UCSF
UC San Francisco Previously Published Works

Title

Variability in testis biopsy interpretation: implications for male infertility care in the era of intracytoplasmic sperm injection

Permalink

Journal

ISSN

Authors

Cooperberg, Matthew R
Chi, Thomas
Jad, Amir
et al.

Publication Date

DOI

Variability in testis biopsy interpretation: implications for male infertility care in the era of intracytoplasmic sperm injection

Matthew R. Cooperberg, M.D.,[a] *Thomas Chi, B.A.,*[a] *Amir Jad, M.D.,*[a] *Imok Cha, M.D.,*[b] *and Paul J. Turek, M.D.*[a]

[a] Department of Urology and [b] Department of Pathology, University of California San Francisco, San Francisco, California

Objective: To determine whether center to center discrepancies in the ability to locate sperm in infertile testes with abnormal histology stems in part from inconsistencies in pathologists' readings of testis biopsies.

Design: Prospective cohort study.

Setting: Academic male infertility practice.

Patient(s): Consecutive series of azoospermic men referred with testis biopsy slides between 1998 and 2003.

Intervention(s): Testis biopsy histologies on azoospermic patients referred for infertility care were re-reviewed by a single pathologist blinded to the original reading. Subsequent infertility care was guided by the findings from the second histologic reading.

Main Outcome Measure(s): Agreement between the outside and in-house review of testis biopsy readings was assessed with the kappa statistic. Pregnancy outcomes that resulted from clinical decisions informed by the second histologic readings were also assessed.

Result(s): Among 113 histologic specimens, re-review was complicated by fixation artifacts in 18 cases (16%) and insufficient biopsy sample size in 13 cases (12%). The kappa score for interobserver agreement in readings was 0.43 (95% CI 0.32–0.054). Mixed histology patterns in particular were underappreciated by outside pathologists (13% of cases on original reading, 36% of cases on review). In 27% of all cases, the differences in biopsy readings had a significant impact on clinical management.

Conclusion(s): A correlation between independent testis histology readings in azoospermic men demonstrates frequent inconsistencies. These differences contribute to inaccurate phenotyping of male infertility and can significantly impact the direction of infertility care. These findings highlight the need for a standardized approach to testis histologic review. (Fertil Steril® 2005;84:672–7. ©2005 by American Society for Reproductive Medicine.)

Key Words: Testis biopsy, azoospermia, male infertility, spermatogenesis

In infertility due to azoospermia, paternity can be achieved with spermatozoa retrieved from the testis in combination with in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI) (1). In men with excurrent duct obstruction, sperm retrieval is not difficult; however, among men with severe testis failure or nonobstructive azoospermia, sperm for ICSI are not successfully retrieved in 25% to 50% of cases (2). In men with nonobstructive azoospermia, clinical features including testicular size, history of ejaculated sperm, serum follicle stimulating hormone level, or testis biopsy histology do not accurately predict whether sperm will be recovered during testicular exploration (2). However, among these variables, successful testicular sperm retrieval is most closely related to the type or pattern of spermatogenic defect observed on testicular biopsy (3).

A close examination of the literature suggests that there is a wide variation in clinicians' ability to locate sperm in testes that share commonly recognized histologic patterns. For example, in cases of Sertoli cell-only (germ cell aplasia) pathology, successful sperm retrieval occurs in 5% to 60% of cases (3–6). Although this may reflect differences in surgical technique among individuals and centers that specialize in male infertility care, we hypothesize this variation may also stem from inconsistent interpretation of biopsy histology by pathologists. To address the issue of interobserver variability in testis biopsy interpretation, we prospectively compared the findings from two independent testis histology readings performed on a cohort of patients with azoospermia. Our goal was to assess how uniformly pathologists read testis biopsy histology and delineate biopsy patterns that exhibit the widest variability in interpretation.

MATERIALS AND METHODS

As part of the evaluation of azoospermic patients referred to the Male Reproductive Health Center at the University of California at San Francisco, testis biopsy slides prepared at the referring institution are routinely procured and re-reviewed by a

single pathologist (I.C.). Histology re-reviews were performed prospectively, and were blinded to the original pathologic diagnosis. For the study, a third independent review summarized the written findings from the two readings and assessed each of the following: [1] specimen adequacy (>25 seminiferous tubule cross-sections) (7), [2] presence or absence of fixation artifact (vacuole formation due to air-drying or fixation in formalin, or specimen overstaining or understaining), and [3] biopsy pattern.

As described by Levin (8), biopsy patterns were categorized as normal spermatogenesis, hypoplasia or hypospermatogenesis (HYP), complete or early maturation arrest (EMA), Sertoli cell only (SCO), incomplete or late maturation arrest (LMA), or other (including sclerosis). When referral reports did not use this terminology, they were assigned to the category best fitting the recorded narrative description. If a biopsy contained a single histologic pattern throughout the specimen, it was deemed a pure pattern. If biopsies exhibited two or more patterns, then a mixed pattern was assigned. Mixed readings were further characterized in terms of the primary, or predominant, pattern and the secondary pattern.

The frequency of each histologic pattern on paired readings was determined, and the concordance rate was calculated for each pattern. Overall agreement between original and second readings was assessed using the simple kappa statistic for interobserver agreement. Additionally, in each case, a determination was made whether the difference in histologic diagnosis between the two readings was clinically significant, defined as a difference that would confer a change in clinical management of the patient.

The criteria for a clinically significant change include the following: [1] a change from normal to any abnormal pattern (i.e., change in presumed etiology from obstructive to nonobstructive azoospermia); [2] a change from any abnormal pattern to a normal pattern (i.e., change in presumed etiology from nonobstructive to obstructive azoospermia); [3] any pure pattern diagnosis without mature sperm (i.e., EMA, SCO) changed to a pure or mixed pattern that suggests presence of mature sperm (i.e., HYP or normal); and [4] any mixed pattern diagnosis without mature sperm (i.e., EMA, SCO) changed to a pure or mixed pattern that suggests mature sperm is present (i.e., HYP or normal).

Finally, fertility pathways and outcomes for patients who met the criteria for a clinically significant change in management were examined to better understand the ultimate impact of these clinical decisions.

RESULTS

Over a 5-year period, 132 testis biopsies were evaluated. Of these, a formal histologic diagnosis from a referring institution was available for 113 (86%) cases. The median interval between the date of the biopsy procedure and the date of re-review was 5.5 months. Fixation artifact was noted in 18 cases (16%), making re-review difficult, and in 13 cases (12%) testis biopsy sample size was considered suboptimal for interpretation.

Concordance Between Readings

The concordance between histologic patterns assigned in the two readings—original and review—is outlined in Table 1. Several observations are evident from this comparison. The general frequency distribution of histologic patterns is similar between biopsy readings (except for mixed pattern readings). There is a 23% reduction in the diagnosis of normal spermat-

TABLE 1

Concordance in testis histologic patterns between two readings.

	Review diagnosis								
	Normal	Hypo	EMA	LMA	SCO	Mixed	Other	Total	Agree (%)
Original diagnosis									
Normal	19	5	0	0	0	15	0	39	49
Hypo	9	11	0	1	0	4	0	25	44
EMA	0	1	5	1	0	6	0	13	38
LMA	0	1	1	1	0	2	0	5	20
SCO	0	0	0	0	10	1	0	11	91
Mixed	1	1	0	0	1	12	0	15	80
Other	1	0	0	0	0	1	3	5	60
Total	30	19	6	3	11	41	3	113	
Agree (%)	63	58	83	33	91	29	100		54

Note: Hypo = hypoplasia or hypospermatogenesis; EMA = early maturation arrest; LMA = late maturation arrest; SCO = Sertoli cell only.

Cooperberg. Variability in testis biopsy analysis for infertility. Fertil Steril 2005.

Examples of mixed testicular histology that resulted in changes in clinical care. (**A**) An area of normal spermatogenesis (*arrow*) is identified in a large biopsy sample originally read as pure sclerosis. (**B**) An area of late maturation arrest is identified (*arrow*) in a biopsy sample originally read as pure early maturation arrest.
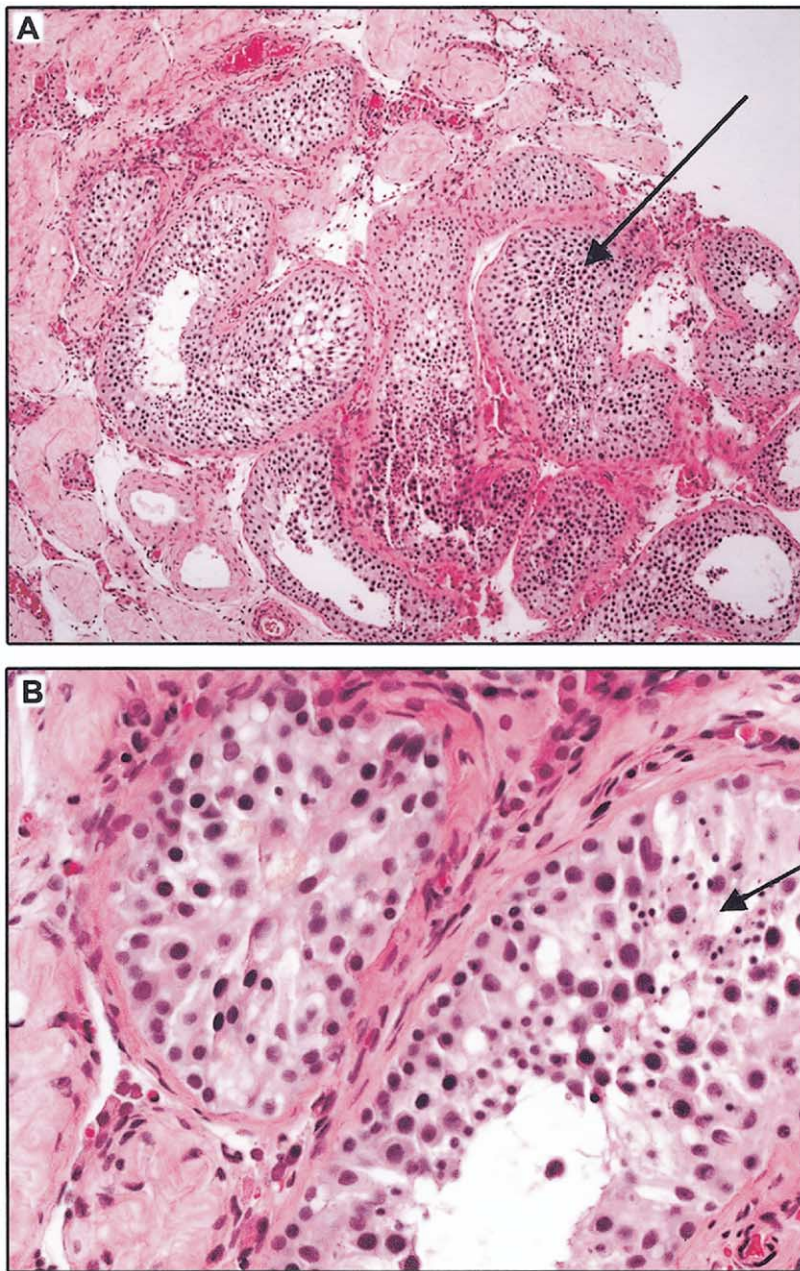
Cooperberg. Variability in testis biopsy analysis for infertility. Fertil Steril 2005.

ogenesis when original (n = 39, 35%) and review (n = 30, 27%) readings are compared. The best concordance is observed with the SCO pattern, identified in 11 cases on both original and review examination, and in agreement in 91% (10/11) of cases. The overall simple kappa score for agreement between the original and review diagnoses is 0.43 (95% CI 0.32–0.54).

Discordance Between Readings

There are also interesting observations in the patterns of discordance among biopsy-sample readings. Among biopsy samples originally read as normal (n = 39), almost half (49%, n = 19) were confirmed as normal on re-review, whereas 12% (n = 5) demonstrated HYP and 39% (n = 15)

mixed histology. Clinically, this represents a significant shift in diagnosis, as many men considered normal (i.e., "obstructed") on the original review were not obstructed on re-review. However, the greatest source of discordance between the biopsy readings was the underappreciation of mixed histology by the original pathologist: 15 cases (13%) were felt to demonstrate mixed histology on the original reading whereas 41 cases (36%) were interpreted as mixed pattern on re-review, representing an almost threefold difference in detection rate. In addition, although most biopsies originally read as mixed pattern were re-reviewed as mixed pattern (80%), three others (20%) had pure pattern interpretations that included normal.

Readings With Mixed Histologic Patterns

Of the 41 cases read as mixed on review, 10 (24%) had a normal primary pattern, and an additional 10 (24%) had a normal secondary pattern. The primary pattern was HYP in 11 cases (27%), EMA in 7 (17%), LMA in 3 (7%), and SCO in 10 (24%). Of the 29 cases read as mixed on review but not on original reading, in 15 cases (52%) the review diagnosis correlated with the primary pattern of the mixed histology. Primary and secondary patterns in the mixed cases existed in virtually every potential combination, with the exception that EMA never coexisted with normal histology in the same testis. In 3 cases, three or four discrete histologic patterns were identified on review; only 1 of these cases was originally read as mixed.

Significance of Differences in Readings

Concerning the significance of the differences in biopsy readings, in 27% of the cases the differences in histology readings had significant impact on clinical management. In 14 cases (12%), an original reading of normal was reread as abnormal, suggesting nonobstructive azoospermia. In 11 cases (10%) an initial reading of abnormal was re-read as normal, suggesting obstructive azoospermia amenable to reconstructive surgery. In 5 cases (5%), a biopsy read as pure EMA, SCO, or other was reread as containing at least some normal, hypospermatogenic, or LMA spermatogenic elements, thus raising the potential for sperm retrieval. Figure 1 illustrates a case for which reinterpretation significantly changed the treatment plan.

Fertility Follow-up Among Patients With Significantly Different Readings

To determine whether altered clinical decisions among these 30 patients were accurate and actually resulted in improved clinical outcomes, information regarding subsequent fertility procedures was available for 26 couples (86%), and birth outcomes for 23 (77%). The mean partner age in this cohort was 33 years, and the overall birth rate was 63% in couples with known pregnancy outcomes. Among 10 patients with a change in diagnosis from nonobstructive to obstructive etiology, 3 underwent bilateral vasoepididymostomy, and all 3

achieved successful singleton pregnancies. Six patients elected microsurgical epididymal sperm aspiration (MESA) with IVF, after which 3 had successful pregnancies, 2 were unsuccessful, and 1 had an unknown outcome. One patient, whose spouse was 41 years of age, opted for testicular sperm aspiration (TESA) and did not achieve pregnancy.

Of the 14 patients whose diagnoses changed from obstructive to nonobstructive azoospermia, one underwent vasoepididymostomy with a diagnosis of hypospermatogenesis and did not achieve a pregnancy. Two underwent varicocelectomy, resulting in one pregnancy and one adoption. Two opted for MESA, with one successfully conceiving after IVF. The remainder (n = 9) elected testis mapping with fine-needle aspiration followed by TESA or testicular sperm extraction (TESE). Among 6 patients with known birth outcomes, 4 had successful pregnancies. Finally, 2 patients harbored mixed patterns on both original and review readings, but mature sperm were found only on re-review. Both patients proceeded to sperm retrieval with IVF/ICSI, and both had successful pregnancies (Table 2).

DISCUSSION

Although no clinical parameter correlates perfectly with the potential for sperm retrieval for ICSI, testis histology serves as one of the better surrogate markers. Testis histology as represented by a single biopsy site certainly does not accurately reflect overall testis biology, as significant spermatogenic variation has been shown to exist in infertile testes (9). Despite this limitation, an accurate assessment of testicular histology has been important for planning infertility treatment; the suggestion of even small areas of complete spermatogenesis makes successful sperm retrieval very likely.

Despite the routine role of the testis biopsy in the diagnosis of azoospermia, the relationship between sperm retrieval success and histologic pattern on diagnostic biopsy varies widely from center to center. We hypothesized that some of this variation may be because of inconsistencies in histologic interpretation by pathologists. Indeed, this study confirms that there is wide interobserver discrepancy in the interpretation of samples from routine testis biopsies performed for azoospermia.

One reason for the large interobserver discrepancy observed in testis biopsy interpretation could stem from the review of poorly prepared or insufficiently sized biopsy samples. We noted that over 25% of specimens showed such features, constituting a significant variable in reporting quality.

Another explanation for interobserver discrepancy is that no particular method of testis biopsy evaluation is uniformly applied and serves as standard procedure for pathologists. Several testis biopsy interpretation algorithms have been reported in the last 3 decades. In general, they show large variations in the quantitative and qualitative assessment of spermatogenesis (7, 8, 10–12). The three most commonly

TABLE 2

Changes in diagnosis affecting clinical management.

Original diagnosis	Review diagnosis	Procedures	Fertility outcome	Partner age (y)
Change in diagnosis from nonobstructive to obstructive				
Mixed (LMA + EMA)	Normal	MESA	Unknown	42
Hypo	Normal	VE, MESA	Singleton-natural	32
Hypo	Normal	VE, MESA	Singleton	31
Hypo	Normal	TESA	No pregnancy	41
Hypo	Normal	MESA	Twins	29
Hypo	Normal	MESA	No pregnancy	39
Hypo	Normal	MESA	No pregnancy	29
Hypo	Normal	MESA	Singleton	29
Hypo	Normal	MESA	Singleton	27
Hypo	Normal	VE, MESA	Singleton-natural	Unknown
Change in diagnosis from obstructive to nonobstructive				
Normal	Mixed (Normal + Hypo)	Varicocelectomy, IVF	Singleton	35
Normal	Mixed (Normal + Hypo)	TESE	Unknown	32
Normal	Mixed (Normal + LMA)	MESA	Twins	28
Normal	Mixed (Normal + LMA)	TESE	No pregnancy	31
Normal	Mixed (Hypo + Normal)	TESE	Singleton	36
Normal	Mixed (Hypo + Normal)	VE	No pregnancy	Unknown
Normal	Mixed (Hypo + Normal)	TESE	Singleton	28
Normal	Mixed (Hypo + LMA)	TESE	Twins	26
Normal	Mixed (Hypo + LMA)	TESA	Unknown	30
Normal	Mixed (Hypo + LMA)	TESA	Singleton	35
Normal	Mixed (LMA + Normal)	TESE	Singleton	35
Normal	Hypo	TESA	No pregnancy	30
Normal	Hypo	MESA	No pregnancy	37
Normal	Hypo	Varicocelectomy	No pregnancy	Unknown
No change in assumed etiology, but mature elements found on review only				
EMA	Hypo	TESE	Singleton	40
EMA	LMA	Varicocelectomy	Twins	30

Note: Hypo = hypospermatogenesis; LMA = late maturation arrest; EMA = early maturation arrest; SCO = Sertoli cell only; VE = vasoepididymostomy; MESA = microsurgical epididymal sperm aspiration; TESA = testicular sperm aspiration; TESE = testicular sperm extraction.

Cooperberg. Variability in testis biopsy analysis for infertility. Fertil Steril 2005.

cited scoring systems are those by Johnsen (11), Levin (8), and Silber and Rodriquez-Rigau (12). The Johnsen system is based on the idea that testis damage causes the successive disappearance of the most mature cell type (11); therefore, the scoring system involves a detailed, quantitative assessment of germ cell types and is considered too laborious for routine clinical use by most practicing pathologists. Based on an assessment of 21 patients, Silber and Rodriquez-Rigau (12) proposed a simpler quantitative scoring system that counts only mature spermatids. Finally, the scoring system proposed by Levin (8) is the most qualitative scoring system, involving the recognition of certain spermatogenic patterns that occur quite typically in infertility cases. No method is considered the standard, despite a great need for a uniform approach to testis biopsy sample interpretation.

The issue of interobserver variability in testis biopsy interpretation is of increased importance in the era of ICSI. Indeed, we observed that clinical management was altered in 27% of patients when biopsy histology was re-reviewed by a single in-house pathologist. Such changes in clinical care treatment decisions vary from whether certain genetic risks are discussed with the patient, to the level of difficulty required for testis sperm retrieval or the ability to provide surgical reconstruction for obstructive azoospermia in cases previously deemed nonobstructive in nature. The validity of this analysis is supported by the comprehensive assessment of pregnancy outcomes in couples in whom fertility care was altered by histologic re-review, which demonstrated excellent overall success rates.

Study limitations include the fact that that biopsy re-review was subject to similar, though albeit better controlled, restrictions defined earlier for any biopsy scoring system used in the analysis. However, the spirit of the study is to encourage clinicians to develop and partake in a uniformly acceptable method of assessing fertility potential in male infertility.

The lack of a standard approach to testis biopsy sample histologic examination by pathologists may contribute to inaccurate phenotyping of male infertility. Such inaccuracy may significantly impact on the clinical care of patients. One area of variability in testis biopsy readings is the accurate identification of mixed spermatogenic patterns, as even a small number of tubules with complete spermatogenesis may yield viable spermatozoa for ICSI. Another concern is that the definition of normal spermatogenesis varies widely depending on the scoring system, a fact that can significantly alter the care of infertile men. We recommend that the clinical care of male infertility involve an experienced pathologist, working in close association with infertility clinicians.

REFERENCES

1. Schlegel PN, Palermo GD, Goldstein M, Menendez S, Zaninovic N, Veeck LL, et al. Testicular sperm extraction with intracytoplasmic sperm injection for nonobstructive azoospermia. Urology 1997;49:435–40.
2. Tournaye H, Verheyen G, Nagy P, Ubaldi F, Goossens A, Silber S, et al. Are there any predictive factors for successful testicular sperm recovery in azoospermic patients? Hum Reprod 1997;12:80–6.
3. Su LM, Palermo GD, Goldstein M, Veeck LL, Rosenwaks Z, Schlegel PN. Testicular sperm extraction with intracytoplasmic sperm injection for nonobstructive azoospermia: testicular histology can predict success of sperm retrieval. J Urol 1999;161:112–6.
4. Tournaye H, Liu J, Nagy PZ, Camus M, Goossens A, Silber S, et al. Correlation between testicular histology and outcome after intracytoplasmic sperm injection using testicular spermatozoa. Hum Reprod 1996;11:127–32.
5. Seo JT, Ko WJ. Predictive factors of successful testicular sperm recovery in non-obstructive azoospermia patients. Int J Androl 2001;24:306–10.
6. Meng MV, Cha I, Ljung BM, Turek PJ. Relationship between classic histological pattern and sperm findings on fine needle aspiration map in infertile men. Hum Reprod 2000;15:1973–7.
7. Steinberger E, Tijoe D. A method for quantitative analysis of human seminiferous epithelium. Fertil Steril 1968;19:959–61.
8. Levin HS. Testicular biopsy in the study of male infertility: its current usefulness, histologic techniques, and prospects for the future. Hum Pathol 1979;10:569–84.
9. Turek PJ, Ljung BM, Cha I, Conaghan J. Diagnostic findings from testis fine needle aspiration mapping in obstructed and nonobstructed azoospermic men. J Urol 2000;163:1709–16.
10. Meinhard E, McRae CU, Chisholm GD. Testicular biopsy in evaluation of male infertility. BMJ 1973;3:577–81.
11. Johnsen SG. Testicular biopsy score count—a method for registration of spermatogenesis in human testes: normal values and results in 335 hypogonadal males. Hormones 1970;1:2–25.
12. Silber SJ, Rodriguez-Rigau LJ. Quantitative analysis of testicle biopsy: determination of partial obstruction and prediction of sperm count after surgery for obstruction. Fertil Steril 1981;36:480–5.