

Results of the American Association of Bioanalysts national proficiency testing programme in andrology

Brooks A.Keel^{1,4}, Patrick Quinn²,
Charles F.Schmidt Jr³, Nicholas T.Serafy Jr³,
Nicholas T.Serafy Sr³ and Tammie K.Schalue¹

¹Department of Obstetrics and Gynecology, Women's Research Institute, University of Kansas School of Medicine-Wichita, Wichita, Kansas, ²Sage BioPharma, San Clemente, California and ³American Association of Bioanalysts Proficiency Testing Service, Brownsville, Texas, USA

⁴To whom correspondence should be addressed at: Women's Research Institute, University of Kansas School of Medicine-Wichita, 1010 N. Kansas, Wichita, KS 67214, USA

Proficiency testing samples for antisperm antibodies (ASAB), sperm count, morphology and vitality were mailed to participating laboratories. The majority participating utilized Immunobead ASAB procedures (81 versus 14% mixed antiglobulin reaction and 5% 'other'), and there was 95.6 ± 1.2% agreement on the presence or absence of ASAB. The majority of laboratories utilized manual (79%) versus computer assisted semen analysis (CASA; 15%) methods. Approximately 64% used the haemocytometer and 26% used the Makler counting chambers for manual counts. Coefficients of variation (CV) in sperm counts ranged from 24 to 138%, with CASA displaying lower overall CV (53 ± 8%) than manual methods (80 ± 9%). A wide variation in the reports of percent normal morphology was noted (CVs calculated from arc sin transformed means ranged from 15 to 93%). Participants using American Society of Clinical Pathologists (ASCP) criteria reported sperm morphology values that were clustered in the 'normal' range (11 out of 12 samples), while those using strict criteria were clustered in the 'abnormal' range (10 out of 12 samples). Good agreement was observed in sperm vitality (overall mean CV = 18%). These data highlight the urgent need for improvement in overall quality of andrology testing and indicate that practical proficiency testing programmes can be made available on a large scale.

Key words: antisperm antibodies/proficiency testing/sperm count/sperm morphology/sperm vitality

Introduction

Semen analysis is arguably the most important clinical laboratory test available in the evaluation of male fertility. In theory, semen analysis is a very simple test to perform; one merely places a drop of semen on a slide and determines the relative number, size and shape, and mobility of spermatozoa. In practice, however, the careful analysis of sperm concentration,

movement and morphology obviously requires a great deal of technical expertise and procedural care and meticulous quality control. It is the test's lack of standardization that has made semen analysis inaccurate and unreliable (Chong *et al.*, 1983; Mortimer *et al.*, 1986; Ombelet *et al.*, 1997), and has accommodated wide variation within and between laboratories (Jequier and Ukombe, 1983; Ayodejit and Baker, 1986; Neuwinger *et al.*, 1990; Walker, 1992). This has raised an urgent need for quality control (Mortimer *et al.*, 1986; Dunphy *et al.*, 1989; Mortimer, 1994; Clements *et al.*, 1995; Cooper *et al.*, 1999), and has caused some to refer to it as the 'neglected test' (Chong *et al.*, 1983). Although several authors have advocated the standardization of the methods utilized in the andrology laboratory (Chong *et al.*, 1983; Jequier and Ukombe, 1983; Mortimer *et al.*, 1986; Baker *et al.*, 1994) and have indicated the importance of quality control (Dunphy *et al.*, 1989; Cooper *et al.*, 1992; Muller, 1992; Mortimer, 1994; Clements *et al.*, 1995; Coetzee *et al.*, 1999), in many cases andrology testing is not comprehensive, technology and technical expertise is minimal and quality is compromised (Jequier and Ukombe, 1983; Baker *et al.*, 1994; Ombelet *et al.*, 1997).

The inaccuracies and lack of standardization associated with andrology testing have made it difficult, and in many cases impossible, for physicians to compare semen analysis results among laboratories. This is especially problematic when treating infertile couples referred from other clinics, who may have had fertility testing performed in other andrology laboratories. For example, owing to disagreements between laboratories, a patient could be classified as normal by one laboratory and infertile by another (Neuwinger *et al.*, 1990). Improvement in inter-laboratory agreement of test results is one of the hallmarks of a national proficiency testing (PT) programme. PT is a process of external, inter-laboratory quality control whereby simulated patient samples are tested by participating laboratories, and the performance of the individual laboratory (i.e. the test result) is compared with the collective performance of all participants (Stull *et al.*, 1998). Organized PT was first introduced in the United States the mid-1940s (Sunderman, 1992). With the advent of the Clinical Laboratory Improvement Act of 1966 and the Amendments of 1988, all clinical laboratories in the United States engaged in moderate or high complexity testing are now required to enrol in a government approved PT programme, if such a programme is available, and failure to achieve satisfactory performance in PT may result in sanctions against the laboratory (Keel, 1998). Since the introduction of PT, numerous reports have indicated that participation in organized PT programmes has resulted in a decrease in inter-laboratory standard deviations and coefficients

Table I. Number of laboratories participating in each proficiency testing programme

Challenge	ASAB	Count	Morphology	Vitality
May 1996	67	129	98	61
October 1996	75	150	114	71
May 1997	117	459	289	130
October 1997	116	465	288	139
May 1998	138	586	372	108
October 1998	135	598	372	169

ASAB = anti-sperm antibodies.

of variation with PT samples (Hanson, 1969; Hain, 1972; Rickman *et al.*, 1993), and a marked improvement in PT performance over time (Taylor and Fulford, 1981; Nakamura and Rippey, 1985; Rickman *et al.*, 1993; Tholen *et al.*, 1995). Thus, PT has caused a dramatic improvement in the quality of clinical laboratory testing and has served to ensure better agreement of results among laboratories.

Attempts at developing a multicentre, external inter-laboratory PT programme in andrology testing are limited (Neuwinger *et al.*, 1990; Walker, 1992; Matson, 1995; Cooper *et al.*, 1999). The American Association of Bioanalysts (AAB) Proficiency Testing Service (Brownsville, Texas, USA) began offering comprehensive external quality control PT programmes in 1949. In May of 1996, the AAB PT Service made available PT programmes for the clinical laboratory specialties of andrology and embryology. In this report, the results of this nationwide survey in andrology PT are presented.

Materials and methods

Proficiency testing (PT) programmes were developed for the detection of antisperm antibodies (ASAB), and for the determination of sperm count, sperm morphology and sperm vitality. Samples for quality control assessment (two samples per distribution) were obtained by contract from commercial vendors (Fertility Solutions, Inc., Cleveland, OH, USA and Dr Nina Desai, Cleveland, OH, USA) and were mailed to participating laboratories in the months of May and October each in the years of 1996, 1997 and 1998 (six testing distributions, 12 samples total). Individual laboratories opted to enrol in one or more of the programmes. The number of laboratories participating in each programme for each testing event is provided in Table I. Each individual PT sample was prepared or aliquotted from a single large well-mixed pool of human serum or semen. Thus, within practical limitations, each laboratory performed analysis on the same specimen. The samples were coded in such a fashion that the participating laboratory did not know whether the specimen would be 'normal' or 'abnormal'.

ASAB programme

Each testing event consisted of two aliquots (vials) of pooled, heat inactivated human serum. The samples were prepared so that one contained measurable tiers of ASAB (i.e. 'positive') and the other sample did not (i.e. 'negative'). The participating laboratories were instructed to allow the specimens to reach room temperature just prior to testing, and to perform the ASAB testing according to the laboratory's usual method. Results were recorded as either negative or positive only. Only one reporting method was allowed, and the

method was coded as 'Immunobead' (Clarke *et al.*, 1983), 'MAR' (mixed antiglobulin reaction; Jager *et al.*, 1978), or 'other'.

Sperm count programme

Each testing event consisted of two aliquots (vials) of pooled stabilized (formalin) suspensions of human spermatozoa. The samples were prepared so that one contained 'low' concentrations of spermatozoa and the other sample contained 'high' concentrations (realizing that these terms are relative). The participating laboratories were instructed to remove the specimens from the refrigerator, warm to room temperature, vortex for a minimum of 10 s or until completely in suspension, and count the spermatozoa according to the laboratory's usual method. Results were recorded as $\times 10^6$ spermatozoa/ml in whole numbers. Only one reporting method was allowed, and the method was coded as CASA (computer assisted semen analysis), manual or other. The laboratory was requested to indicate the type of counting chamber used, and these data were coded as haemocytometer (laboratories did not specify the type), Makler (Sefi-Medical Instruments, Haifa, Israel), Cell-VU (Millennium Sciences Inc., New York, NY, USA), or Micro-Cell (Conception Technologies, San Diego, CA, USA).

Sperm morphology programme

Each testing event consisted of two unstained glass slides of semen smears. The smears were fixed with CytoPrep (Fisher Scientific, Pittsburgh, PA, USA) prior to shipping. Laboratories were instructed either to stain immediately with Papanicolaou stain, or if using a Wright Giemsa based stain, to first remove the fixative by soaking the slides in 95% ethanol for a minimum of 20 min, followed by staining the slides by the laboratory's usual method. The laboratory was then instructed to perform the morphological analysis of the stained smears by the usual method. Results were reported as percentage normal forms in whole numbers. Two reporting methods were allowed in the event that a screening and a more definitive (strict) method was performed. The methods were coded as American Society of Clinical Pathologists (ASCP) (Adelman and Cahill, 1989), strict (Kruger *et al.*, 1986, 1988), WHO2 (World Health Organization, 2nd edn; WHO, 1987), WHO3 (3rd edition; WHO, 1992), or other. The laboratory was also requested to indicate the type of stain employed.

Sperm vitality programme

Each testing event consisted of two glass slides of semen smears that were stained with eosin-nigrosin prior to shipment. The laboratories were instructed to perform sperm vitality assessment according to their usual method, and to record percentage viable in whole numbers.

Statistical analysis

The values for reported morphology were given in percentage and were subjected to arc sin transformations (arc sin of the square root of the proportion) to achieve a Gaussian distribution (Neuwinger *et al.*, 1990) prior to calculation of CV.

Results

Approximately 80% of laboratories indicated that they utilized the Immunobead procedure for detecting antibodies, compared with ~15% of laboratories using the MAR method and 5% using 'other' procedures (data not shown). Percent agreement among the participating laboratories of the presence or absence of ASAB in the PT samples was high, averaging 95% agreement overall (Table II).

Table II. Percentage agreement of laboratories participating in the antisperm antibody proficiency testing programme on the presence (positive) or absence (negative) of antibodies

Sample	% agreement	Result
1	97	Positive
2	100	Negative
3	97	Negative
4	97	Positive
5	97	Negative
6	97	Positive
7	84	Positive
8	97	Negative
9	92	Positive
10	98	Negative
11	98	Negative
12	93	Positive
Mean \pm SE	95.6 \pm 1.2	
95% confidence interval	93.2–97.9	

Table III. Number (%) of laboratories participating in the sperm count proficiency testing programme according to the method and counting chamber used for counting spermatozoa

Method	Chamber	Number	% of total	% of method
Manual	Haemocytometer	291	53	64
	Makler	116	21	26
	Micro-Cell	15	3	3
	Cell-Vu	32	6	7
CASA	Makler	53	10	55
	Micro-Cell	41	7	2
	Cell-Vu	2	<1	2

Data represent the results from the most recent (October 1998) challenge only (representative data). CASA = computer assisted semen analysis.

Approximately 15% of participating laboratories used the CASA system. Because differences in estimating the number of spermatozoa in the PT sample may be related to the choice of counting chamber, participants were further asked to indicate the type of chamber used in their determinations (Table III). Data from the most recent PT challenge indicate that of the laboratories reporting sperm count by CASA, 55% used the Makler chamber, compared with 43% using Micro-Cell and a small percentage using Cell-Vu chambers. The re-usable counting chambers were the choice of a majority of laboratories performing a manual sperm count (64% used the haemocytometer and 26% used the Makler chamber). The coefficient of variation (CV) of reported sperm counts for laboratories performing a manual sperm count ranged widely from 30–138% compared with a CV range of 24–99% for laboratories reporting sperm counts by CASA (Table IV). The reported sperm concentrations according to counting chamber type and method, for the most recent challenge (representative data), are shown in Table V. These data illustrate the large variation and range (10- to 100-fold) of reported values for this PT programme.

More laboratories (40%) reported using the strict criteria than any other single method for determining normal sperm morphology. Interestingly, ~8% of participating laboratories continue to use the older version of the WHO method (2nd

Table IV. Coefficient of variation (CV) and mean sperm concentration ($\times 10^6/\text{ml}$) for the sperm count proficiency testing programme

Sample	Manual	CASA	Concentration
1	30	99	5.5
2	81	29	42.4
3	101	24	35.1
4	91	26	9.2
5	35	62	49.9
6	59	87	19.3
7	80	81	21.4
8	138	84	47.6
9	113	31	30.4
10	105	27	57.8
11	72	28	45.9
12	54	59	15.7
Mean \pm SE	80 \pm 9	53 \pm 8	
95% confidence interval	62–98	37–69	

edition), while 8% of participating laboratories still use less stringent ASCP criteria. The participating laboratories which used the less stringent ASCP criteria for judging sperm morphology tended to report morphology values in the 'normal range' for that method, while those laboratories who used the more stringent strict criteria tended to report 'abnormal values' for sperm morphology in the challenge samples (Table VI). The possible exception to this observation is in sample 8, which was classified as markedly teratozoospermic by all criteria. The extreme variation according to the criteria used by laboratories for reporting 'normal' and 'abnormal' sperm morphology is further evident when the overall mean results for each of the 12 samples are presented (Table VI). The CV of reported sperm morphologies ranged widely from 7 to 56%, depending upon the method and sample (data not shown). The less stringent ASCP method yielded the lowest overall CV (21%), while the more stringent WHO and strict methods resulted in much higher CV (~33 and 50% respectively).

The reported range of percentage viable spermatozoa in the 12 challenge samples ranged from a low of 6.3% viable to a high of 82.3% viable. Good agreement among the laboratories participating in this programme was observed, with an overall CV of ~18% (Table VII).

Discussion

The importance of PT in this area of clinical laboratory medicine has been discussed (Byrd, 1992; Gerrity, 1993; Keel, 1998; Cooper *et al.*, 1999) and several preliminary attempts, involving a relatively small number of participating laboratories, have been made to implement inter-laboratory comparisons of various andrology testing procedures (Jequier and Ukombe, 1983; Ayodeji and Baker, 1986; Neuwinger *et al.*, 1990; Walker, 1992; Matson, 1995; Coetzee *et al.*, 1999). The results presented herein represent the first report of a large-scale nationwide proficiency testing programme in andrology.

The routine semen analysis involves the morphological evaluation of live, motile cells. Preparing live, motile sperm samples in large enough numbers to offer wide distribution

Table V. Reported sperm concentrations in the sperm count PT programme according to the method and counting chamber used to count spermatozoa

Method	Chamber	Sample 11			Sample 12		
		Mean	Range	SD	Mean	Range	SD
Manual	Haemocytometer	49.6	3–492	37.8	16.5	2–147	9.9
	Makler	44.9	5–320	28.1	14.7	3–31	4.2
	Micro-Cell	35.1	24–53	9.6	14.2	6–40	8.0
	Cell-Vu	37.7	29–47	4.5	12.8	6–16	2.3
CASA	Makler	41.7	15–72	11.4	16.8	8–84	11.6
	Micro-Cell	38.6	4–82	11.5	13.6	8–26	4.0
	Cell-Vu	36.5	27–46	13.4	15.5	15–16	0.7

Data are the results of sample 11 and sample 12 from the most recent (October 1998) challenge only (representative data). Sperm count is shown as the mean reported concentration and range ($\times 10^6/\text{ml}$) and SD of all laboratories using the indicated method and counting chamber.

Table VI. Reported mean sperm morphology (% normal forms) in the sperm morphology proficiency testing programme according to the criteria used for judging morphologically normal spermatozoa

Sample	ASCP	Strict	WHO2	WHO3
1	80.7	10.8	35.2	28.8
2	70.0	11.1	43.1	30.2
3	73.2	7.6	41.6	26.0
4	79.8	10.8	50.4	33.4
5	77.8	8.8	42.2	34.9
6	68.8	9.9	40.3	35.3
7	81.5	15.0	54.0	43.6
8	47.5	2.7	23.0	14.8
9	68.7	5.3	36.0	24.7
10	72.5	15.7	49.5	41.9
11	74.9	13.9	51.8	41.6
12	70.0	12.2	47.2	39.3

ASCP = American Society of Clinical Pathologists.

For criteria groups, see Materials and methods section on Sperm morphology programme.

Table VII. Coefficient of variation (CV), mean \pm SE and 95% confidence interval (CI) of percentage vital sperm for the sperm vitality PT programme samples

Sample	CV%	% vitality	95% CI
1	14	44.5 \pm 0.8	(42.9–46.1)
2	32	6.3 \pm 2.5	(1.4–11.2)
3	13	47.7 \pm 0.7	(46.3–49.1)
4	10	64.5 \pm 0.8	(62.9–66.1)
5	18	28.6 \pm 0.4	(27.8–29.4)
6	10	67.3 \pm 0.6	(66.1–68.5)
7	24	53.5 \pm 1.1	(51.3–55.6)
8	30	28.3 \pm 0.7	(26.9–29.7)
9	22	21.7 \pm 0.4	(20.9–22.5)
10	8	65.9 \pm 0.5	(64.9–66.9)
11	6	82.3 \pm 0.4	(81.5–83.1)
12	25	24.4 \pm 0.5	(23.4–25.4)
Mean \pm SE	18 \pm 2		
95% CI	14–22		

presents a unique challenge for proficiency testing and inter-laboratory comparisons. Rather than using cryopreserved specimens (Neuwinger *et al.*, 1990), we and others chose to send stable suspensions of fixed spermatozoa for sperm count-

ing and fixed semen smears for morphology determinations. Although the use of batched, premade semen smears eliminates the potential variation associated with individual slide preparation, it is recognized that this approach also precludes the comparison of the effect that this variation may have on individual laboratory performance. Furthermore, suspensions of fixed spermatozoa can be problematic for the participating laboratories, in that spermatozoa cell settling and clumping necessitates mixing of the specimen prior to analysis, and failure to mix the specimens carefully could result in erroneous results. The use of cryopreserved semen would appear to be a viable alternative of sample preparation, in that the 'live' nature of the specimen is preserved. However, several reports have indicated a lack of consistency between aliquots of frozen semen when used for inter- and intra-laboratory variation determinations (Cooper *et al.*, 1992; Muller, 1992; Clements *et al.*, 1995). While the assessment of sperm motility necessitated the use of cryopreserved specimens in two previous studies (Neuwinger *et al.*, 1990; Walker, 1992), the high cost and inconvenience of cryopreserving and shipping the frozen semen makes this approach unrealistic on a large scale (Cooper *et al.*, 1999). The use of videotapes of fresh ejaculates, which can be duplicated from a master tape and distributed relatively inexpensively, is being evaluated as a suitable alternative for PT programmes in sperm motility.

The vast majority of laboratories participating in the sperm count programme employed manual, non-automated methods for counting spermatozoa. Furthermore, more laboratories used the standard haemocytometer than any other counting chamber. Baker *et al.* (Baker *et al.*, 1994) surveyed 129 acute care community hospitals in the United States and found that when these laboratories performed semen analyses, they tended to use more conventional methods, including the use of the haemocytometer for sperm counts. Only 1.6% of hospital clinical laboratories surveyed performed automated semen analysis and only 3.1% used a Makler counting chamber (Baker *et al.*, 1994). Thus, the preference of methods for sperm counting and the selection of counting chambers may, in many cases, reflect the type of clinical services and level of expertise provided by the participating laboratory.

Wide variations in sperm concentrations between laboratories (CV ranging from 10 to 65%) have been previously reported for both manual (Jequier and Ukombe, 1983; Neuwinger *et al.*, 1990; Walker, 1992; Matson, 1995) and CASA (Walker, 1992) methods. The results of the sperm count PT programme presented herein, from participants representing a wide spectrum of clinical laboratory settings and expertise, demonstrated alarmingly high CV and wide ranges in reported sperm concentrations. Indeed, reported sperm concentrations among the participating laboratories varied by as much as two orders of magnitude, indicating a sperm concentration of $3 \times 10^6/\text{ml}$ in one laboratory, and $492 \times 10^6/\text{ml}$ in another for the same sample. This variation appeared to be greater when results were compared among laboratories using manual methods versus CASA, which may reflect the expertise of the reporting laboratory. However, it should be pointed out that the type of CASA system used by the participating laboratory was not requested. Unless standardized procedures are agreed upon and strictly followed (Davis and Katz, 1992), differences may exist between CASA systems in the ability to provide accurate values for sperm concentrations (Gill *et al.*, 1988; Mahony *et al.*, 1988; Agarwal *et al.*, 1992; ESHRE Andrology Special Interest Group, 1998). This may account for at least part of the observed variation. Improper specimen handling (see above) and potential clerical and data entry/recording errors (see below) notwithstanding, the data presented herein suggest gross unreliability in the results of sperm concentrations reported by some clinical laboratories. They also highlight the urgent need for thorough technician training and the use of careful, standardized procedures and regular internal quality control evaluations (Mortimer *et al.*, 1986; Mortimer, 1994).

It has been argued that of all individual semen parameters, sperm morphology is most closely related with fertility potential (Kruger *et al.*, 1986, 1988; Ombelet *et al.*, 1995). However, it is also recognized that there is considerable variation in this determination (Jequier and Ukombe, 1983; Dunphy *et al.*, 1989; Neuwinger *et al.*, 1990; Baker *et al.*, 1994). Reasons for this variation include lack of standardization (Chong *et al.*, 1983; Dunphy *et al.*, 1989), differing techniques of smear preparation and staining procedures (Davis and Gravance, 1993), and the level of technical expertise (Dunphy *et al.*, 1989; Neuwinger *et al.*, 1990). Lack of standardization can make it difficult if not impossible to compare results from one laboratory to another. Although some investigators have supported the use of computerized morphological assessments (i.e. CASA) as a means to establish standardization and reduce variability (Barroso *et al.*, 1999; Kruger and Coetzee, 1999), others have concluded that the current generation of CASA instruments is not capable of analysing human sperm morphology in a manner adequate for routine clinical applications (ESHRE Andrology Special Interest Group, 1998). Complicating the standardization issue is the fact that little consensus exists on the most appropriate classification system. In a recent survey of 410 fertility centres from all over the world, a wide and complex variation was found in different sperm morphology classification systems employed (Ombelet *et al.*, 1997). In the current study, approximately two-thirds of the participating laboratories utilized the more standardized World

Health Organization or strict criteria for morphology determinations. However, the remainder of participating laboratories used the much less stringent ASCP criteria, or some other undefined protocol. It is evident from the data that overall, those laboratories using the ASCP criteria tended to classify the PT samples as normal while those laboratories using the strict criteria tended to classify the PT samples as teratozoospermic.

Considerable variation exists when comparing results of sperm morphology both between and within laboratories (Jequier and Ukombe, 1983; Ayodeji and Baker, 1986; Neuwinger *et al.*, 1990; Clements *et al.*, 1995; Matson, 1995; Coetzee *et al.*, 1999). In this study, a high degree of variation among laboratories participating in the morphology programme is reported, with CV ranging from 15 to 93%. Interestingly, the greatest variation was observed among laboratories using the most stringent criteria (strict) while the least variation was found among laboratories using the less stringent (ASCP). It should be pointed out that one must use caution when computing and comparing CVs among criteria that use vastly different percentages as normal ranges. The use of strict criteria results in a significant reduction in the mean values obtained, which results in an increase in the CV (Clements *et al.*, 1995) analogous to the precision profiles of assays in which increased variation is observed at lower analyte concentrations (Cooper *et al.*, 1999). This fact, coupled with the statistical concerns of calculating the variance of proportions, necessitated the use of data transformation prior to determining CV. Thus, the range of CV reported herein would be even greater if the data had not been transformed before analysis. Along these same lines, in order to obtain the same statistical confidence, more spermatozoa must be counted when using a criterion representing a low percentage of normal forms compared with a criterion which established a higher percentage of normal forms (Davis and Gravance, 1993; Coetzee *et al.*, 1999). We did not determine or standardize the number of spermatozoa counted by the individual laboratories participating in this PT programme, and this may have contributed to some of the observed variation.

There appears to be only one other report which evaluated the performance of sperm vitality determinations among different laboratories (Walker, 1992). In that study, large CVs were reported, ranging from 42 to 90%. In contrast, relatively low CV were noted among the laboratories participating in this PT programme. This previous study (Walker, 1992) collected data from a relatively few laboratories, and each participant was required to prepare and stain their own semen smears. In contrast, in the current study, premade and prestained smears were provided, which would have eliminated the variation associated with this process. Nevertheless, the data indicate that good agreement and interpretation of viable versus nonviable spermatozoa can be obtained among laboratories.

There are recognized shortcomings of PT. The results reported by PT participants represent a concurrence, or agreement upon a certain value, rather than a reflection of the actual value of the analyte measured. For example, a large majority of participating laboratories could agree on a certain value, yet be incorrect in estimating the true value (i.e. PT measures

precision, not accuracy). The use of reference values, determined by a small group of expert referee laboratories, may help to circumvent this shortcoming. Nevertheless, PT does represent an effective mechanism to ensure accurate comparison of reported values between laboratories. In addition, there are several variables unique to the PT process which are unrelated to normal clinical laboratory performance and can lead to unreliable PT results (Stull *et al.*, 1998). Clerical and mathematical errors in reporting results to PT programmes are a potential source of variation that is difficult to measure. Although the PT report forms used herein did not allow for the use of decimals, and the ability of laboratories to apply correct mathematical calculations prior to reporting was not addressed, values were accepted as submitted and no attempt was made to ensure accuracy of recording results. Limitations on the use of PT as an indicator of routine laboratory performance is also hampered by the fact that a PT event is a non-random sample of the work performed at a given testing site and is subject to biases inherent in such a process (Stull *et al.*, 1998). In other words, the individual processing the PT sample is often aware of the fact that the PT sample is unique and the results are subject to greater scrutiny, which may result in increased pressure to give the specimen 'special handling' (Boone, 1992). Such special handling may involve repeated analysis and averaging of multiple replicates, the use of the most trained technician or multiple technicians, and use of methods other than those routinely used (Boone, 1992). Thus, in many cases, PT performance, as measured through mailed-in samples that are known to the laboratory as regulatory challenges, will reflect the best performance that laboratory is capable of providing, and not necessarily its typical or routine performance (Boone, 1992). Although PT participants are instructed to use routine procedures in analysing PT specimens and avoid 'special handling', if PT performance truly represents the best analytical work a laboratory is capable of producing (Stull *et al.*, 1998), then the data in this study would indicate that large variations exist in the results of semen analysis from the laboratories participating in this programme even under the best of circumstances.

The data presented here support the urgent plea for standardization of semen analysis methodologies expressed by others (Chong *et al.*, 1983; Mortimer, *et al.*, 1986; Mortimer, 1994; Ombelet *et al.*, 1997). It is strongly recommended that laboratories performing semen analyses reduce the variation observed herein by adhering to accepted standards, such as those proposed by the World Health Organization (WHO, 1992), and by employing active programmes of internal quality control. Numerous reports have indicated that participation in organized PT programmes has resulted in a decrease in interlaboratory SD and CV with PT samples (Hanson, 1969; Hain, 1972; Rickman *et al.*, 1993), and a marked improvement in PT performance over time (Taylor and Fulford, 1981; Nakamura and Rippey, 1985; Rickman *et al.*, 1993; Tholen *et al.*, 1995). Thus, it is hoped that PT in the field of clinical laboratory andrology will help to improve the quality of clinical laboratory testing and serve to ensure better agreement of results among laboratories.

Acknowledgements

This study is supported in part by the Women's Research Institute, Wichita, Kansas and the American Association of Bioanalysts Proficiency Testing Service, Brownsville, TX, USA. B.A.K. and P.Q. are on the Board of Scientific Advisors of the American Association of Bioanalysts Proficiency Testing (AABPT) Program; N.T.S. Jr, N.T.S. Sr and C.F.S. are employed by the AABPT Program; P.Q. has a vested interest in Sage BioPharma.

References

- Adelman, M.M. and Cahill, E.M. (eds) (1989) *Atlas of Sperm Morphology*. ASCP Press, Chicago, IL.
- Agarwal, A., Ozturk, E. and Loughlin, K.R. (1992) Comparison of semen analysis between the two Hamilton-Thorn semen analysers. *Andrologia*, **24**, 327–329.
- Ayodeji, O. and Baker, H.W. (1986) Is there a specific abnormality of sperm morphology in men with varicoceles? *Fertil. Steril.*, **45**, 839–842.
- Baker, D.J., Paterson, M.A., Klaassen, J.M. *et al.* (1994) Semen evaluations in the clinical laboratory. How well are they being performed? *Lab. Med.*, **25**, 509–514.
- Barroso, G., Mercan, R., Ozguc, K. *et al.* (1999) Intra- and inter-laboratory variability in the assessment of sperm morphology by strict criteria: impact of semen preparation, staining techniques and manual versus computerized analysis. *Hum. Reprod.*, **14**, 2036–2040.
- Boone, D.J. (1992) Literature review of research related to the Clinical Laboratory Improvement Amendments of 1988. *Arch. Pathol. Lab. Med.*, **116**, 681–693.
- Byrd, W. (1992) Quality assurance in the reproductive biology laboratory. *Arch. Pathol. Lab. Med.*, **116**, 418–422.
- Chong, A.P., Walters, C.A. and Weinreb, S.A. (1983) The neglected laboratory test. The semen analysis. *J. Androl.*, **4**, 280–282.
- Clarke, G.N., Stojanoff, A., Cauchi, M.N. *et al.* (1983) Detection of antispermatzoal antibodies of IgA class in cervical mucus. *Am. J. Reprod. Immunol.*, **5**, 61–65.
- Clements, S., Cooke, I.D. and Barratt, C.L.R. (1995) Implementing comprehensive quality control in the andrology laboratory. *Hum. Reprod.*, **10**, 2096–2106.
- Coetzee, K., Kruger, T.F., Lombard, C.J. *et al.* (1999) Assessment of interlaboratory and intralaboratory sperm morphology readings with the use of a Hamilton Thorne Research integrated visual optical system semen analyzer. *Fertil. Steril.*, **71**, 80–84.
- Cooper, T.G., Neuwinger, J., Bahrs, S. *et al.* (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.
- Davis, R.O. and Gravance, C.G. (1993) Standardization of specimen preparation, staining, and sampling methods improves automated sperm-head morphometry analysis. *Fertil. Steril.*, **59**, 412–417.
- Davis, R.O. and Katz, D.F. (1992) Standardization and comparability of CASA instruments. *J. Androl.*, **13**, 81–86.
- Dunphy, B.C., Kay, R., Barratt, C.L.R. *et al.* (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.
- Gerrity, M. (1993) Legislative efforts affecting the reproductive biology laboratory. *Curr. Opin. Obstetric. Gynecol.*, **5**, 623–629.
- Gill, H.S., Van Arsdalen, K., Hypolite, J. *et al.* (1988) Comparative study of two computerized semen motility analyzers. *Andrologia*, **20**, 433–440.
- Hain, R.F. (1972) Proficiency testing in the physician's office laboratory: an ounce of prevention. *S. Med. J.*, **65**, 608–610.
- Hanson, D.J. (1969) Improvements in medical laboratory performance. *Postgrad. Med.*, **46**, 51–56.
- Jager, S., Kremer, J. and van Slochteren-Draaisma, T. (1978) A simple method of screening for antisperm antibodies in the human. Detection of spermatozoal surface IgG with the direct mixed antiglobulin reaction carried out on untreated fresh human semen. *Int. J. Fertil.*, **23**, 12–21.
- Jequier, A.M. and Ukombe, E.B. (1983) Errors inherent in the performance of a routine semen analysis. *Br. J. Urol.*, **55**, 434–436.

- Keel, B.A. (1998) The assisted reproductive technology laboratories and regulatory agencies. *Infert. Reprod. Med. Clin. N. Am.*, **9**, 311–330.
- Kruger, T.F. and Coetzee, K. (1999) The role of sperm morphology in assisted reproduction. *Hum. Reprod. Update*, **5**, 172–178.
- Kruger, T.F., Menkveld, R., Stander, F.S.H. *et al.* (1986) Sperm morphologic features as a prognostic factor in *in vitro* fertilization. *Fertil. Steril.*, **46**, 1118–1123.
- Kruger, T.F., Acosta A.A., Simmons, K.F. *et al.* (1988) Predictive value of abnormal sperm morphology in *in vitro* fertilization. *Fertil. Steril.*, **49**, 112–117.
- Mahony, M.C., Alexander, N.J. and Swanson, R.J. (1988) Evaluation of semen parameters by means of automated sperm motion analyzers. *Fertil. Steril.*, **49**, 876–880.
- Mortimer, D. (ed.) (1994) *Practical Laboratory Andrology*. Oxford University Press, New York, NY.
- Mortimer, D., Shi, 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.
- Matson, P.L. (1995) Quality control assessment for semen analysis and sperm antibody detection: results of a pilot scheme. *Hum. Reprod.*, **10**, 620–625.
- Muller, C.H. (1992) The andrology laboratory in an assisted reproductive technologies program. *J. Androl.*, **13**, 349–360.
- Nakamura, R.M. and Rippey, J.H. (1985) Quality assurance and proficiency testing for autoantibodies to nuclear antigen. *Arch. Pathol. Lab. Med.*, **109**, 109–114.
- Neuwing, J., Behre, H.M. and Nieschlag, E. (1990) External quality control in the andrology laboratory: an experimental multicenter trial. *Fertil. Steril.*, **54**, 308–314.
- Ombelet, W., Menkveld, R., Kruger, T.F. *et al.* (1995) Sperm morphology assessment: historical review in relation to fertility. *Hum. Reprod. Update*, **1**, 543–557.
- Ombelet, W., Pollet H., Bosmans, E. *et al.* (1997) Results of a questionnaire on sperm morphology assessment. *Hum. Reprod.*, **12**, 1015–1020.
- Rickman, W.J., Monical, C. and Waxdal, M.J. (1993) Improved precision in the enumeration and absolute numbers of lymphocyte phenotypes with long-term monthly proficiency testing. *Ann. N. Y. Acad. Sci.*, **677**, 53–58.
- Stull, T.M., Hearn, T.L., Hancock, J.S. *et al.* (1998) Variation in proficiency testing performance by testing site. *JAMA*, **279**, 463–467.
- Sunderman Sr, F.W. (1992) The history of proficiency testing/quality control. *Clin. Chem.*, **38**, 1205–1209.
- Taylor, R.N. and Fulford, K.M. (1981) Assessment of laboratory improvement by the Center for Disease Control Diagnostic Immunology Proficiency Testing Program. *J. Clin. Microbiol.*, **13**, 29.
- Tholen, D., Lawson, N.S., Cohen, T. *et al.* (1995) Proficiency test performance and experience with College of American Pathologists' programs. *Arch. Pathol. Lab. Med.*, **119**, 307–311.
- Walker, R.H. (1992) Pilot surveys for proficiency testing of semen analysis. *Arch. Pathol. Lab. Med.*, **116**, 432–424.
- World Health Organization (1987) *WHO Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, 2nd edn. Cambridge University Press, Cambridge, UK.
- World Health Organization (1992) *WHO Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, 3rd edn. Cambridge University Press, Cambridge, UK.

Received on September 3, 1999; accepted on December 2, 1999