysis/year). Of interest, mean bias values for concentration dropped significantly at the third trial for low activity laboratories. In conclusion, lack of agreement of results of semen analysis in Tuscany is mainly because of the activity and the experience of the laboratory. Our study points out the importance of participating in EQC programmes and periodical teaching courses as well as the use of WHO recommended standardized procedures to increase precision and to allow the use of WHO reference values.

INTRODUCTION

Number, percentage and quality of motility and shape of spermatozoa are considered not only diagnostically significant but sometimes even normative for assessing the fertility status of the male. All these parameters can be evaluated by semen analysis. However, semen analysis is very poorly predictive of the male fertility status, mainly giving information about the status of male genital tract and thus only indirect indications of male fertility potential. Despite this fact, the diagnosis of male infertility is quite often based only on semen analysis, that is used by clinicians, in conjunction with analyses concerning the female partners, for health-care decisions regarding the appropriate assisted reproduction technique (ART) for the couple. Considering the elevated costs of ARTs, it is desirable that results of semen analysis are as accurate and precise as possible, male infertility is indeed becoming an epidemic problem affecting almost 5% of the male population (Jungwirth *et al.*, 2012). Another epidemic condition with a similar prevalence in general population is diabetes mellitus where evaluation of blood glycaemia has a normative value for diagnostic purposes (Hill *et al.*,

2011) as it occurs for semen analysis in male infertility. However, the percentage coefficient of variation (CV, the ratio between standard deviation and the mean ×100) for testing glycaemia is ranging from 2 to 4% (Harris *et al.*, 1998), whereas the CV for sperm analysis is greatly exciding 30%, according to previous surveys (Jorgensen *et al.*, 1997; Gandini *et al.*, 2000; Cooper *et al.*, 2002).

In Tuscany, there are 71 laboratories (33 in public and 38 in private structures) which perform about 20 000 semen analysis/year. Although clinical examination of the male partner of infertile couples is recommended by WHO clinical manual for the management of the infertile male (Rowe *et al.*, 2000), it is routinely performed only in a small subset of ART centres. Thus, the diagnosis of male factor resides just in sperm analysis in most ART centres.

In the ideal condition, each laboratory performing semen analysis should apply the same methodology and adhere to the same criteria for the different measurements. In theory, this condition could be ensured by following the primary reference for methods of semen analysis represented by the World Health Organization (WHO) laboratory manuals (WHO, 1980, 1987, 1992, 1999, 2010), which provided guidelines for standardized procedures in the last 30 years to allow comparison of data for clinical and research purposes. In addition, adherence to standardized methods allows association of data with reference values, which have been recently revisited by WHO (Cooper *et al.*, 2010; WHO, 2010). The questions that arise concern whether all the laboratories performing semen analysis in Tuscany adhere to WHO guidelines, and, even in this case, whether results obtained from the different technicians (even within the same laboratory) are really comparable, considering the subjectivity of the measurements. For these reasons, the necessity of external quality control (EQC) and standardization of the procedures has been highlighted in several studies (Neuwinger *et al.*, 1990; Jørgensen *et al.*, 1997; Gandini *et al.*, 2000; Cooper *et al.*, 2002; Keel *et al.*, 2002; Palacios *et al.*, 2012), and the last two editions of the WHO manual (1999 and 2010) dedicated an entire Chapter to quality assurance and quality control. EQC has the aim to detect and possibly reduce systematic errors and variability in results. With the exclusion of a pilot study (Gandini *et al.*, 2000) performed about 10 years ago and involving only 20 laboratories, in Italy there are no EQC programmes for semen analysis. The idea of performing an EQC for semen analysis among the Tuscany laboratories came to our mind because of several variants in the applied methods and in the reference values as could be evinced from reports produced in different labs. Here, we report the results of the first three trials of an EQC developed in the Tuscany region, thanks to funding from Tuscany government and Azienda Ospedaliera Universitaria Careggi (AOUC) and involving the 71 laboratories in the region that perform semen analysis. We have also evaluated the effect of a teaching course (performed between the second and the third trial) on variability in the results and in the techniques used by the participating laboratories.

MATERIALS AND METHODS

A request was sent to the 106 laboratories performing clinical biochemistry analysis in public and private structures in Tuscany asking whether they were performing semen analysis and, in this case, if they were willing to participate in an EQC trial. A total of 71 laboratories declared to perform semen analysis (33 in public and 38 in private structures) and all of them agreed to participate in the EQC.

Biological material for EQC programme was collected within the semen analysis Laboratory of the Unit of Sexual Medicine and Andrology of the University of Florence. Semen was collected, after informed consent, from healthy volunteers who underwent, before semen collection, serum analysis for detection of HIV, HBV and HCV and resulted negative for all the above markers. Semen samples (3 samples/trial) were mixed and, after formalin (1%, final concentration) addition and continuous gentle agitation in a rotator, aliquoted in vials of 250 μL and sent to the 71 laboratories for evaluation of sperm concentration and percentage of normal morphology. For sperm motility assessment, a DVD containing records from three different samples (from 6 to 8 fields/sample of 20–30 sec/field) was obtained using a digital video camera and sent to the laboratories. The images for DVD were collected at 40× magnification without any grid or scale for calibration. Participants were instructed to assess sperm parameters exactly as they were currently doing for their own clinical specimens. In addition, they were requested to indicate the method used to perform each analysis giving the choice among Neubauer, Burker, Cell-Vu, Makler cytometers or other methods (unspecified) for sperm count, manual or automatic methods for sperm motility, Diff-Quick, Papanicolau, May-Grunwald, Haematoxilin-Eosin, TestSimplets or others (unspecified) staining for morphology. At the end of the first trial, all participants were invited to a meeting to view the results. On that occasion, the differences between the WHO IV (1999) and V (2010) edition in evaluating sperm count, motility and morphology were illustrated to the participants, to whom the new manual (Italian version by Società Italiana di Andrologia e Medicina della Sessualità – SIAMS) was given. During the meeting, participants were recommended to follow WHO (2010) procedures to execute semen analysis. At the end of the second trial, participants were invited to a teaching course, during which they participated to a test for motility and morphology evaluation both at the entry and at the end of the course. Details of the methods of evaluation of the different parameters according to the WHO V edition were illustrated during the course.

A secret code was assigned to each laboratory for data processing. Results of each trial were collected by the Excel software. We obtained a response with the results from 52, 66 and 56 laboratories (of the 71 contacted), respectively, for the first, second and third trial. At the end of each trial, results of the performance of each laboratory were sent specifying, for each parameter, the mean value, the standard deviation and the coefficient of variation (CV%) of all the participating laboratories and of those performing the analysis with the same method of the laboratory if used by at least 5 laboratories. In addition, histograms of the distribution, marking in bold the result of the laboratory, were included.

Statistical analysis

Statistical analysis was performed using the programme Microcal Origin software 6.1 version (MicroCal Software Inc, Northampton, MA, USA, www.origin.com) to evaluate, for each parameter, the mean, standard deviation (SD) and median values of all the results together and after grouping on the basis of the used method for evaluating sperm count and sperm

morphology. Differences between the mean CV and mean difference between each laboratory result and the median value of the results expressed as percentage (from herein denominated 'bias') values were analysed using ANOVA followed by Bonferroni test to take into account the multiplicity of testing, considering statistically significant the values of $p < 0.05$/number of analysed groups.

RESULTS

Sperm concentration

Table 1A shows the different techniques used by the participating laboratories to evaluate sperm concentration during the three trials. The main tools used were the Makler and the Burker cytometers. Despite recommendation, at the end of the first trial and during the teaching course after the second trial, of using the improved Neubauer cytometer, only few laboratories are currently using such WHO recommended technique.

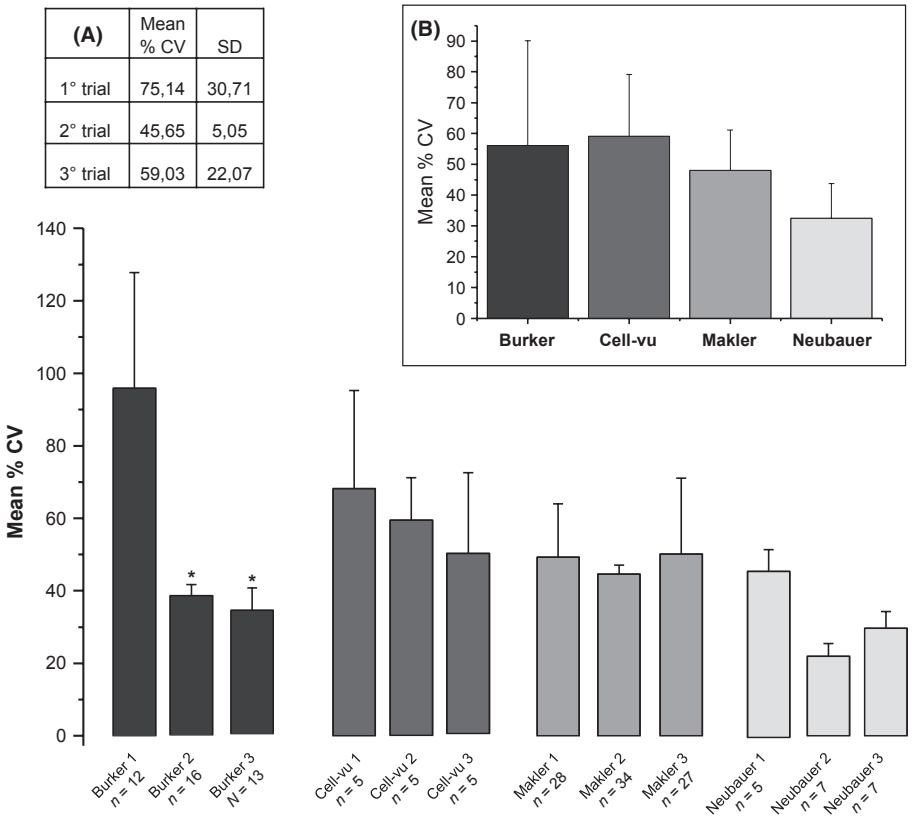
Mean CV ± SD values of sperm count of the three trials are reported in table A inserted in Fig. 1. No significant differences were observed in mean CV values from the first to the second and third trial. When data were divided according to the used technique (inset B of Fig. 1), the mean CVs showed no significant differences, although ANOVA analysis revealed a statistical significance among the different methods ($F = 2.98$, $p = 0.046$). A significant decrease in mean CVs was observed, at both second and third trials respect to the first for Burker cytometer, indicating an improvement of the performance, whereas no differences were observed for Cell-vu, Makler and Neubauer.

We calculated the mean bias value of each trial of the laboratories that took part in at least 2 trials. We found that several laboratories experienced a decrease in the bias value both at the second as respect to the first (30/49, 61%) and at third as respect

Table 1 Counting chambers and staining methods used by the laboratories participating in the EQC scheme for the evaluation of the sperm concentration (A) and the morphology (B) in the three trials. *N* indicates the number of laboratories

	1° Trial *N* (%)	2° Trial *N* (%)	3° Trial *N* (%)
A: Concentration			
Burker	12 (23.1)	16 (24.2)	13 (23.2)
Cell-vu	5 (9.6)	5 (7.5)	5 (8.9)
Neubauer	5 (9.6)	7 (10.6)	7 (12.5)
Makler	28 (53.8)	34 (51.5)	27 (48.2)
Others	2 (3.8)	4 (6.06)	4 (7.1)
Total	52	66	56
B: Morphology			
Diff quick	7 (13.5)	8 (11.9)	5 (8.9)
Haematoxilin/Eosin	4 (7.7)	6 (8.9)	6 (10.7)
May-Grunwald	8 (15.4)	16 (23.8)	14 (25.0)
Papanicolau	12 (23.1)	10 (14.9)	8 (14.2)
Fresh	2 (3.8)	1 (1.4)	1 (1.8)
TestSimplets	18 (34.6)	21 (31.3)	17 (30.3)
Others	1 (1.9)	5 (7.4)	5 (8.9)
Total	52	67	56

Figure 1 CV values of the trials for sperm count. Mean CV values of the trials (1, 2, 3) are shown according to the methods used by the participating Laboratories. *indicates significance ($p < 0.016$) within trials with the same cytometer. The inserted Table 1A reports mean ± SD overall CV values for the three trials. The inset B shows the mean CV value of the three trials for the different methods. Error bars represent SD values.



to the second trial (33/53, 62%). Interestingly, when results were divided according to tertiles of declared activity of the laboratories (number of semen analysis performed/year, see Fig. 2 for the distribution of laboratories activity), a significant decrease in the bias value was found at the third trial for the laboratories with low activity (<than 79 semen analysis/year) (Fig. 3A).

Sperm motility

All the participating laboratories declared to perform analysis of sperm motility manually. No differences were observed between mean CV values of percentage progressive [grade 'a', speed >25 μm/sec + grade 'b', speed < 25 μm/sec according to WHO manual IV Edition (WHO, 1999)], non-progressive [grade 'c', <5 μm/sec according to WHO manual IV Edition (WHO, 1999)] and immotile [grade 'd' according to WHO manual IV Edition (WHO, 1999)] spermatozoa of the three trials (Fig. 4). The highest CV values were observed for the non-progressive motility. Of note, mean CV for progressive motility and immotile spermatozoa was the lowest among all the parameters analysed in the EQC scheme. We calculated the bias of the different laboratories for percentage progressive motility, non-progressive motility and immotile spermatozoa of the three trials. For all types of motility there was a decrease in the mean bias value of the laboratories in the third respect to the second trial. In particular, 34/53 (64%) and 36/53 (67%) laboratories demonstrated an improvement, respectively, for progressive and non-progressive motility.

Figure 3B shows the mean ± SD bias values according to tertiles of laboratory activity for progressive motility. The higher bias values for progressive motility were observed for laboratories of the second tertile.

Figure 2 Activity of the different laboratories participating in the trials. Activity is expressed as number of semen analysis/year (ordinate, log scale).

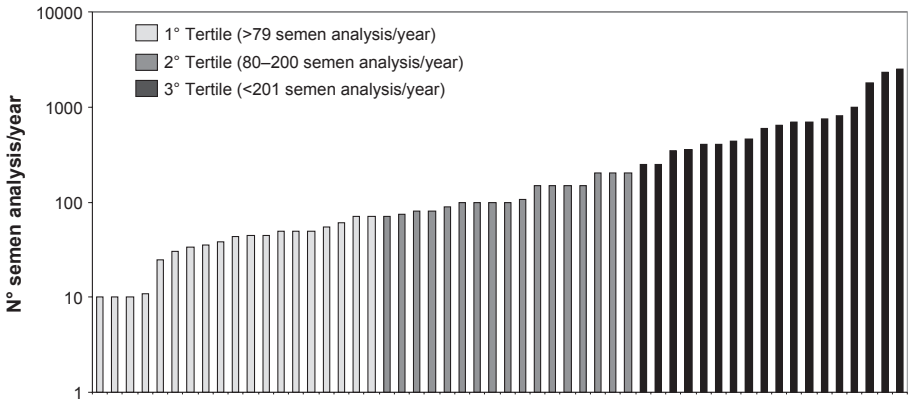


Figure 3 Bias values of the three trials in the different laboratories divided in tertiles of activity. Laboratories were divided in tertiles according to activity (number of semen analysis/year, abscissa), and the mean bias values for each trial for sperm concentration (A), progressive sperm motility (B) and sperm morphology (C) are reported on ordinates. *indicates significance ($p < 0.016$) respect to first trial for sperm concentration in the low activity tertile. °indicates significance ($p < 0.016$) respect to the same trial of first tertile of activity for sperm concentration. Error bars represent SD values.

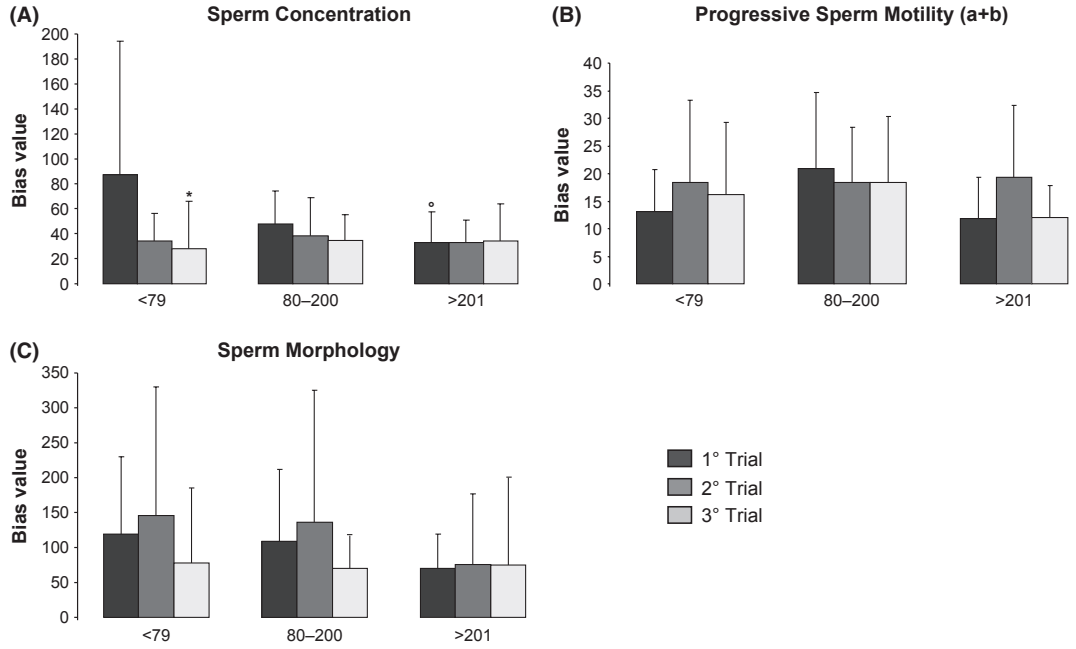
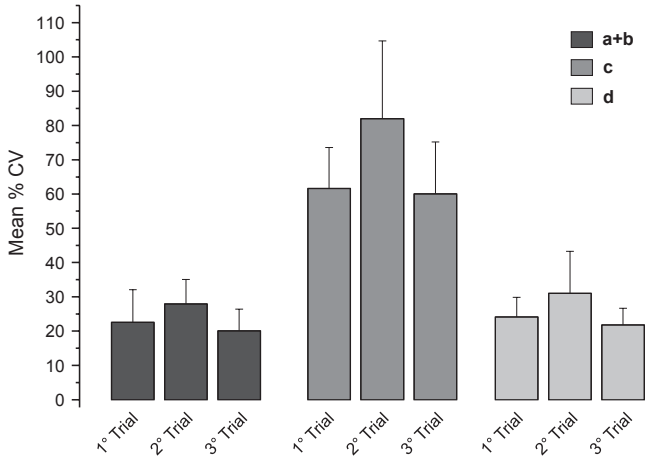


Figure 4 CV values of the trials of EQC for sperm motility. Mean CV values of the three trials are shown according to the type of motility [a + b: progressive motility; c: non-progressive motility; d: immotile spermatozoa according to WHO (1999)]. Error bars represent SD values.



Sperm morphology

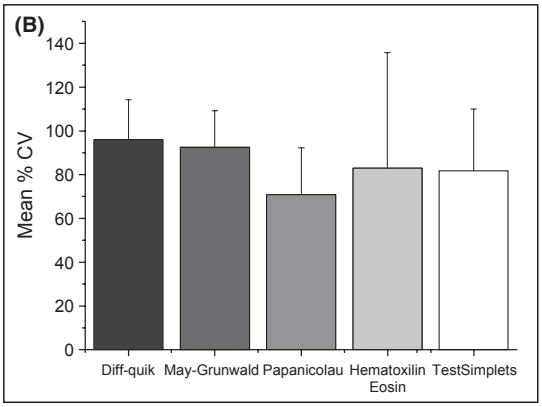
Table 1B reports the different techniques used by the participating laboratories to evaluate sperm morphology during the three trials. The most used method of evaluation resulted Test-Simplets, followed by Papanicolau and May-Grunwald staining.

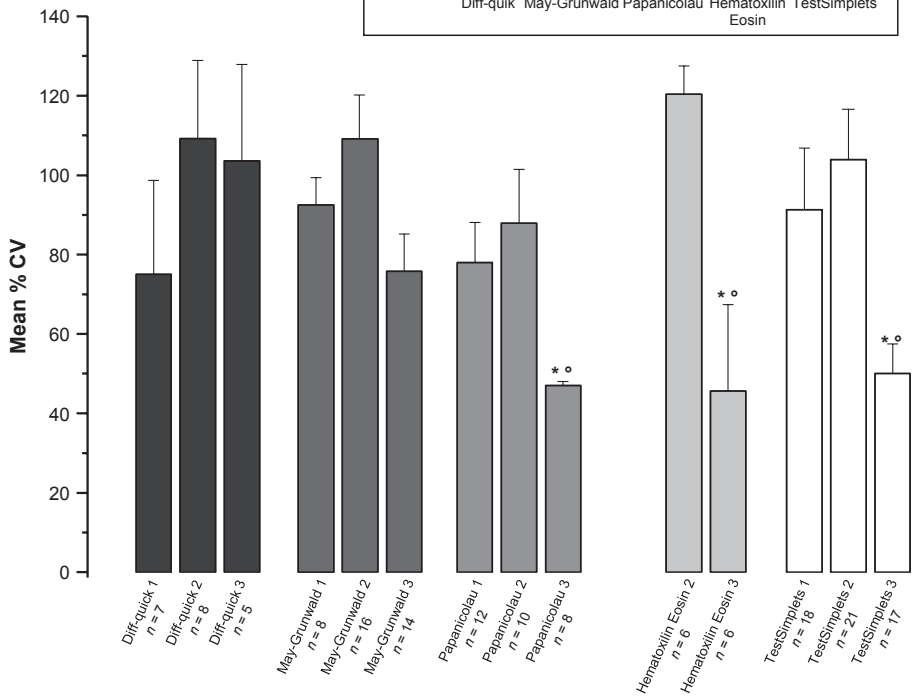
Mean ± SD CV values for sperm morphology of the three trials are shown in table A inserted in Fig. 5. No significant differences were present. Similarly, no differences were observed for mean CV values when the different methods of evaluation were considered (inset B of Fig. 5). Conversely, a significant decrease in mean CV values was observed in the third trial respect to the second for Papanicolau, Haematoxilin-Eosin and Testsimplet staining. Moreover, at the third trial, CV values of Papanicolau, Haematoxilin-Eosin and Testsimplet staining were significantly lower respect to those of Diff-Quick.

Thirty-one of 50(62%) laboratories demonstrated a decrease in the mean bias value in the second trial respect to the first and 26/50 (48%) in the third respect to the second. When results

Figure 5 CV values of the trials of EQC for sperm morphology. Mean CV values of the trials (1, 2, 3) are shown according to the methods used by the participating Laboratories. *indicates significance ($p < 0.016$) within trials with the same method. °indicates significance ($p < 0.016$) of Diff-Quick vs. Papanicolau, Haematoxilin-Eosin and TestSimplets at the same trial. The table A reports mean ± SD overall CV values for the three trials. The inset B shows the mean CV value of the three trials for the different methods. As only two laboratories performing haematoxilin-eosin participated in the first trial, the data have been omitted. Error bars represent SD values.

(A)	Mean % CV	SD
1° trial	92,82	6,91
2° trial	105,6	0,58
3° trial	88,62	0,96





were divided according to tertiles of activity of the laboratories, no differences in the mean bias value were present among laboratories with different declared activity (Fig. 3C).

DISCUSSION

Quality assurance and external and internal quality control are required for all types of biomedical analyses; in particular, they are considered essential in Andrology Laboratories (WHO, 1999, 2010; Pacey, 2006) in view of the fact that results are entirely operator dependent. EQC has the purpose to correct systematic and random errors that can be present and arise during execution of semen analysis. In addition, EQC allows the laboratory to have an idea of the type of errors present as well as the level of 'certainty' of their results (Pacey, 2006). Results of this study, which analysed the performance of semen analysis laboratories in the Tuscany region, demonstrate a huge variability both in the results and in the applied procedures of the different laboratories. Lack of agreement of semen analysis parameters may have important consequences for patients and for infertile couples undergoing ARTs, such as wrong choice of treatment modality possibly resulting in low/absent fertilization. In addition, in Tuscany region, where ART are guaranteed by the regional Health Care System, possible errors in the procedures arising from incorrect semen analysis may result in great waste of public money.

In general, the large variability in the results of our EQC is in agreement with previous studies aimed at analysing the performance of Andrology Laboratories (Matson, 1995; Jørgensen *et al.*, 1997; Gandini *et al.*, 2000; Cooper *et al.*, 2002; Alvarez *et al.*, 2005; Mallidis *et al.*, 2012). Although performed in a group of laboratories confined in a small region like Tuscany, our study identified possible sources of variability in the results and indicates possible solutions which may be of general interest.

One possible source of variability is represented by the methods used for the analyses. For sperm count, the lowest mean CV values of the three trials, although not reaching statistical significance, were obtained with the Neubauer cytometer (Fig. 1 inset B), which is the recommended counting methods by WHO (2010). Although at the first trial, mean CV values of results with the Makler cytometer were similar to those of the Neubauer, two considerations should be made on this result: one concerns the higher number of laboratories using Makler (which may have contributed to lower the mean CV value) and the second concerns the fact that spermatozoa to be counted in our EQC test are, at difference with fresh samples, immotile thus possibly making easier the evaluation in this chamber and lowering the variability. Although WHO discourages the use of the Makler cytometer, as also pointed out by us during the teaching course organized at the end of second trial, it remains the most used counting method in Tuscany Laboratories.

Concerning morphology, our results indicate that the variability is independent of the technique used to stain spermatozoa as mean CV values for the various methods employed by the laboratories were not different (inset B of Fig. 5). However, our study evidenced, at the third trial, lower CV values for Papanicolau, Haematoxilin-Eosin and TestSimplets respect to Diff-Quick, thus indicating better performance for the former three methods. However, it must be considered that, at least in case of TestSimplets, the high number of laboratories employing this methods may have contributed to decrease in the CV value, and, at the same time, the low number of laboratories using Diff-Quick to increase it. For sure, laboratories employing Papanicolau, Haematoxilin-Eosin and TestSimplets demonstrate to have a benefit from the teaching course between second and third trial (see also below). In any case, the huge variability in morphology data even at the third trial evidences the need for standardization of the procedures and increased adherence to WHO recommendation.

Both at the meeting after the first trial and during the course at the end of the second, we recommended adherence to WHO (2010) indicating the problems arising by using uncorrected procedures. Despite this, most laboratories continued to use non-recommended techniques (such as Makler chamber for counting and Testsimplet slides for morphology evaluation) even at the third trial. One important consequence is the fact that reference values of WHO should not be used by these laboratories, because, in case they are used, they may induce the clinician to an incorrect diagnosis.

In case of motility assessment, the mean CV values are the lowest respect to the other parameters. This result may be influenced by the fact that the sample does not need any procedure by the laboratory (as a DVD is sent), thus limiting errors caused by handling of the sample.

Another important source of variability is represented by the activity of the different laboratories. If we consider the results of the first trial (which was not influenced by our recommendations as in the case of second and, even more, the third one which was preceded by a teaching course), the lowest bias values of results were obtained, at least for morphology and concentration, in laboratories of the third tertile of activity which declare to perform over 200 semen analysis/year. This result reinforces the concept that not only operator's experience but also continuous practice are important in performing semen analysis. Most importantly, it should alarm about reliability of results from laboratories with low activity, although, according to our results, these laboratories showed better performances after teaching course, demonstrating commitment to ameliorate their activity.

Concerning possible solutions to decrease variability in the results among the different Laboratories, our study confirms the importance of teaching courses as also previously reported (Björndahl *et al.*, 2002). The decrease in the mean CV of morphology results at the third trial for most methods (Fig. 5) indeed indicates that the teaching course performed at the end of the second trial has been useful for evaluation of this parameter. It must be considered, however, that, although useful, the course was not sufficient to limit variability in the results of morphology, as the overall average CV of the laboratories remained very high, possibly because of the above-mentioned differences in the methods used to determine the parameter. Although not significantly different respect to the first and second, lower CV values were also observed, at the third trial, for sperm motility, whereas results of sperm count were apparently unaffected by the course. It should be noted that, although for sperm morphology several images can be used in a teaching course, allowing a better representation of the 'normal' spermatozoon according to WHO (2010), illustrating how to evaluate sperm motility and how to perform sperm counting may be more difficult, as only directives on the procedures to be used can be given.

In conclusion, our study indicates that variability in the procedures and laboratory activity (number of analysis performed) is

the main sources of lack of agreement of results of semen analysis in Tuscany region. Possible solutions to the problem are represented by persisting in participating in EQC programmes (Cooper *et al.*, 1999; Björndahl *et al.*, 2002), participating periodically to teaching courses and use of standardized WHO recommended procedures to increase precision and to allow referring to WHO reference values.

ACKNOWLEDGEMENTS

Supported by Regione Toscana and Azienda Ospedaliera-Universitaria Careggi (AOUC). We thank all semen analysis operators who participated in the course.

REFERENCES

Alvarez C, Castilla JA, Ramírez JP, Vergara F, Yoldi A, Fernández A *et al.* (2005) External quality control program for semen analysis: Spanish experience. *J Assist Reprod Genet* 22, 379–387.

Björndahl L, Barratt CL, Fraser LR, Kvist U & Mortimer D. (2002) ESHRE basic semen analysis courses 1995–1999: immediate beneficial effects of standardized training. *Hum Reprod* 17, 1299–1305.

Cooper TG, Atkinson AD & Nieschlag E. (1999) Experience with external quality control in spermatology. *Hum Reprod* 14, 765–769.

Cooper TG, Björndahl L, Vreeburg J & Nieschlag E. (2002) Semen analysis and external quality control schemes for semen analysis need global standardization. *Int J Androl* 25, 306–311.

Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM *et al.* (2010) World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 16, 231–245.

Gandini L, Menditto A, Chiodo F & Lenzi A. (2000) From the European Academy of Andrology. Italian pilot study for an external quality control scheme in semen analysis and antisperm antibiotics detection. *Int J Androl* 23, 1–3.

Harris MI, Flegal KM, Cowie CC, Eberhardt MS, Goldstein DE, Little RR *et al.* (1998) Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care* 21, 518–524.

Hill NR, Oliver NS, Choudhary P, Levy JC, Hindmarsh P & Matthews DR. (2011) Normal reference range for mean tissue glucose and glycemic variability derived from continuous glucose monitoring for subjects without diabetes in different ethnic groups. *Diabetes Technol Ther* 13, 921–928.

Jørgensen N, Auger J, Giwercman A, Irvine DS, Jensen TK, Jouannet P *et al.* (1997) Semen analysis performed by different laboratory teams: an intervariation study. *Int J Androl* 20, 201–208.

Jungwirth A, Giwercman A, Tournaye H, Diemer T, Kopa Z, Dohle G *et al.* (2012) EAU Working Group on Male Infertility. European Association of Urology Guidelines on Male Infertility: the 2012 Update. *Eur Urol* 62, 324–332.

Keel BA, Stembridge TW, Pineda G & Serafy NT Sr. (2002) Lack of standardization in performance of the semen analysis among laboratories in the United States. *Fertil Steril* 78, 603–608.

Mallidis C, Cooper TG, Hellenkemper B, Lablans M, Uckert F & Nieschlag E. (2012) Ten years' experience with an external quality control program for semen analysis. *Fertil Steril* 98, 611–616.

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

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

Pacey AA. (2006) Is quality assurance in semen analysis still really necessary? A view from the andrology laboratory *Hum Reprod* 21, 1105–1109.

Palacios ER, Clavero A, Gonzalvo MC, Rosales A, Mozas J, Martínez L *et al.* (2012) Acceptable variability in external quality assessment programmes for basic semen analysis. *Hum Reprod* 27, 314–322.

Rowe PJ, Comhaire FH, Hargreave TB & Mahmoud AMA. (2000) *WHO Manual for the Standardized Investigation, Diagnosis and Management of the Infertile Male*. Cambridge University Press, Cambridge.

World Health Organization (1980) *WHO Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, 1st edn. Press Concern, Singapore.

World Health Organization (1987) *WHO Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, 2nd edn. Cambridge University Press, Cambridge.

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.

World Health Organization (2010) *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. World Health Organisation Press, Geneva.

