Journal of Andrology, Vol. 25, No. 4, July/August 2004
Copyright © American Society of Andrology

Standardized Methods for Semen Evaluation in a Multicenter Research Study

CHARLENE BRAZIL,* SHANNA H. SWAN,† ERMA Z. DROBNIS,‡ FAN LIU,† CHRISTINA WANG,§ J. BRUCE REDMON,‖ JAMES W. OVERSTREET,* AND THE STUDY FOR FUTURE FAMILIES RESEARCH GROUP

*From the *Division of Reproductive Biology, Department of Obstetrics and Gynecology, University of California, Davis, California; the †Department of Family and Community Medicine and the ‡Department of Obstetrics and Gynecology, University of Missouri, Columbia, Missouri; the §Department of Medicine, Harbor-UCLA Medical Center, Los Angeles, California; and the ‖Departments of Medicine and Urologic Surgery, University of Minnesota, Minneapolis, Minnesota.*

ABSTRACT: Semen evaluation methodology is complex and difficult to standardize. Rigorously standardized laboratory protocols and strict quality control (QC) are essential for meaningful comparison of data from multiple sites. We describe the methods used for determination of semen volume, sperm concentration, and percent sperm motility in the Study for Future Families, a multicenter study of semen quality in the United States. Each of these 3 semen parameters was assessed using 2 techniques, which provided the opportunity to compare precision and assess suitability for multicenter studies. Detailed protocols were used, and technicians were centrally trained. A total of 509 semen evaluations were performed. Semen volume measured by weight was greater ($P < .0001$) than that determined by pipetting (3.7 ± 1.6 mL vs 3.2 ± 1.6 mL). Sperm concentration determined using hemacytometer chambers was consistently higher ($P < .001$) than that using disposable MicroCell chambers (81.0×10^6/mL vs 65.9×10^6/mL). Precision was slightly greater for the MicroCell chamber. The percentage of motile sperm was assessed by a simple counting technique as well as by the World Health Organization categorical method that assigns individual motile sperm to "a," "b," and "c" categories on the basis of progression. When these 3 categories were collapsed, the methods provided values that were not statistically different ($P > .05$), although the collapsed values tended to be higher (58.1% vs 51.6%) and less precise (CV 7.7% vs 4.1%) for the categorical method than for motility determined using the simple method. The data obtained in this study demonstrate the critical need for rigorous standardization of protocols and techniques for multicenter studies.

Key words: Sperm concentration, semen volume, sperm motility, quality control, observer variation, precision.

J Androl 2004;25:635–644

Semen evaluation is the single most important laboratory test for assessing male fertility (Jarow et al, 2002). However, the methodology is complex, and standardization is difficult. For example, the first large-scale, nationwide proficiency testing program for clinical andrology laboratories in the United States reported that the interlaboratory coefficient of variation for manual sperm concentration determination was 80%, with a range for 1 semen specimen of $3–492 \times 10^6$/mL (Keel et al, 2000). Some of this difference is likely attributable to the use of different sperm counting chambers, and there is considerable disagreement regarding the relative accuracy and precision of various chambers. The well-known variability among laboratories in the determination of sperm concentration and other semen parameters has contributed to uncertainty in the clinical interpretation of semen analysis results (Neuwinger et al, 1990; Keel et al, 2000).

Recently, limitations of semen evaluation methodology have been brought into sharp focus by controversies raised in the epidemiological literature. Some investigators have presented analyses based on data from historical studies suggesting that sperm concentrations have decreased significantly during the past 50 years (Carlsen et al, 1992; Swan et al, 1997, 2000). Other studies suggest sperm counts have not decreased (Fisch et al, 1996; Paulsen et al, 1996; Vierula et al, 1996). While studies suggest considerable geographical variation in sperm concentration, these studies have used a variety of study populations and methods (Fisch and Goluboff, 1996). To be credible, the data generated by multicenter studies must be obtained using comparable study populations and protocols of strict quality control (QC) for andrology labo-

This work was supported by grant R01 ES09916 to the University of Missouri from the National Institutes for Environmental Health Sciences (NIEHS), National Institutes of Health; grant MO1-RR00400 to the University of Minnesota General Clinical Research Center; and grant MO1-RR0425 to the Research and Education Institute at the Harbor-UCLA Medical Center and the Cedars-Sinai Research Institute from the National Center for Research Resources, National Institutes of Health.

Correspondence to: Dr James W. Overstreet, Center for Health and the Environment, University of California, One Shields Ave, Davis, CA 95616 (e-mail: jwoverstreet@ucdavis.edu).

Received for publication January 20, 2004; accepted for publication March 2, 2004.

ratory procedures (Auger et al, 2000). Standardization of techniques, combined with training and practice, has been shown to decrease variation in estimates of semen parameters (Björndahl and Kvist, 1998; Punjabi and Spiessens, 1998).

In this communication, we report the methods used for determination of semen volume, sperm concentration, and sperm motility in the Study for Future Families (SFF), a multicenter study of semen quality conducted in 4 cities in the United States (Swan et al, 2003). The semen evaluations for the SFF also included assessments of sperm motility by computer-aided sperm analysis (CASA) and manual assessments of sperm morphology, both of which will be discussed in a subsequent publication. Semen volume, sperm concentration, and sperm motility were each assessed using 2 techniques. The use of 2 methods was necessary to enable comparison with data obtained in the European multicenter study of fertile men (Jørgensen et al, 2001) as well as studies of fertile and infertile men in the United States (Guzick et al, 2001). The use of multiple methods provided the opportunity to compare the precision of the different methods and to assess their suitability for use in multicenter studies.

Materials and Methods

Design of the SFF

The SFF was designed to compare semen quality of fertile males in 4 cities in the north, east, west, and south-central regions of the United States by using identical methods for semen analysis. Women who conceived a pregnancy without medical intervention were recruited at prenatal clinics of university hospitals in Los Angeles, Calif, Minneapolis, Minn, Columbia, Mo, and New York, NY. The partners of these women provided 2 semen samples for analysis. The study subjects gave written informed consent, and all study protocols were approved by the institutional review board (IRB) of each participating institution. Details of the study protocols and the results obtained for the first 493 men in the study are published elsewhere (Swan et al, 2003).

A central andrology laboratory at the University of California, Davis had responsibility for standardizing supplies and equipment, providing detailed written protocols for standardized semen analysis, training laboratory technicians from each center, and providing QC for the study. The latter activity included preparation and regular distribution of QC materials, analysis of results, and follow-up activities. The semen donors at the central andrology laboratory were healthy young volunteers who gave written informed consent; research activities were approved by the IRB of the University of California, Davis. The primary data reported in this study are the results obtained following analysis of the first semen samples collected by 509 men at the 4 study sites. In addition to the SFF data, results are presented for several ancillary studies that were performed at the central andrology laboratory.

Supplies and Equipment

All study sites used the same laboratory supplies, including MicroCell chambers (Conception Technologies, San Diego, Calif). Equipment was closely matched and included Olympus phase-contrast microscopes fitted with eyepiece reticles (KR-402; Klarman Rulings, Litchfield, NH) and stage warmers. Each site purchased Improved Neubauer Phase hemacytometer chambers (Hauser Scientific Inc, Horsham, Pa) at the beginning of the study. These chambers were used only for study subjects. Positive displacement pipettors were used to prepare all dilutions. All supplies that came in contact with live sperm were pretested for toxicity by assessing survival of sperm suspensions in contact with the test material. Briefly, toxicity testing was performed by exposing ''swim-up'' sperm to the laboratory supplies. One hundred fifty microliters of sperm suspension was placed in contact with the test material and covered with oil. Overnight survival after incubation at 37°C was assessed and compared to the control, which was incubated in a Corning 15-mL tube. Both percent motility and progression were considered. Test materials were considered nontoxic if percent motility was within 10%–15% of the control and had similar progression. None of the materials used for this study negatively influenced sperm motility. Microbead suspensions (Accubead; Hamilton Thorne Research, Beverly, Mass) were used in ancillary studies of counting chambers by the central andrology laboratory.

Technician Training

Technicians from the study sites were trained at the central andrology laboratory during a weeklong training session, regardless of prior experience. Five technicians provided data for the present study. The details of technician training and QC activities by the central andrology laboratory are described in another publication (Brazil et al, 2004).

Laboratory Protocols

Standard Analysis Methods—Following collection by the study subjects, semen was allowed to liquefy at room temperature and was analyzed within 30–45 minutes. Immediately before each semen aliquot was removed for analysis or dilution, the entire semen specimen was mixed thoroughly 10 times by aspiration and expulsion with a disposable transfer pipette. Following dilution with positive displacement pipettors, diluted sperm suspensions for concentration assessment were vortexed for 20 seconds; an aliquot was immediately removed and loaded into the chamber. Separate dilutions were used to load the 2 sides of the counting chambers. Similarly, replicate slides and chambers were prepared for motility assessment. After the 2 initial assessments of MicroCell concentration, hemacytometer concentration, or percent motility were performed, the percent difference between the 2 values was determined. If they differed by less than 10%, the mean value was used. If, however, the difference was greater than 10%, the semen was mixed again, a third measurement from a new aliquot of semen was made, and the median of the 3 values was used.

Semen Volume Determinations—Volume was assessed initially by semen weight, assuming a density of 1.0 g/mL. The specimen container was weighed before and after semen collection, and the weights were recorded to the nearest hundredth gram.

The difference between these weights was considered the volume, by weight, of the ejaculate. In addition, the entire ejaculate was aspirated into a 1-, 5-, or 10-mL pipette, and the volume was recorded to the nearest tenth milliliter.

Sperm Concentration Determination With the MicroCell Chamber—Sperm were immobilized with diluent (World Health Organization [WHO], 1999) using 1 part diluent : 1 part semen; the MicroCell chamber was loaded with 7 μL of this suspension. An eyepiece reticle was used to score sperm from 6 or 12 fields in each side of the chamber. For sperm concentrations equal to 20×10^6/mL or greater, 6 fields were counted; for sperm concentrations less than 20×10^6/mL, 12 fields were counted. The fields were chosen according to a prescribed pattern: 3 (or 6) fields evenly spaced left to right and 3 (or 6) fields evenly spaced top to bottom, forming a plus sign centered in the middle of the chamber, excluding the areas 2–3 mm from the chamber edges. To avoid selection bias for or against fields of a particular density or regularity, all fields were located without looking through the microscope. On average, 200–300 sperm were scored for each count.

Ancillary studies were conducted at the central andrology laboratory using donor semen to compare chamber loading procedures currently suggested by the MicroCell manufacturer (3 μL of raw semen) to those suggested by the manufacturer in the past (7 μL of raw semen) and those used in this study (7 μL of immobilized sperm). To assess possible loading bias within the chamber, comparisons were also made of sperm densities of different microscopic fields within the MicroCell chamber.

Sperm Concentration Determination With the Hemacytometer Chamber—Procedures recommended by WHO (1999) guidelines for hemacytometry were followed, except that dilutions were always 1:20, regardless of sperm concentration, and the area of the hemacytometer grid counted was altered to ensure that adequate sperm were counted, regardless of the concentration (see below). These changes were made to minimize technical errors in dilution or in estimating sperm concentration before dilution, as well as to minimize errors in calculating the final sperm concentration after counting. An additional deviation from the WHO recommendation was that both pillars of the hemacytometer were wet with water to secure the coverslip to the chamber, as opposed to the WHO (1999) recommendation of wetting only 1 pillar.

For this study, the hemacytometer pillars were each wetted with a 1- to 2-μL drop of water before applying the coverslip. Ten microliters of the diluted sperm suspension were used to load one side of the hemacytometer chamber; 10 μL from a separate dilution were used to load the other side. After 5–10 minutes in a humid chamber to allow the sperm to settle, sperm were counted from the top row in the central large square of the hemacytometer grid. On the basis of the number of sperm in the first row, up to 10 additional rows were counted, as necessary, to obtain a total of approximately 225 sperm per hemacytometer count.

In ancillary experiments with donor semen at the central andrology laboratory, comparisons were made for different methods of wetting the hemacytometer pillars. Studies were conducted comparing pillar wetting droplets of 1.5 and 10 μL of water on each pillar. Additional studies assessed the effect of having neither, one, or both hemacytometer pillars wet. For these studies, the size of the wetting droplet was standardized to 1.5 μL, a volume that just spread to the ends of the pillars when the coverslip was applied.

Comparison of MicroCell Chambers and Hemacytometers Using Standard Concentrations of Microbeads—During a training session at the central andrology laboratory, 5 trainees counted the standard Accubead concentrations (18×10^6/mL and 35×10^6/mL), using both the MicroCell chamber and the hemacytometer. At the time, none of the trainees were familiar with the product, which is designed for QC of CASA measurement, so they were not biased as to the expected concentration outcome. These comparisons were made before the Accubead manufacturer's instructions were changed to indicate an expected concentration difference between the MicroCell and hemacytometer chambers.

Determination of Simple Percent Sperm Motility and Progression—Raw semen was loaded into a prewarmed (37°C) MicroCell chamber and placed on the heated (37°C) microscope stage. Using a multikey hand counter and looking through the reticle grid in the microscope eyepiece, the technician enumerated the nonmotile sperm and then the motile sperm from portions of the grid for 5 or more fields, until at least 100 sperm had been scored. The portion of the grid scored at one time was larger for low-concentration specimens and was smaller for specimens with high concentrations. Sperm with any evidence of flagellar movement, whether twitching or progressing, were scored as motile. Progression was scored for the overall sample as described previously (Overstreet and Brazil, 1997). Briefly, a 1–4 scale was used, where a "1" indicates that no sperm in the sample were progressive and "4" means that most of the motile sperm in the sample had progressive motility. Half points were given for progression falling between 2 categories.

Categorical Sperm Motility Assessment—Sperm motility was also scored according to methods recommended by WHO (1999) guidelines. A 10-μL aliquot of raw semen was pipetted onto a prewarmed (37°C) plain glass slide, a coverslip was applied, and the slide was placed on the microscope stage warmer. The reticle grid was used as described above for simple motility assessments, with small sections of the grid scored at one time. All motile sperm were classified into categories: "a" for rapid progressive motility, "b" for slow or sluggish progressive motility, and "c" for nonprogressive motility (WHO, 1999). Because of the complexity of calculating the percent difference between assessments of sperm in different categories when more than 100 sperm are scored, the percent difference between the values obtained for the first 2 aliquots was not determined. Instead, 3 aliquots of semen were routinely scored; a total of 300 or more sperm were thus categorized. Percentages of sperm in each category were determined from the mean of the 3 replicates during data analysis.

Data Analysis

Paired *t* tests were used to compare mean values for methods of determining semen volume, sperm concentration, and sperm motility. The variance in sperm concentration determined by MicroCell and hemacytometer was compared by the *F* test. For the ancillary studies, mean values of paired comparisons were

Table 1. *Semen volume determined by pipetting and by weight at 4 sites participating in a multicenter study*

Site	n	Pipetted Volume* (mL)	Volume by Weight* (mL)	Difference (mL)
1	130	3.4 ± 1.7	3.7 ± 1.8	0.3
2	159	3.1 ± 1.4	3.9 ± 1.5	0.8
3	160	3.3 ± 1.6	3.8 ± 1.7	0.5
4	40	2.8 ± 1.1	3.2 ± 1.2	0.4
All sites	509	3.2 ± 1.6	3.7 ± 1.6	0.5

* Values are mean ± SD. The pipetted volume was consistently lower than the volume by weight ($P < .0001$ at each site).

tested by the paired *t* test. Repeated-measures analysis of variance was used to evaluate hemacytometry with dry, one-wet, or both-wet pillars. Analysis of variance with technician as a random factor was used to test the effect of chamber type on counts of microspheres.

Results

Semen Volume

The pipetted volume was consistently lower than the calculated volume from weight ($P < .0001$ at each site), although the mean difference between the 2 volume measurements varied among sites from a high of 0.8 mL to a low of 0.3 mL (Table 1). At mid-study, the difference between the 2 measurements at one site (site 2, Table 1) was greater than at the other sites, and efforts were made to understand this large difference between the 2 techniques. Measurement of the weight of 1.000 mL of semen (measured with a positive displacement pipettor) from multiple semen specimens gave nearly the same weight at each site (Table 2), ruling out different density levels of semen at different sites or site-specific weighing errors. A detailed questionnaire was administered to each technician asking about specific aspects of the weighing and pipetting procedures used for study subjects. It was discovered that at site 2, the identification labels for the semen collection containers were applied after the empty container had been weighed, which caused an overestimation of the semen weight of approximately 0.1 g (label weight, 0.08 g). It was also revealed that at site 2, the semen specimen was often tilted or tipped when received by the andrology laboratory, after transport from the collection area. This likely resulted in more semen clinging to the sides of the container, with less semen available for aspiration into the pipette; therefore, the pipetted volume was underestimated. After changing these 2 factors at site 2 (ie, the time of label application and the position of the specimen [ensuring that it was upright during transportation]), the average pipetted volume was 0.4 mL less than the volume determined by weight, making it comparable to the relative weight vs volume values obtained at the other sites.

Table 2. *Weight of 1.000 mL of semen determined by technicians at 3 study sites*

Site	Weight of 1000 μL of Semen (g): n*	Mean
1	30	1.013
2	25	1.014
3	8	1.004
All sites†	63	1.012

* n indicates no. of semen samples for which measurement was made.
† Site 4 was no longer active during this data collection.

Sperm Concentration

At all sites, the sperm concentration determined by the MicroCell chamber was lower than that determined by the hemacytometer ($P < .001$) (Table 3). The variability between the 2 replicates for each concentration method, indicated by the coefficient of variation for the replicate measures, was similar for the 2 methods, although precision with MicroCell was slightly higher (CV 3.9% vs 4.4% from replicate counts). The mean values from 508 study subjects showed remarkable consistency between the replicate dilutions for both the MicroCell chamber and the hemacytometer. Mean value for the first MicroCell count was 65.5 million/mL; mean value for the second MicroCell count was 66.1 million/mL. The mean values for the 2 hemacytometer counts from these same 508 subjects were identical—81.2 million/mL for both the first and second hemacytometer dilutions. Figure 1 shows mean concentration values for replicate counts by site for both counting techniques. The mean percent difference between replicates for individual specimens (the value used to determine if a third count was necessary) was 6.2% for MicroCells and 6.6% between hemacytometer replicates.

Commercially available microbead preparations of known concentration were counted using both chambers (Table 4). As with sperm suspensions, the concentration of microbeads determined by MicroCell was lower than that determined by hemacytometer ($P < .0001$). The MicroCell values were closer to the standard concentration stated by the manufacturer; however, it must be noted that the concentrations provided by the manufacturer were

Table 3. *Difference in sperm concentration using MicroCell and hemacytometer chambers at 4 sites participating in a multicenter study; coefficients of variation were determined from replicate counts for each chamber for each specimen analyzed**†

		MicroCell		Hemacytometer	
Site	n	Mean (million/mL)	CV From Replicates	Mean (million/mL)	CV From Replicates
1	130	72.3	3.2	82.7	4.0
2	158	73.6	4.0	99.4	5.1
3	179	51.9	4.8	58.9	4.0
4	40	77.5	1.4	100.7	5.1
All sites	507	65.9	3.9	81.0	4.4

* At all sites, concentration determined by the MicroCell chamber was lower than that determined by the hemacytometer ($P < .001$).
† CV indicates coefficient of variation.

determined in part using MicroCell chambers (McNamara and Thorne, personal communication to C.B.).

In ancillary studies on the MicroCell chamber, when concentrations from undiluted semen (ie, with motile sperm, as currently suggested by the manufacturer) were compared with concentrations from fixed, diluted semen (as used in our study), the values were slightly, though significantly, lower ($P < .001$) for live than for killed sperm (65.7 ± 5.5 and 69.4 ± 5.7, respectively; mean ± SEM, n = 28).

In additional studies, when sperm concentrations were determined after loading MicroCell chambers with 2 different volumes of fixed, diluted semen (3 μL, current manufacturer's recommendation, and 7 μL, manufacturer's earlier recommendation), the concentration determined following the 7-μL loading was slightly higher ($P < .05$) than that following the 3-μL loading (59.3 ± 6.6 and 56.8 ± 6.2, respectively; mean ± SEM, n = 14).

The results of these 2 ancillary studies suggest that both of our deviations from manufacturer's current recommendation for the MicroCell (ie, use of a loading volume of 7 μL instead of 3 μL and use of immobilized sperm instead of live sperm) led to slightly higher MicroCell values. This means the concentration differences between the MicroCell chamber and the hemacytometer chamber may even be slightly greater than we reported, if the MicroCell chamber is used exactly per manufacturer's recommendation.

Because loading the MicroCell chamber involves capillary filling of a viscoelastic solution into a chamber only 20 μm deep, it is possible that the semen does not load evenly and that the sperm concentration thus differs in different areas of the chamber. In both this study and QC activities (Brazil et al, 2004), individual counts were recorded by technicians for each of the 6 fields scored within the MicroCell chamber. Most technicians recorded the individual field data in a standardized pattern onto the data sheets. For example, these technicians always analyzed top to bottom and then left to right and consistently recorded these values in order on the data sheets. Other technicians analyzed the same field locations but did not consistently record them in any particular order on the data sheet. When we looked at the data for those technicians who had consistently followed a pattern in recording their data, the average values from the 6 locations were remarkably consistent (Table 5). In an ancillary study, counts were made from areas very near the edges of the chamber (ie, the areas not routinely scored). The

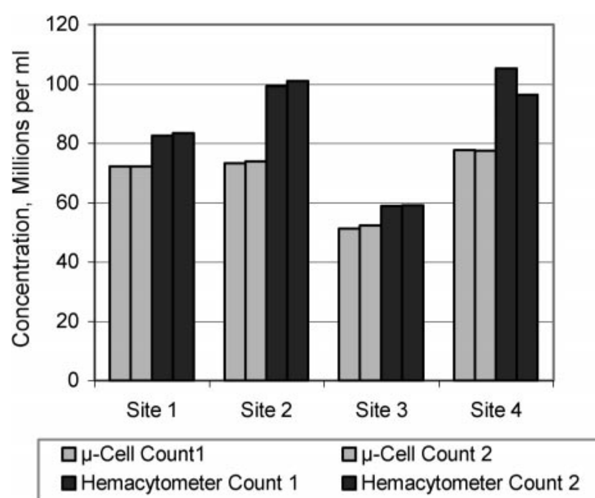

Figure 1. Replicate values for sperm concentration obtained with the MicroCell counting chamber and with the hemacytometer chamber at 4 different study sites. Mean values are shown for first and second replicate sperm counts obtained with the MicroCell counting chamber and with the hemacytometer at each site. N = 130 semen samples at site 1, 158 at site 2, 179 at site 3, and 40 at site 4.

Table 4. *Concentration of Accubeads using MicroCell and hemacytometer chambers determined during training at the central andrology laboratory*

Standard Bead Concentration (million/mL)	Concentration Counted* (million/mL)	
	MicroCell	Hemacytometer
18	17.9 ± 1.9	23.5 ± 3.9
35	36.0 ± 3.3	45.4 ± 4.9

* Values are mean ± SD for 14 blinded counts, 2–3 from each of 5 study technicians.

Table 5. *Distribution of sperm in different fields routinely counted in the MicroCell chamber**

Field 1, Nearest Opening for Loading	Field 2, Center of Chamber	Field 3, Distal End Opposite the Loading Area	Field 4, Top	Field 5, Center	Field 6, Bottom
45 ± 32†	45 ± 32	45 ± 31	45 ± 31	45 ± 32	45 ± 32

* All 6 fields are taken from the main chamber area, avoiding the outer 2–3 mm of each edge.
† Mean ± SD for MicroCell counts made during the multicenter study, n = 420. Includes data only from those technicians who were confident they had recorded a given field location in a specific data entry pattern.

extreme top and bottom edges of the chamber showed a small increase in sperm concentration, and the innermost edge showed a rather large increase in sperm concentration relative to the more centrally located areas (data not shown). However, these areas of differing sperm concentration were very small, were restricted to within 1 microscopic field of the edges of the chamber, and amounted to less than 5% of the total chamber area. Even if these areas were included in routine counts, they would account for less than a 10% difference in the determined sperm concentration. Thus, loading bias alone cannot fully explain the differences in sperm concentration determined with MicroCell and hemacytometer chambers.

WHO (1999) guidelines have recommended that the pillars of the hemacytometer be wetted before placing the cover glass on them. The resulting capillary force is believed to hold the cover glass closer to the pillars, providing a more standard chamber depth. In an ancillary study, sperm concentration was not significantly different ($P > .05$) if both pillars were dry (manufacturer's recommendation and common practice), if 1 pillar was wet with 1.5 μL of water before loading (similar to the recommendations of WHO [1999]), or when both pillars were wet with 1.5 μL of water, as in our study (86.3 ± 8.6, 84.0 ± 8.2, and 80.0 ± 8.2 × 10^6/mL, respectively; mean ± SEM, n = 15). In a separate study, determined concentrations were not significantly different when pillars were wet with 1.5 vs 10 μL of water (82.3 ± 9.0 and 85.8± 8.8 × 10^6/mL, respectively; mean ± SEM, n = 15). Although neither of the pillar-wetting studies yielded differences that were statistically significant, there seemed to be a trend toward lower concentrations when both pillars were wet with small volumes of water and a suggestion of greater sperm concentration with either greater pillar wetting volume or no wetting of the pillars (ie, too much or no water on the pillars).

Sperm Motility

Mean simple percent motilities were lower, though not significantly so, than those derived by enumerating sperm in the 4 WHO motility categories, considering percent motile sperm to be the sum of "a," "b," and "c" (51.6% vs 58.1%) (Table 6). During semen analysis, the WHO motility was performed immediately before the simple motility assessments. Although a minute or two separated each motility assessment, it does not appear that timing contributed to the lower values for the simple method, as the overall mean values for the triplicate WHO motility measurements for 509 subjects, in order of analysis, were 57.6%, 58.2%, and 58.5%. Immediately following these 3 assessments, the 2 consecutive simple motility assessments averaged 51.6% and 51.5%. Simple motility was more precise, having lower coefficients of variation between replicates ($P < .0001$) than the collapsed values obtained with the categorical method.

Table 6. *Comparison of sperm motility determined by progression categories and by simple percent at 4 sites participating in a multicenter study*

		Categorical Motility* "a + b + c"		Simple Motility	
Site	n	Mean (%)	Mean per Sample CV† (%)	Mean (%)	Mean per Sample CV† (%)
1	130	58.5	5.0	53.7	3.3
2	159	58.8	8.1	51.7	5.3
3	180	56.9	10.1	48.7	3.9
4	40	58.8	3.5	56.2	2.2
All sites	509	58.1	7.7	51.6	4.1

* Categories "a, b, and c" as defined by the World Health Organization (WHO, 1999).
† Mean sample CV is the average of the coefficient of variation values for the replicate determinations made on each individual semen sample. Simple motilities were lower than those derived by collapsing WHO categories "a + b + c," although the values were not significantly different.

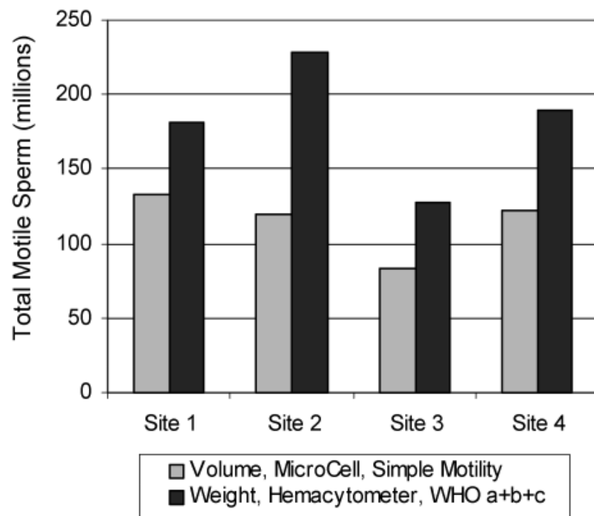

Figure 2. Mean total motile sperm (TMS) determined using different techniques at 4 different study sites. TMS as measured by pipetted volume, MicroCell concentration, and simple percent motility was much lower than TMS measured by weight, hemacytometer concentration, and collapsed categorical motility ($P < .0001$ at each site).

Total Motile Sperm

To demonstrate the consequence of measuring semen parameters by different methods, the mean number of total motile sperm (TMS) per ejaculate was calculated by multiplying the semen volume, sperm concentration, and percent motility. First, the calculation was performed using the data for pipetted volume, MicroCell concentration, and simple percent motility. The calculation was performed again for the same semen samples using the data for volume by weight, hemacytometer concentration, and collapsed categorical motility (a + b + c). At all sites, TMS, as measured by pipetted volume, MicroCell concentration, and simple motility, was much lower than TMS measured by weight, hemacytometer concentration, and collapsed categorical motility ($P < .0001$ at each site, Figure 2).

Discussion

Because of its complexity and subjectivity, semen evaluation is extremely difficult to standardize across laboratories. In this multicenter study, we emphasized detailed laboratory protocols, extensive one-on-one training of technicians, continuous proficiency testing with reliable QC materials, and thorough review of all data forms and materials received at the central andrology laboratory (Brazil et al, 2004). To our knowledge, our data comparing the results obtained with 2 counting chambers using standardized techniques are the most extensive available to date. We obtained a high level of precision with both methods, although concentrations were much higher using the hemacytometer chamber.

The literature provides conflicting reports regarding the accuracy, precision, and comparability of the various chambers used to determine sperm concentration. Latex microbeads (Accubeads) have been used as a standard for comparison and have been reported to have the same concentration when analyzed using MicroCell chamber or hemacytometer (Ginsburg and Armant, 1990). However, some investigators have reported that hemacytometers are more accurate in measuring Accubead concentrations (Sokol et al, 2000), and yet others have reported that MicroCell chambers are the more accurate chamber (Johnson et al, 1996). These conflicts demonstrate the need for studies with large numbers of samples and strict QC procedures.

The higher values for sperm concentrations determined by hemacytometer, as reported in the present study, are consistent with our QC data obtained from routine proficiency testing during the study (Brazil et al, 2004) as well as with data obtained with Accubeads during training. Although we modified the manufacturer's current recommendation on use of the MicroCell chamber, both of our modifications, if anything, seemed to cause a slight increase in the MicroCell final concentration; without these changes, the difference between the 2 chambers may even be slightly greater. Others have reported similar higher counts with hemacytometers when counting sperm or latex beads (Seaman et al, 1996; Keel et al, 2000; Tomlinson et al, 2001), although there is at least 1 report that hemacytometer counts are lower than MicroCell counts (Mahmoud et al, 1997). Some studies comparing counting chambers have shown that the hemacytometer has lower variability than other chambers tested (Peters et al, 1993; Shiran et al, 1995; Sokol et al, 2000; Tomlinson et al, 2001), yet others have reported the hemacytometer to have higher variability (Seaman et al, 1996). Clearly, laboratories vary in their application of these sperm counting chambers, causing apparent conflicting conclusions.

The results of the present study show that, under tightly controlled conditions, both the MicroCell and the hemacytometer can provide very precise results. We found the precision associated with the MicroCell to be slightly better than that of the hemacytometer. Possible explanations for the greater precision of the MicroCell include the fact that the chamber is fixed and requires less manipulation than the hemacytometer. The hemacytometer is an open chamber and, as such, is less stable and more likely to be affected by variations in the evaporation that results from environmental conditions as well as the amount of time elapsed between filling and analyzing the chamber. In addition, for our study, the MicroCell required a 1:1 dilution, which is less likely to result in errors than the 1:20

Table 7. *Potential sources of variability between MicroCell chambers and hemacytometers*

Source of Variation	Possible Effect
Loading bias of MicroCell chamber	When loading the MicroCell, the flow dynamics of the chamber cause a collection of sperm along the distal wall. This could lead to an underestimation of the sperm concentration.
Manufacturer allowed 10% variability in MicroCell chamber depth	Can cause a 10% over- or underestimation of the concentration
Hemacytometer pillar wetting or not	May cause an overestimation if pillars are not secured
Hemacytometer pillar wetting volume	May cause an additional overestimation if wetting volume is too high
Dilutions	Pipettor inaccuracies are less likely to cause erroneous results with the 1 + 1 dilution used for the MicroCell than with the 1 + 19 dilution used for the hemacytometer.
No. sperm counted	In this study, the number of sperm counted with the 2 chambers was different. Using our hemacytometer protocol, approximately 200–225 sperm were counted for each count, regardless of the concentration. The average number of sperm counted for each count using our MicroCell protocol was approximately 250 sperm; however, the number counted varied according to the sperm concentration and was highest when the sperm concentration was highest.

dilution used for the hemacytometer. The variability between replicates for both chambers is surprisingly low, given that it reflects the sampling error (from the semen specimen as well as from both dilution vials) as well as the intratechnician variability associated with performing the technique. However, it must be noted that these replicate counts were not performed blindly, and it is possible that the second value was influenced by the first. Given the closeness of the replicate count and motility assessments, the need for such duplicate assessments in population studies can be questioned, but this precision is likely attainable only with strict training and protocol standardization.

The hemacytometry method used in this study was modified from WHO (1999) protocol to simplify the dilutions and calculations for determining the final sperm concentration. The WHO (1999) recommendation was followed with respect to wetting the pillars, but in our study, both pillars were wet. Technicians were instructed and trained to use a 1- to 2-μL drop of water to secure the coverslip. However, later evaluations determined that some technicians modified this instruction and either used a drop of water on their finger to wet the pillar or made their best estimation of a 1- to 2-μL droplet from a pipettor set to a higher volume. While the differences were not significant, the trends seen in our ancillary studies suggest that the wetting of the pillar, both in terms of the volume of water applied and whether one or both pillars are wet, is an important detail that requires further investigation. In retrospect, the details of wetting the pillars should have been better standardized for this study and given more emphasis during training. Variability from this source may have contributed to the lower precision of the hemacytometer. Historically, the recommendation to wet the pillars comes from a time when the pillars were smooth glass, and the coverslips were lightweight (Hauser Scientific, personal communication). According to the hemacytometer manufacturer, the coverslips used currently are heavy enough that wetting of the pillars is no longer necessary to hold the coverslip in place. Additionally, many hemacytometers manufactured currently have ground glass pillars (including those used for the current study). It is unclear how much these changes have affected the ability to form a seal. Certainly, the recommendation to observe Newton rings (WHO, 1999) is not possible when using hemacytometers with ground glass pillars.

The reason(s) that hemacytometer chambers give consistently higher sperm concentration values than those from MicroCell chambers is unknown. Both chambers are based on the principle of enumerating the sperm in a known volume within the chamber grid. It is interesting that although all sites showed this difference between the chambers, the relative percent difference between the 2 chambers was not consistent among sites. Table 7 details some differences between the techniques that may have contributed to the discrepancies in sperm concentration determined with the 2 chambers. Although we cannot conclude which chamber is more accurate, we found that the MicroCell technique is best suited to multicenter studies for 3 reasons: 1) the slightly greater precision obtained with this chamber from 509 semen evaluations, 2) the intra- and intertechnician coefficient of variation data obtained during the QC activities for the study (Brazil et al, 2004), and 3) the ease and simplicity of learning, using, and standardizing the MicroCell method. It should be noted that although the manufacturer's recommendation is to use the MicroCell with live sperm for concentration as-

sessment, in our experience, it is much more challenging to enumerate moving sperm, which would likely present bigger problems for standardization in a multicenter study. Therefore, we recommend using the MicroCell with immobilized sperm for concentration assessment in multicenter studies.

Our data suggest that the least variable semen volume assessments can be determined by weighing the semen specimen in its original collection container and subtracting the previously determined empty container weight. If pipetting is used to determine semen volume, extreme care must be taken to keep the specimen container upright after collection, until volume measurements are made.

We found categorical assessment of the percentage of motile sperm to be less precise than simple percent motility determination. It must be emphasized that each of these methods counts all sperm with evidence of motility as motile and is therefore not comparable to an estimate of the percentage of *progressively motile* sperm (ie, a + b by the WHO guidelines). In the present study, we showed that the simple method provided data that were comparable to the percentage of sperm counted by the categorical method in the combined ''a plus b plus c'' categories defined by WHO guidelines. This finding suggests that data obtained by these 2 methods of motility assessment can be compared between studies—for example, between the SFF and the European studies of fertile men (Jørgensen et al, 2001). A similar approach of collapsing categories has been used previously to compare data obtained in different laboratories (Dunphy et al, 1989; Jørgensen et al, 1997; Auger et al, 2000).

The data presented in this communication demonstrate that caution must be used when comparing semen data collected in different laboratories or in the same laboratory in which methods and/or personnel have changed over time, or even when data are evaluated from multicenter studies where strict standardization and QC were not followed. This point is well illustrated using data from the current study that were obtained when the same semen samples were evaluated by different methods for measuring semen volume, sperm concentration, and percent motility. At all sites, TMS, as measured by pipetted volume, MicroCell concentration, and simple percent motility, was much lower than TMS measured by weight, hemacytometer concentration, and collapsed categorical motility (a + b + c). These data clearly demonstrate the critical need to rigorously standardize protocols and techniques for multicenter studies. For these types of studies, we recommend assessing semen volume by weight, sperm concentration by MicroCell chamber, and sperm motility by simple categorization as motile or nonmotile.

Acknowledgments

In addition to the authors, the Study for Future Families Research Group includes: from the University of Missouri, Columbia: B. S. Carter, RN; D. J. Kelly, RN; R. L. Kruse, PhD; S. L. Stewart, BA; and T. M. Simmons, BS; from the University of California, Davis: C. Treece, CLA; and C. Tollner, BS; from the Harbor-UCLA Medical Center, Torrance: R. S. Swerdloff, MD; L. Lumbreras, PA; S. Villanueva, RN, NP; M. Diaz-Romero, MD; A. Victoroff, PA; R. Sandoval, BA; S. Baravarian, PhD; A. Leung, HTC; and A. L. Nelson, MD; from the Cedars-Sinai Medical Center, Los Angeles: C. Hobel, MD; and B. Brock, MD; from the Mt Sinai School of Medicine, New York City: M. Hatch, PhD; M. Pfeiffer, MS; L. Quinones, BA; K. Polgar, PhD; and A. Brembridge, MPH; and from the University of Minnesota, Minneapolis: C. Kwong, MPH; A. Muehlen, BA; T. Perrier, MLT; T. Srb, BS; J. Pryor, MD; and C. De Jonge, PhD.

We also acknowledge the valuable scientific guidance provided by Dr Gwen Collman, National Institute of Environmental Health Sciences, National Institutes of Health, and the contributions of the physicians, midwives, and staff of the University Physicians Clinic, Columbia, Mo; the Fairview Riverside Women's Clinic, Minneapolis, Minn; the Harbor-UCLA Medical Center, Torrance, Calif; the Cedars-Sinai Medical Center, Los Angeles, Calif; the Mt Sinai Medical Center, New York, NY; and participants in the Study for Future Families.

References

Auger J, Eustache F, Ducot B, et al. Intra- and inter-individual variability in human sperm concentration, motility and vitality assessment during a workshop involving ten laboratories. *Hum Reprod.* 2000;15:2360–2368.

Björndahl L, Kvist U. 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.* New York, NY: Parthenon Publishing Group; 1998:91–101.

Brazil C, Swan S, Tollner C, Treece C, Drobnis E, Redmon JB, Wang C, Overstreet J. Quality control of laboratory methods for semen evaluation in a multicenter research study. *J Androl.* 2004;25:645–656.

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

Dunphy BC, Kay R, Barratt CLR, Cooke ID. Quality control during the conventional analysis of semen, an essential exercise. *J Androl.* 1989; 10:378–385.

Fisch H, Goluboff E. Geographic variations in sperm counts: a potential cause of bias in studies of semen quality. *Fertil Steril.* 1996;65:1044–1046.

Fisch H, Goluboff ET, Olson JH, Feldshuh J, Broder SJ, Barad DH. Semen analyses in 1,283 men from the United States over a 25-year period: no decline in quality. *Fertil Steril.* 1996;65:1009–1014.

Ginsburg G, Armant D. The influence of chamber characteristics on the reliability of sperm concentration and movement measurements obtained by manual and videomicrographic analysis. *Fertil Steril.* 1990; 53:882–887.

Guzick DS, Overstreet JW, Factor-Litvak P, et al. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med.* 2001:345:1388–1393.

Jarow JP, Sharlip ID, Belker AM, et al. Best practice policies for male infertility. *J Urol.* 2002;167:2138–2144.

Johnson JE, Boone WR, Blackhurst DW. Manual versus computer-automated semen analysis. Part 1. Comparison of counting chambers. *Fertil Steril.* 1996;65:150–155.

Jørgensen N, Anderson AG, Eustache F, et al. Regional differences in semen quality in Europe. *Hum Reprod.* 2001;16:1012–1019.

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

Keel B, Quinn P, Schmidt C, Serafy N Jr, Serafy N Sr, Schalue T. Results of the American Association of Bioanalysts national proficiency testing programme in andrology. *Hum Reprod.* 2000;15:680–686.

Mahmoud H, Depoorter B, Piens N, Comhaire F. The performance of 10 different methods for the estimation of sperm concentration. *Fertil Steril.* 1997;68:340–345.

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

Overstreet JW, Brazil CK. Semen analysis. In: Lipshultz L, Howards S, eds. *Infertility in the Male.* St Louis, Mo: Mosby; 1997:487–490.

Paulsen CA, Berman NG, Wang C. Data from men in greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. *Fertil Steril.* 1996;65:1015–1020.

Peters AJ, Zaneveld LJ, Jeyendran RS. Quality assurance for sperm concentration using latex beads. *Fertil Steril.* 1993;60:702–705.

Punjabi U, Spiessens C. 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.* New York, NY: Parthenon Publishing Group; 1998:107–113.

Seaman E, Goluboff E, BarChama N, Fisch H. Accuracy of semen counting chambers as determined by the use of latex beads. *Fertil Steril.* 1996;66:662–665.

Shiran E, Stoller J, Blumenfeld Z, Feigin P, Makler A. Evaluating the accuracy of different sperm counting chambers by performing strict counts of photographed beads. *J Assist Reprod Genet.* 1995;12:434–442.

Sokol R, Shulman P, Paulson R. Comparison of two methods for the measurement of sperm concentration. *Fertil Steril.* 2000;73:591–594.

Swan SH, Brazil CK, Drobnis E, et al. Geographic differences in semen quality of fertile US males. *Environ Health Perspect.* 2003;111:414–420.

Swan SH, Elkin EP, Fenster L. Have sperm densities declined? A reanalysis of global trend data. *Environ Health Perspect.* 1997;105:1228–1232.

Swan SH, Elkin EP, Fenster L. The question of declining sperm density revisited: an analysis of 101 studies published 1934–1996. *Environ Health Perspect.* 2000;108:961–966.

Tomlinson M, Turner J, Powell G, Sakkas D. One-step disposable chambers for sperm concentration and motility assessment: how do they compare with the World Health Organization's recommended methods? *Hum Reprod.* 2001;16:121–124.

Vierula M, Niemi M, Keiski A, Saaranen M, Saarikoski S, Suominen J. High and unchanged sperm counts of Finnish men. *Int J Androl.* 1996;19:11–17.

World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction.* 4th ed. Cambridge, United Kingdom: Cambridge University Press; 1999.