Variability in the morphologic assessment of human sperm: use of the strict criteria recommended by the World Health Organization in 2010

Yongxin Wang, M.Sc.,[a] Jiali Yang, B.Sc.,[a] Yanping Jia, B.Sc.,[a] Chengliang Xiong, M.D.,[b] Tianqing Meng, M.D.,[b] Huangtao Guan, M.D.,[b] Wei Xia, M.D.,[b] Mingyue Ding, Ph.D.,[a,c] and Ming Yuchi, Ph.D.[a]

[a] Key Laboratory of Image Processing and Intelligent Control of the Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, [b] Research Institute of Family Planning, Tongji Medical College, and [c] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei, People's Republic of China

Objective: To determine the variability in the recognition of normal sperm and various sperm defects using the strict criteria recommended by the World Health Organization (5th edition, 2010).
Design: Sperm morphologic assessment by three experienced evaluators.
Setting: Image processing laboratory and reproduction research institute.
Patient(s): Semen donors from a sperm bank.
Intervention(s): The morphology of 5,296 sperm was evaluated using statistical analyses of variability.
Main Outcome Measure(s): The proportion and coefficients of variation (CVs) of normal sperm, defects of specific parts, and the categories of defects were measured. The degree of agreement between any two of the three evaluators was calculated. The multiple anomalies index, teratozoospermia index, sperm deformity index, and the CVs were also measured.
Result(s): The CVs of normal sperm, multiple anomalies index, teratozoospermia index, and sperm deformity index were 4.80%, 4.14%, 5.75%, and 6.81%, respectively. A broader range (4.80%–132.97%) of CVs was observed for the recognition of various defects. The coefficients of the degree of agreement concerning specific morphologic parts of sperm varied (0.387–0.607), with lower relative values for the head and mid-piece than for the tail and cytoplasm.
Conclusion(s): The sperm head is more difficult to evaluate than the other parts using the criteria recommended by the World Health Organization in 2010. The degree of agreement concerning specific parts and various defects varied in broad ranges. A stricter definition for each defect is needed. (Fertil Steril® 2014;101:945–9. ©2014 by American Society for Reproductive Medicine.)
Key Words: Sperm morphology assessment, evaluation variability, agreement analysis

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/wangy-evaluation-variability-sperm-morphology-who/

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.

Sperm morphology is positively associated with and able to predict fertilization (1). Morphology provides the most independent and stable semen assessment parameters (2). However, sperm morphology assessment can be influenced by many factors, including differences in semen samples, smearing preparation techniques, interpretation, classification systems, and technician experience (2). Eustache and Auger (3) studied the assessment variability by evaluators in 2003. Microscopic images of 100 sperm were assessed by 62 evaluators who were divided into two subgroups according to experience level. David's morphology classification method was used. The results showed a high variability (range 6%–39%, coefficient of variation [CV]

Received August 18, 2013; revised December 5, 2013; accepted December 26, 2013; published online February 10, 2014.
Y.W. has nothing to disclose. J.Y. has nothing to disclose. Y.J. has nothing to disclose. C.X. has nothing to disclose. T.M. has nothing to disclose. H.G. has nothing to disclose. W.X. has nothing to disclose. M.D. has nothing to disclose. M.Y. has nothing to disclose.
Supported by the Hubei Province Natural Science Foundation (grant 2008BCC007).
Reprint requests: Ming Yuchi, Ph.D., Huazhong University of Science and Technology, School of Life Science and Technology, Department of Biomedical Engineering, 1037 Luoyu St., Wuhan, Hubei Province, 430074, People's Republic of China (E-mail: m.yuchi@hust.edu.cn).

Fertility and Sterility® Vol. 101, No. 4, April 2014 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2013.12.047

40%) in the assessment of normal sperm and a lower variability for the multiple anomalies index (MAI) (CV 12%), percentages (CV 23%), absent tails (CV 25%), and abnormal or absent acrosomes (CV 26%). The authors concluded that the role that experience played and the importance of following the recommended methodologies for relevant and comparable results were important variables for providing an accurate assessment.

In 2010 the World Health Organization (WHO) published the latest (5th edition) laboratory manual for the examination and processing of human semen (4), which recommended a lower reference limit for normal morphology (4%) than the 4th edition (14%) (5). Defects recommended as criteria by the 5th edition manual are also different from the 4th edition manual used by Eustache and Auger. Can the new criteria be accurately grasped and applied strictly by evaluators? How does the corresponding variability in the morphologic assessment change with the new criteria? With the application of strict criteria, can the difficulties in sperm morphologic assessment, especially for various defects, be improved upon? The main objective of the present study was to explore these questions by investigating the statistical variability of sperm assessment by studying a large group of sperm.

MATERIALS AND METHODS

The semen samples were provided by the Human Sperm Bank in Wuhan, Hubei Province, China. Each volunteer signed an informed consent form, and this experiment was approved by the institutional review board. The semen samples were obtained according to the standard protocol for human sperm banking initiated by the Chinese Ministry of Health. The sperm are from anonymous donors between the ages of 22 and 44 years. The donors must be in good health, both physically and psychologically, and have no history of genetic disease (6). The semen samples were processed using the modified Papanicolaou staining method recommended by the 5th WHO manual. Slide images (average 10.71 sperm per image) were captured using a digital camera with a resolution of 1,280 × 960 pixels.

Three experts (A, B, and C) with more than 10 years of experience from three different hospitals assessed 5,296 individual sperm. All of the three hospitals are among the top clinical research organizations for sperm morphology assessment in China. During the assessment, the criteria recommended by the 5th WHO manual were strictly followed.

Software was developed for the sperm morphology assessment. The software labeled each sperm with a unique number. The software provides a user-friendly interface so that the evaluators can assess each sperm and mark all defects, including the following: tapered, pyriform, round, amorphous, vacuolated, small acrosome area on head; bent neck, asymmetrical, thick insertion, thin in mid-piece; short, bent, coiled in tail; and excess residual cytoplasm (ERC), as recommended by the 5th manual from the WHO. In the event that a sperm is considered to be abnormal but cannot be categorized by a defect included in the 5th WHO manual, the evaluator can mark the sperm as dubious. The recorded assessment result can be directly used for the analyses of agreement and variability.

Statistical analyses were performed on the basis of the parameters suggested by the 5th WHO manual, including the proportion of normal sperm (PNS), the teratozoospermia index (TZI), MAI, the sperm deformity index (SDI), and the rates of a deformed head, mid-piece, tail, and cytoplasm from individual sperm. To measure the variability generated by evaluator subjectivity, the means, CVs, and degrees of agreement for defects were calculated as well.

The number of sperm with a defect observed in the head was divided by 5,296 to calculate the rate of defects in the head. The rate of defects in the mid-piece, the rate of defects in the tail, the rate of defects in the cytoplasm, and the rate of various other defects were all calculated by the same criteria. The TZI, MAI, and SDI were obtained following the method reported by Auger (7). The corresponding mean, SD, and CV of each parameter were also calculated. The results are shown in Table 1. The CVs reported by Eustache and Auger in 2003 (3) are also listed in the rightmost column of Table 1.

To analyze evaluator variability, we needed to identify whether evaluators provide the same assessment for duplicate sperm. Normally slides were evaluated by one or more evaluators with a fixed amount of sperm (usually 200 or 400) (8). Because the sperm assessed may not be consistent, it is difficult to ensure that all evaluators are assessing identical sperm. In this study, three evaluators assessed a fixed amount of 5,296 sperm independently, allowing us to measure the evaluators' views on each specific sperm and to analyze the degree of agreement for specific defects.

The degree of agreement between two evaluators (e.g., evaluators A and B) can be obtained by calculating the kappa (κ) coefficient with the following formula (9): $\kappa = \frac{P_o - P_e}{1 - P_e}$, in which κ is the coefficient of agreement (0–1), P_o is the proportion of units in which the judges agreed, and P_e is the proportion of units in which agreement is expected by chance.

A high κ value corresponds to a high degree of agreement between the evaluators. Normally, a six-level classification mode is used for evaluating the κ coefficient: almost perfect (κ: 0.81–1.00), substantial (κ: 0.61–0.80), moderate (κ: 0.41–0.60), fair (κ: 0.21–0.40), slight (κ: 0.00–0.20), or poor ($\kappa < 0.00$).

Specific parts of sperm in this study include the following: head, mid-piece, tail, and cytoplasm. The results of the agreement on overall aspects and specific parts of the sperm are shown in Table 2.

Specific defects in this study include the following: head aspect: tapered, round, pyriform, amorphous, small acrosomal area, vacuolated, dubious for head; mid-piece: asymmetrical, bent beck, thick insertion; tail: short, coiled, bent, dubious for tail; cytoplasm: ERC. The results of agreement on specific defects are shown in Table 3.

RESULTS

All of the PNS values were much higher than the lower limit recommended by the WHO in the 5th manual (4%). The PNS mean was 20.87%. The CV of the proportion of normal sperm was 4.80%, which is lower than the values reported by other studies (2, 3). Considering that the semen samples were from the donors of the Human Sperm Bank, not from patients with infertility, the result of the proportion of normal sperm is reasonable and credible.