ORIGINAL ARTICLE: ANDROLOGY

Scoring human sperm morphology using Testsimplets and Diff-Quik slides

Ilaria Natali, Ph.D.,[a] Monica Muratori, Ph.D.,[b] Valentina Sarli, B.S.,[a] Mario Vannuccini, Ph.D.,[c] Sonia Cipriani, Sc.D.,[d] Laura Niccoli, M.D.,[a] and Claudia Giachini, Ph.D.[a]

[a] Sterility Center, Obstetric and Gynecology Unit, S.S. Cosma and Damiano Hospital, Pescia; [b] SOD Sexual Medicine and Andrology, University of Florence, Florence; [c] Eureko, Pistoia; and [d] Dipartimento Materno-Infantile, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy

Objective: To compare two staining methods to assess sperm morphology: Diff-Quik (DQ), which is the fastest of the recommended techniques, and Testsimplets (TS), a technique that uses prestained slides and is quite popular in in vitro fertilization (IVF) centers.
Design: Prospective study.
Setting: Patients at the Sterility Center of the Obstetrics and Gynecology Unit of the Hospital of S.S. Cosma and Damiano (Azienda USL 3 of Pistoia, Italy).
Patient(s): 104 randomly enrolled male patients evaluated by the seminology laboratory.
Intervention(s): None.
Main Outcome Measure(s): Statistical comparison of sperm morphology results obtained after staining of semen samples both with DQ and TS.
Result(s): Our data show that TS gives a statistically significantly lower number of normal forms than DQ (median: 6% [range: 0–29%] vs. 12% [range: 0–40%], respectively) as well as an overestimation of sperm head defects (median: 92.0% [range: 67%–100%] vs. 82.3% [range: 55%–100%], respectively).
Conclusion(s): The two staining methods should not be considered equivalent. Specifically, the lower reference limit established by the World Health Organization is not appropriate when sperm morphology is assessed by TS. The routine application of TS in the evaluation of sperm morphology is therefore not recommended because it leads to an overestimation of patients with sperm morphology values below the lower reference limit (4%), thus potentially influencing clinical decisions. (Fertil Steril® 2013;■:■–■. ©2013 by American Society for Reproductive Medicine.)
Key Words: Diff-Quik, sperm morphology, staining methods, Testsimplets

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/natalii-scoring-human-sperm-morphology/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

For routine semen analysis, reference values for sperm concentration, motility, and morphology are provided by the World Health Organization (WHO) laboratory manual for the examination and processing of human semen (1). In the most recent version of the WHO manual, the reference values were determined by a large multicenter study that recruited 4,500 men from 14 different countries whose partners had a time to pregnancy ≤12 months (2). According to this study, the lower reference limit for each of the semen parameters corresponded to the 5th percentile (95% confidence interval [CI]) observed in the study population. For sperm morphology, the lower reference limit value decreased from 15% (3) to 4% (1).

Received July 20, 2012; revised and accepted November 26, 2012.
I.N. has nothing to disclose. M.M. has nothing to disclose. V.S. has nothing to disclose. M.V. has nothing to disclose. S.C. has nothing to disclose. L.N. has nothing to disclose. C.G. has nothing to disclose.
Reprint requests: Ilaria Natali, Ph.D., Sterility Center, Obstetric and Gynecology Unit, S.S. Cosma and Damiano Hospital, Via Cesare Battisti 32, Pescia, Italy (E-mail: i.natali@usl3.toscana.it).

Fertility and Sterility® Vol. ■, No. ■, ■ 2013 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2012.11.047

Spermatozoa are defined as morphologically normal on the basis of those recovered from the female reproductive tract, especially in uterine lumen, the fallopian tubes (4), and the postcoital endocervical mucus (4–6). In addition, it has been reported that normal spermatozoa according to the strict criteria (7) can be selected in vitro by the presence of the hyaluronic acid (HA) binding sites on their surface (8).

The association between normal sperm morphology and the pregnancy rate in vitro and in vivo has been reported by several investigators (9, 10), thus suggesting that morphology assessment is critical predictive

information for supporting clinical decisions. However, other investigators have questioned the role of sperm morphology in in vitro fertilization (IVF) because the percentage of normal spermatozoa assessed by strict criteria does not correlate with the outcome of the IVF cycles (11–13).

According to Mortimer and Menkveld (14), Papanicolaou staining is the recommended method for the best morphology assessment in routine application. This method was originally introduced for vaginal cytology and then was modified to enable the evaluation of sperm morphology (15). The World Health Organization manual (1) admits also the use of Shorr (16) and Diff-Quik (17) staining techniques. These techniques have the advantage of being fast, but they provide fewer sperm details than Papanicolaou staining (14). No statistically significant difference has been observed when comparing the Papanicolaou and the DQ staining methods (17, 18) using both washed and nonwashed samples (18), and extensive agreement has also been found between Papanicolaou and Shorr stained smears (19). All these methods accurately stain each region of the spermatozoa: the sperm head stains pale blue in the acrosomal region and dark blue in the postacrosomal region, the midpiece may show some red staining, and the tail stains blue or reddish (1).

Testsimplets (TS), another commercially available staining method (20), is a simple, rapid procedure that uses precolored slides and does not require the use of chemical reagents. Although TS is not included among the methods recommended by the WHO manual, it has become popular in IVF centers (21), including Italian centers, as revealed by the recent survey within the External Quality Control (EQC) program for seminology performed in the Tuscany region (personal communication by the Quality and Security Unit, AOUC Careggi, Florence, Italy). Only a few studies have compared TS with the recommended staining methods (19, 22), so further data are needed. We compared DQ and TS, two staining methods used for sperm morphology assessment, to evaluate the consequences of using TS in male infertility workups.

MATERIALS AND METHODS

Our study enrolled 104 male patients attending the Seminology Laboratory at the Sterility Center of the Obstetrics and Gynecology Unit of the Hospital of S.S. Cosma and Damiano (Azienda USL 3 of Pistoia, Italy). The evaluation of seminal fluid was performed as part of a couple's infertility assessment ($n = 95$) or during an assessment for other andrologic diseases ($n = 9$). The ages of the men were between 19 and 57 years (mean 34.7 $\pm$ 6.1 standard deviation [SD] years). The study protocol was approved by the local ethics committee, and all participants gave their informed consent before inclusion in the study.

Semen samples were collected by masturbation at the hospital after 2 to 7 days' abstinence from intercourse. The semen analysis was performed according to the current WHO guidelines (1). The mean sperm concentration and progressive motility (rapid and slow) determined in our group of samples were 27.9 $\pm$ 25.0 $\times$ 10^6/mL and 41.3% $\pm$ 15.0%, respectively. For the morphology assessment, each sample was stained using the two methods: the Diff-Quik (DQ) staining technique (Diff-Quik staining kit; Origio Italia) and stain-coated Testsimplets (TS) slides (Waldeck GmbH & Co. KG).

For the DQ method, we followed the WHO manual guidelines (1). We smeared 10 μL of semen on a slide, which was fixed by immersion in triarylmethane fixative for 15 seconds after complete air drying. The smears were then consecutively stained by solution 1 (10 seconds), then air-dried and stained by solution 2 (5 seconds). Finally, the slides were washed in running tap water to remove the excess stain (10 to 15 times) (1). The stained slides were read at $\times$1,000 magnification with oil immersion (Leica Microsystems) within 5 hours of their preparation.

For the TS method, 10 μL of semen were placed on the prestained slide and then covered with the coverslip. The sample was read at $\times$1,000 magnification with oil immersion within 1 hour of slide preparation.

We examined 200 spermatozoa in two replicates for each staining method. The evaluation of the percentage of morphologically normal and abnormal spermatozoa was performed according to the WHO guidelines (1). In addition, among the spermatozoa with abnormal morphology, the percentage of head, midpiece, and flagellum defects was monitored.

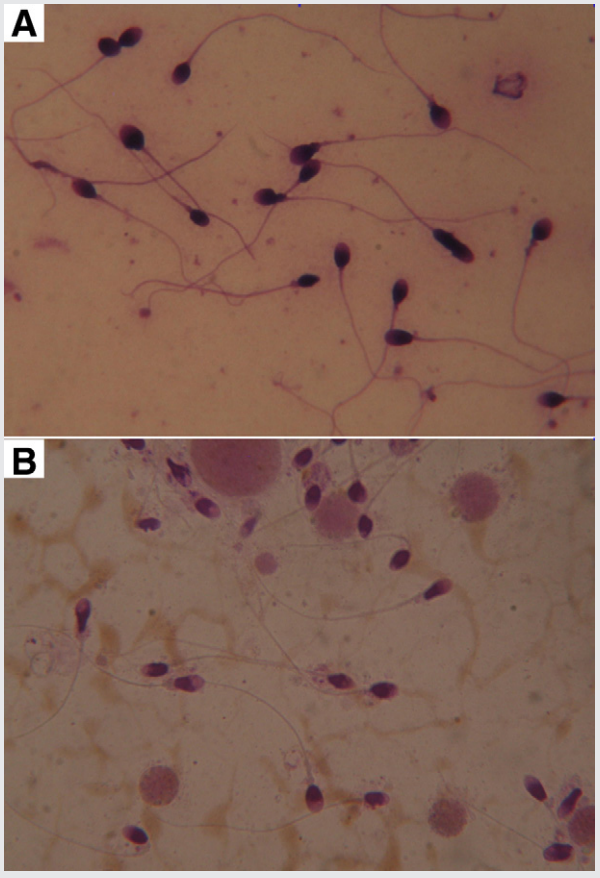
The morphology score for both DQ and TS stained samples was performed through blinded slide analyses by the same operator. The latter participates in the External Quality Assurance Programme operated by the European Society of Human Reproduction and Embryology (ESHRE) Andrology Special Interest Group (SIGA) for the assessment of sperm concentration, motility, morphology, and vitality.

Statistical Analysis

The statistical analysis was performed using SPSS software (version 17.0; SPSS, Inc.) and Analyse-it (Evaluation Edition for Microsoft Excel; Analyse-it Software, Ltd.). The nonparametric Wilcoxon test for paired samples was used to compare the mean values obtained by the two methods. The statistical analysis was performed by Deming regression, which accounts for measurement error in both the independent and dependent variable, yielding a regression line that minimizes the sum of the square of residual in both x and y directions simultaneously (23).

Bland and Altman plots (24) were then constructed, with the difference between the two methods plotted against the average measure of the two methods both for normal sperm morphology (%) and for head defects (%). Agreement between the two methods was determined by Cohen's kappa statistic (25–27), using the lower reference limit established by WHO—that is, 4% (1)—as the threshold value to split the study population into two groups. The kappa statistic examines the amount of agreement between two measurements by calculating the kappa value (K), ranging from -1 to 1: $K = -1$ when the agreement is less than what would be expected by chance; $K = 0$ when the agreement does not differ from what would be expected by chance. The interpretation of kappa values was assessed by the criteria of Landis and Koch (26), who defined six levels of agreement: "poor" ($K < 0.00$), "slight" ($K = 0.00$–0.20),

FIGURE 1



Stained smears of human sperm obtained using (A) Diff-Quik and (B) Testsimplets.

Natali. Testsimplets for sperm morphology assessment. Fertil Steril 2013.

"fair" (K = 0.21–0.40), "moderate" (K = 0.41–0.60), "substantial" (0.61–0.90), and "almost perfect" (K = 0.81–1.00).

RESULTS

Typical staining patterns obtained by the two methods (DQ and TS) are shown in Figure 1. In Table 1 we report the mean, the standard deviation, the median, and the range values of the percentages of normal sperm morphology and head defects as determined using DQ (nDQ and hDQ, respectively) and TS (nTS and hTS, respectively) staining methods. As shown, the percentage of normal morphology was statistically significantly lower using TS than DQ in the same study population. Similarly, the percentage of head defects was statistically significantly higher by TS than DQ (Table 1).

The compared distribution of both normal sperm morphology and head defects as obtained by the two methods is shown in Supplemental Figure 1 (available online). The distribution of normal morphology by TS shifted toward the lower percentage with respect to DQ (Supplemental Fig. 1A). Similarly, the distribution of head anomalies by TS shifted toward higher values when compared with the distribution obtained by DQ (Supplemental Fig. 1B).

To further compare the two methods, we performed a Deming regression analysis of nDQ and nTS as measured in the 104 samples (Fig. 2A). The corresponding regression (Sy|x = 4.62) equation resulted: nTS = −2.29 + 0.76 nDQ. The 95% CI was 0.58–0.94 for the slope and −4.19–0.39 for the intercept. The slope and the intercept were thus statistically significantly different from 1 and 0 (null hypothesis), respectively, indicating that the two methods are not equivalent.

We also performed a Deming regression analysis to compare the results of head defects (%). (see Fig. 2B). The percentage of head defects determined by TS (hTS) was much higher than determined by DQ (hDQ), and the following regression equation resulted: hTS = 33.52 + 0.69 hDQ; 95% CI for the slope: 0.50–0.88; 95% CI for the intercept: 17.09–49.95; Sy|x = 6.50. Hence, the slope and the intercept were statistically significantly different from 1 and 0, respectively.

The Bland-Altman plot of the differences (Fig. 3A) between the two methods confirmed that the normal morphology values obtained by TS were considerably lower than those obtained by DQ (mean difference (nTS − nDQ): −5.42 ± 0.53% [mean ± standard error of the mean]). It is interesting that the shape of the distribution suggests that the difference between the two methods proportionally increases with the magnitude of the measurement.

Similarly, for sperm head defects the Bland-Altman plot of the difference (Fig. 3B) between the methods confirmed the Deming regression results—that is, an overestimation of head defects using TS compared with DQ (difference: 8.17 ± 0.775% [mean ± standard error]). Moreover, similarly to what we had observed for sperm morphology, such a difference between the two methods slightly increases for the lower values of head defects.

The agreement between the two methods for normal sperm morphology was determined by the kappa statistics using 4% as a threshold value (1) to split the study population

TABLE 1

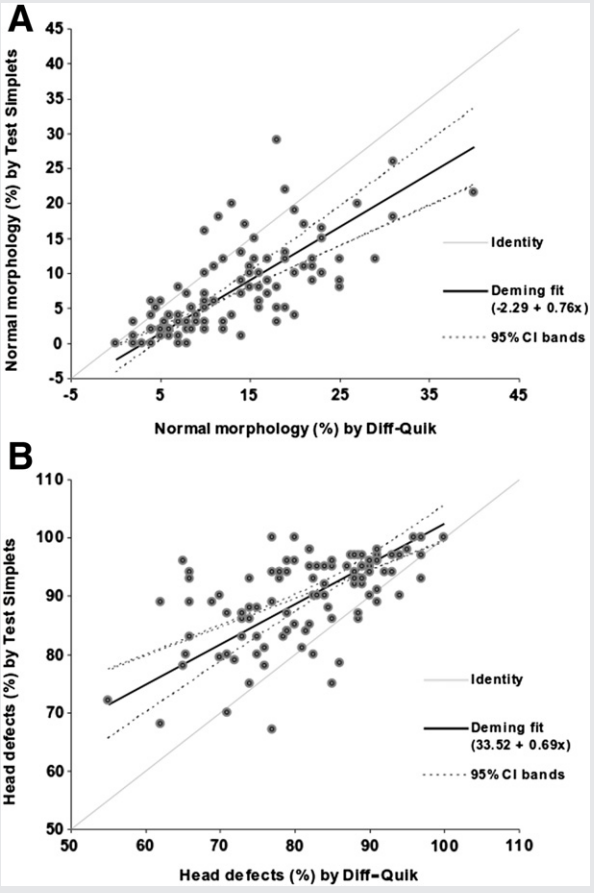
Mean and median values of normal sperm morphology (%) and head defects (%) obtained using Diff-Quik versus Testsimplets.

	Testsimplets				Diff-Quik				
Parameter	**Mean**	**SD**	**Median**	**Range**	**Mean**	**SD**	**Median**	**Range**	***P* values**
Normal morphology (%)	7.6	6.2	6	0–29	13.0	7.6	12.0	0–40	<.001
Head defects (%)	89.8	7.5	92	67–100	81.6	9.4	82.3	55–100	<.001

Note: P values are obtained using the nonparametric Wilcoxon test. SD = standard deviation.

Natali. Testsimplets for sperm morphology assessment. Fertil Steril 2013.

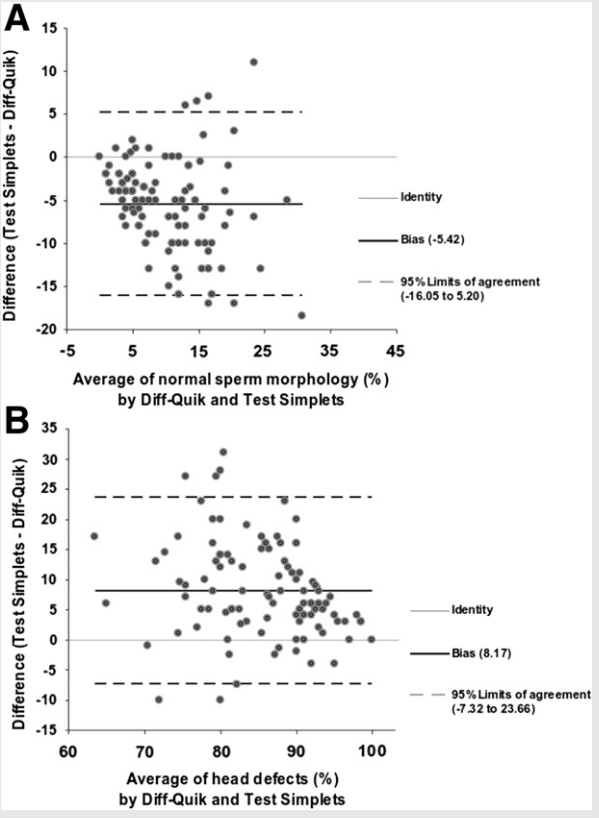
FIGURE 2



Deming's regression analysis for (A) normal sperm morphology and (B) sperm head defects determined by Diff-Quik (DQ) and Testsimplets (TS) staining methods.

Natali. Testsimplets for sperm morphology assessment. Fertil Steril 2013.

FIGURE 3



Bland-Altman plot of the difference versus the average of the values obtained by Diff-Quik (DQ) and Testsimplets (TS) for (A) sperm morphology and (B) head defect assessment. The *continuous line* (bias) represents the mean difference, and *dashed lines* represent the 95% limits of agreement (±1.96 standard deviation).

Natali. Testsimplets for sperm morphology assessment. Fertil Steril 2013.

into two groups. The Cohen K value (K = 0.22) revealed a fair degree of agreement (26) between the two staining methods (Supplemental Table 1, available online).

DISCUSSION

Accurate sperm morphology evaluation is crucial to the routine examination of semen because the percentage of morphologically normal sperm represents an important predictor of male fertilizing potential (9, 28). The reference staining method for sperm morphology assessment is the Papanicolaou technique (14), and it has been used in several studies as the reference to validate other staining methods such as Shorr and Diff-Quik.

However, some investigators have proposed another cell-staining technique, the commercially available Testsimplets (TS), as a valid alternative for the evaluation of sperm morphology (20, 29) because it provides good visualization of neck details and cytoplasmic residues (22). Presently, TS is quite popular among IVF centers (21); its speed and ease of preparation make it particularly attractive under the time constraints of assisted reproduction management. In addition, TS is very popular among IVF centers that do not have separate laboratories for andrology and embryology, who thus need to avoid the use of chemical reagents and in particular of volatile compounds that could contaminate the environment in which embryos are cultured (30).

In Italy, a recent survey within the external quality control program for seminology performed in the Tuscany region reported that 22.4% of laboratories assessing semen were using TS routinely (personal communication by the Quality and Security Unit, AOUC Careggi, Florence, Italy); the percentage rises to 71.4% when only IVF centers are included.

In our study, the TS method was compared with DQ staining, which is the fastest among the WHO-recommended techniques for sperm morphology assessment. After its validation against the reference method, DQ was introduced for general use in the 1992 edition of the WHO manual (31). At present, sperm morphology data obtained from DQ are considered to be comparable to the results of Papanicolaou staining, so DQ represents a valid, time-saving, alternative technique (17, 18).

To compare TS and DQ methods, we first used the Wilcoxon test and Deming regression analysis. Deming

regression analysis is very suitable for the comparison of methods (23). Indeed, it allows measurement error in both the x and y variables (23), and it is thus more appropriate than the most popular linear regression analysis (the latter assuming that only one variable has measurement errors). Alongside the Deming regression analysis, we constructed a Bland-Altman plot to further compare the two methods. The Bland-Altman plot is a graphical method to compare two measurements; in addition, it reveals whether there is a relationship between the differences between the methods and the magnitude of measurements (24). Finally, we used the kappa statistic to evaluate the strength of agreement between the two methods.

Our statistical evaluation revealed that the two methods for detecting sperm morphology, DQ and TS, yielded very different results. The TS method gave a lower number of normal forms compared with the results obtained by the DQ technique (see Table 1, and the Deming regression analysis). Further, this difference increased by the percentage of normal forms (see Fig. 3A). Moreover, TS also provided an overestimation of head defects (see Table 1 and the Deming regression analysis).

These data indicate that the use of TS instead of a validated method such as DQ can result in statistically significantly differences in morphology evaluations. Accordingly, the K statistics revealed a fair degree of agreement between the two staining methods. In addition, this analysis indicated that the lower reference limit for normal sperm morphology established by WHO (i.e., 4%) is not ideal for the assessment of sperm morphology by TS. In particular, using 4% as the lower reference limit for the morphology assessment by TS would result in a statistically significant overestimation of patients with a sperm morphology <5th percentile, with obvious clinical consequences for the workup and treatment of an infertile couple.

A possible reason for the discrepancy between DQ and TS could be the different preparation of the slides in the two procedures. First, TS does not require a smear of the sample. The lack of this step, together with the smaller available area for sperm spreading may provoke a poorer distribution and isolation of sperm than in the DQ method, thus hampering a proper evaluation of sperm details (as requested by the WHO manual).

Further, unlike DQ, TS does not include an air-drying step, which provokes the loss of most of the cytoplasmic droplets, osmotically sensitive vesicles (32, 33) found in a high percentage (over 50%) of normal motile sperm (33). These structures can increase the score of sperm anomalies, as they may be confused with the "excess residual cytoplasm" (34) often retained by abnormally shaped spermatozoa. Accordingly, the air-dried Papanicolaou method yields a smaller number of spermatozoa with cytoplasmic droplets than those obtained in fixed wet preparations (33), and the lack of the air-drying step results in an overestimation of head morphology anomalies in bovine spermatozoa (35). The importance of the air-drying step for sperm morphology assessment is also indicated by Katz et al. (36), who used a videomicrographic system to evaluate sperm size and shape. They reported that the sperm heads were 30% smaller in dried-stained samples than in wet preparations, suggesting that different staining methods may yield different sperm morphology scores by differently detecting cell size (37). Finally, without the step of air-drying or fixation, sperm cells are often still motile, which can be very distracting when assessing morphology.

Additional limitations of the TS procedure are [1] the pale staining of the spermatozoa, which thus poorly stands out against the background, the latter even appearing nonhomogeneous (19) (see Fig. 1), and [2] the fact that spermatozoa can lie on their sides and thus may be confused with tapering forms by inexperienced observers (38).

The finding that TS overestimates the percentage of abnormal morphology and head defects is consistent with previous studies. Ragni et al. (22) reported that the TS technique detects a higher number of sperm anomalies than Papanicolaou, modified Giemsa, Hemaquick, and hematoxylin-eosin staining techniques. Accordingly, Henkel et al. (19) found a significantly lower number of normal spermatozoa with TS compared with the Papanicolaou and Shorr procedures. Overall, the studies to date and ours are consistent in concluding that the values obtained by TS are not comparable to those obtained by the methods recommended by WHO (1) for sperm morphology assessment. Such methods rely on standardized procedures (1) aimed at minimizing the interlaboratory variability and on external quality control programs to check compliance with the standardized procedures and with reference values of normalcy (2). This is not the case with TS, resulting in differences in its execution and interpretation among different laboratories.

We have demonstrated that the TS and DQ staining methods yield statistically significantly different results for sperm morphology, and they should not be considered equivalent procedures. Consequently, the routine application of TS for the assessment of sperm morphology is not recommended.

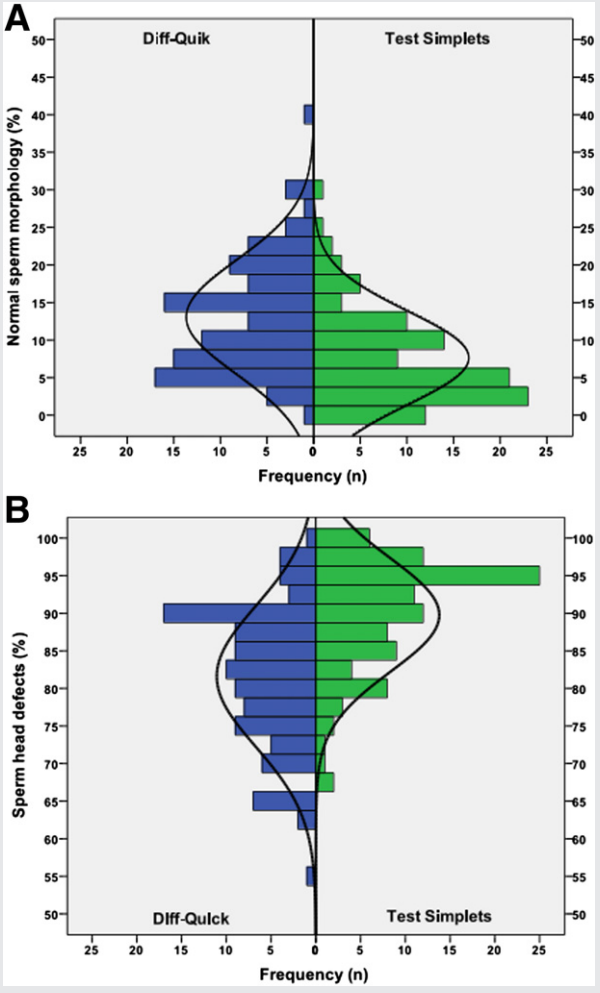
Acknowledgments: The authors thank Professor Oreste Franci for his useful statistical help and advice, and Dr. Erminio Filimberti and Selene Degl'Innocenti for their support with images, and Professor David Natali and Chiara Chianese for editing the manuscript.

REFERENCES

1. World Health Organization. Laboratory manual for the examination and processing of human semen. 5th ed. Geneva: World Health Organization; 2010.
2. Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HWG, Behre HM, et al. World Heath Organization reference values for human semen characteristics. Hum Reprod Update 2010;16:231–45.
3. World Health Organization. Laboratory manual for the examination of human semen and sperm-cervical mucus interaction. 4th ed. New York: Cambridge University Press; 1999.
4. Mortimer D, Leslie EE, Kelly RW, Templeton AA. Morphological selection of human spermatozoa in vivo and in vitro. J Reprod Fert 1982;64:391–9.
5. Fredricsson B, Björk G. Morphological of postcoital spermatozoa in the cervical secretion and its clinical significance. Fertil Steril 1977;28:841–5.
6. Menkveld R, Stander FS, Kotze TJ, Kruger TF, van Zyl JA. The evaluation of morphological characteristics of human spermatozoa according to strict criteria. Hum Reprod 1990;5:586–92.
7. Kruger TF, Menkveld R, Stander FSH, Lombard CJ, Van der Merwe JP, van Zyl JA, et al. Sperm morphologic features as a prognostic factor in in-vitro fertilization. Fertil Steril 1986;46:1118–23.

8. Prinosilova P, Kruger T, Ozkavukcu S, Vigue L, Kovanci E, Huszar G. Selectivity of hyaluronic acid binding for spermatozoa with normal Tygerberg strict morphology. Reprod Biomed Online 2009;18:177–83.
9. Coetzee K, Kruger TF, Lombard CJ. Predictive value of normal sperm morphology: a structured literature review. Hum Reprod Update 1998;4: 73–82.
10. Ombelet W, Flourie F, Vandeput H, Bosmans E, Cox A, Janssen M, et al. Teratozoospermia and in vitro fertilization: a randomized prospective study. Hum Reprod 1994;9:1479–84.
11. Berger DS, AbdelHafez F, Russell H, Goldfarb J, Desai N. Severe teratozoospermia and its influence on pronuclear morphology, embryonic cleavage and compaction. Reprod Biol Endocrinol 2011;9:37.
12. French DB, Sabanegh ES Jr, Goldfarb J, Desai N. Does severe teratozoospermia affect blastocyst formation, live birth rate, and clinical outcome parameters in ICSI cycles? Fertil Steril 2010;93:1097–103.
13. Svalander P, Jakobsson AH, Forsberg AS, Bengtsson AC, Wikland M. The outcome of intracytoplasmic sperm injection is unrelated to "strict criteria" sperm morphology. Hum Reprod 1996;11:1019–22.
14. Mortimer D, Menkveld R. Sperm morphology assessment—historical perspectives and current opinions. J Androl 2001;22:192–205.
15. Papanicolaou GN. A new procedure for staining vaginal smears. Science 1942;95:438–9.
16. Shorr E. A new technique for staining vaginal smears. III: A single differential stain. Science 1941;94:545–6.
17. Kruger TF, Ackermann SB, Simmons KF, Swanson RJ, Brugo SS, Acosta AA. A quick, reliable staining technique for human sperm morphology. Arch Androl 1987;18:275–7.
18. Menkveld R, Lacquet FA, Kruger TF, Lombard CJ, Sanchez Sarmiento CA, de Villiers A. Effects of different staining and washing procedures on the results of human sperm morphology evaluation by manual and computerized methods. Andrologia 1997;29:1–7.
19. Henkel R, Schreiber G, Sturmhoefel A, Hipler U-C, Zermann DH, Menkveld R. Comparison of three staining methods for the morphological evaluation of human spermatozoa. Fertil Steril 2008;89:449–55.
20. Schirren C, Eckhart U, Jachczik R, Carstensen CA. Morphological differentiation of human spermatozoa with Testsimplets slides. Andrologia 1977;9: 191–2.
21. Ombelet W, Pollet H, Bosmans E, Vereecken A. Results of a questionnaire on sperm morphology assessment. Hum Reprod 1997;12:1015–20.
22. Ragni G, Marzioli S, Levenberg A, Guercilena S. Comparison of the various techniques of identifyng human spermatozoa morfology. Acta Eur Fertil 1984;15:437–40.
23. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. Clin Chem 1979;25:432–8.
24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
25. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
27. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirement. Phys Ther 2005;85:257–68.
28. Menkveld R, Wong WY, Lombard CJ, Wetzels AM, Thomas CMG, Merkus HM, et al. Semen parameters, including WHO and strict criteria morphology, in a fertile and sub fertile population: an effort towards standardization of in-vivo thresholds. Hum Reprod 2001;16:1165–71.
29. Calamera JC, Vilar O. Comparative study of sperm morphology with three different staining procedures. Andrologia 1979;11:255–8.
30. Riddell D, Pacey A, Whittington K. Lack of compliance by UK andrology laboratories with World Health Organization recommendations for sperm morphology assessment. Hum Reprod 2005;20:3441–5.
31. World Health Organization. Laboratory manual for the examination of human semen and sperm-cervical mucus interaction. 3rd ed. New York: Cambridge University Press; 1992.
32. Abraham-Peskir JV, Chantler E, Uggerhoj E, Fedder J. Response of midpiece vesicles on human sperm to osmotic stress. Hum Reprod 2002;17:375–82.
33. Cooper TG, Yeung C-H, Fetic S, Sobhani A, Nieschlag E. Cytoplasmic droplets are normal structures of human sperm but are not well preserved by routine procedures for assessing sperm morphology. Hum Reprod 2004;19: 2283–8.
34. Aitken R, Krausz C, Buckingham D. Relationships between biochemical markers for residual sperm cytoplasm, reactive oxygen species generation, and the presence of leukocytes and precursor germ cells in human sperm suspensions. Mol Reprod Dev 1994;39:268–79.
35. Harasymowycz J, Ball L, Seidel G. Evaluation of bovine spermatozoal morphologic features after staining or fixation. Am J Vet Res 1976;37:1053–7.
36. Katz DF, Overstreet JW, Samuels SJ, Niswander PW, Bloom TD, Lewis EL. Morphometric Analysis of spermatozoa in the assessment of human male fertility. J Androl 1986;7:203–10.
37. Barratt CLR. On the accuracy and clinical value of semen laboratory tests. Hum Reprod 1995;10:247–52.
38. Schoenfeld C, Amelar RD, Dubin L, Amelar S. A new staining technique for the rapid determination of the morphologic characteristics of sperm. Fertil Steril 1981;36:408–10.

SUPPLEMENTAL FIGURE 1



Distribution of (A) normal sperm morphology and (B) head defects, percentages obtained by the Diff-Quik and Testsimplets staining methods.

Natali. Testsimplets for sperm morphology assessment. Fertil Steril 2013.

SUPPLEMENTAL TABLE 1

Contingency table 2 × 2 (left part of the table) used for Cohen kappa statistics (right), reporting the number of subject and the frequency for each category (in parentheses).

	Diff-Quik			Kappa statistic			
	Sperm morphology ≥4% n (frequency)	Sperm morphology <4% n (frequency)	Total n (frequency)	Cohen's kappa	Standard error	95% CI	Level of agreement
Testsimplets				0.22	0.077	0.07–0.37	Fair
Sperm morphology ≥4% n (frequency)	69 (0.66)	0 (0)	69 (0.66)				
Sperm morphology <4% n (frequency)	29 (0.28)	6 (0.06)	35 (0.34)				
Total	98 (0.94)	6 (0.06)	104 (1.0)				

Note: The level of agreement was established according to Landis and Koch (26).

Natali. Testsimplets for sperm morphology assessment. Fertil Steril 2013.