Asian Journal of Andrology (2015) 17, 281–284

www.asiaandro.com; www.ajandrology.com

Open Access

ORIGINAL ARTICLE

Male Fertility

Follicle-stimulating hormone as a predictor for sperm retrieval rate in patients with nonobstructive azoospermia: a systematic review and meta-analysis

Qi Yang, Yan-Ping Huang, Hong-Xiang Wang, Kai Hu, Yi-Xin Wang, Yi-Ran Huang, Bin Chen

Noninvasive parameters for predicating sperm retrieval rate (SRR) are desirables. Follicle-stimulating hormone (FSH) has been an important predictor since the first years of testicular sperm extraction. Recent studies showed continuous interests in FSH, with both pros and cons. Thus, we conducted a meta-analysis to evaluate the diagnostic value of FSH as a predictor for patients with nonobstructive azoospermia (NOA) taking testicular sperm retrieval. Eligible diagnosis tests were identified from electronic databases (Cochrane Central Register of Controlled Trials, Medline, and EMBASE) without language restrictions. The database search, quality assessment, and data extraction were performed independently by two reviewers. The reference standard was the sperm retrieval result. Diagnostic value of FSH were explored by area under receiver operation characteristics (ROC) curve using Review Manager, version 5.1.0 (Cochrane Collaboration, Oxford, UK) and Meta-DiSc, version 1.4. Meta regression will be done if there is heterogeneity. Then, we find 11 tests including a total of 1350 patients met the inclusion criteria. Our pooled analysis showed that the area under ROC curve of FSH was 0.72 ± 0.04. Meta regression analyses showed that region and average age have an influence on the diagnostic value. FSH showed more diagnostic value with patients in East Asia and with younger patients. We concluded that FSH had moderate value in independently predicating SRR in men with NOA (area under curve >0.7). More detailed diagnosis tests should be anticipated in the future to confirm the diagnostic value of other noninvasive parameters.
Asian Journal of Andrology (2015) **17**, 281–284; doi: 10.4103/1008-682X.139259; published online: 23 September 2014

Keywords: follicule-stimulating hormone; meta-analysis; nonobstructive azoospermia; testicular sperm retrieval

INTRODUCTION

Azoospermia occurs in 1% of men and 10%–12% of the infertile male population. Nonobstructive azoospermia (NOA), which is caused by testicular failure, represents 60% of all cases of azoospermia.[1,2] Since the first successful surgical sperm retrieval in combination with intracytoplasmic sperm injection (ICSI) in 1994, the use of surgically retrieved sperm from the testis for ICSI has made it possible for patients with NOA to father children.[3,4]

However, the recovery of spermatozoa is successful in only 50% of cases and therefore it would be beneficial to predict the success of sperm retrieval using noninvasive parameters before attempted treatment.[5,6] This would not only decrease the surgical risk and the inconvenience to the patient, but also lower the costs of the infertility workup. Although no single clinical finding or investigation able to accurately predict has been found, follicle-stimulating hormone (FSH) has been an important preoperative serum parameter studied since the first years of testicular sperm extraction (TESE).[7] In general, the serum FSH concentration is inversely related to sperm retrieval rate (SRR).[8,9]

Recent studies showed continuous interests in the value of FSH in prediction, with both pros and cons.[10–25] Therefore, it is necessary to conduct a systematic review and meta-analysis to assess the diagnostic value of FSH as a predictor for SRR in patients with NOA before testicular sperm retrieval.

MATERIALS AND METHODS

Systematic search strategy

We searched the following databases: Cochrane Central Register of Controlled Trials, PubMed (from 1994 to June 2013), and EMBASE (from 1994 to June 2013). The following search terms were used to identify any relevant studies: "FSH" and "sperm retrieval, or TESE, or microdissection TESE (MESE)." In addition, identified reports, reviews of the included studies, and other relevant publications from the American Urological Association, European Association of Urology, and Societe Internationale d'Urologie between 2007 and 2013 were manually searched. Conference abstracts were excluded because of the limited data presented in them.

Identification of articles

Diagnosis tests were included only if they met the criteria of testing the diagnostic value of FSH as a predictor for SRR in patients with NOA before TESE/MESE, with general demographic data like patients' age (average age), excluding the presence of limiting to any particular cause of NOA, such as AZFa deletion, or of usage of any other sperm retrieval technique like sperm aspiration that was

Department of Urology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai Institute of Andrology, Shanghai 200127, China.
Correspondence: Dr. B Chen (dr_binchen@126.com)
Received: 09 March 2014; Revised: 27 April 2014; Accepted: 13 June 2014

obviously less successful. Tests without definitive four-fold table were also excluded.

Quality assessment of included studies

The titles and abstracts of all articles were reviewed by two reviewers according to the inclusion criteria using a standardized form. If inconsistencies existed between the reviewers' data, a third reviewer evaluated the data. Quality assessment was performed using methods adapted from two guidelines on systematic reviews of diagnostic studies.[26,27]

For each study, the following quality criteria were scored as fulfilled or not: (1) independent comparison of FSH level against TESE/MESE results; (2) blinded (single or double) interpretation of test and reference standard results; (3) unsolved data preformed. If no data on the above criteria were reported in the primary studies, we requested the information from the authors. For the purposes of analysis, responses coded as "not reported" were grouped together with "not met." A high-quality study was arbitrarily defined as that which met all three criteria; a medium quality met two of the three criteria; and low quality study met <2/3 criteria.

Outcome

Our primary outcome was the summary receiver operation characteristics (SROC) and the area under ROC curve (AUC) of FSH's diagnostic value as a predictor for SRR in patients with NOA before TESE/MESE, while TESE/MESE result was the reference standard, followed by sensitivity, specificity and diagnosis odds ratio (DOR).

Data synthesis and analysis

All analyses were performed using the Review Manager, version 5.1.0 (Cochrane Collaboration, Oxford, UK) and Meta-DiSc, version 1.4 (Clinical Biostatistics Unit, Ramony Cajal Hospital, Madrid, Spain). $P < 0.05$ was considered to be statistically significant. Four-fold tables of each test were fulfilled with numbers of true positive, true negative, false positive, and false negative.

The categorical data were presented as specificity and sensitivity, both with a 95% confidence interval (CI). Continuous outcomes were presented as SROC, and qualitatively described as AUC. The chi-square test and I^2 statistic were used to analyze the heterogeneity in the results.[28] Meta regression and stratified analyses on year of publication, region, patients' average age and sample size will be performed to identify the source of heterogeneity if necessary.

RESULTS

Study characteristic

The combined search strategies identified 11 diagnosis tests,[10–21] including 1350 patients that met the inclusion criteria. Ten of the studies were reported in English, and one was in Chinese. The characteristics and the quality score of the quality assessment of the 11 studies are presented in **Table 1**. All trials were deemed middle or high quality.

Diagnostic accuracy of follicle-stimulating hormone

Figure 1 displays the sensitivity, specificity and DOR estimates from each of the 11 studies. Both sensitivity and specificity estimates were highly variable. Summary measures were grossly heterogeneous ($P < 0.05$) and therefore would not be appropriately summarized. The SROC curve displays an ROC-type trade-off between sensitivity and specificity. The AUC (**Figure 2**) of the 11 studies was 0.72 ± 0.04, with a sensitivity of 0.70 (0.66–0.73) and a specificity of 0.62 (0.58–0.66).

Table 1: Description of included studies

	Year	*Region*	*Study quality*	*TP*	*FP*	*TN*	*FN*	*Sum*	*Average age*
Ballescá *et al.*	2000	Spain	2	9	4	1	3	17	32.3
Amer *et al.*	2001	Eygpt	2	18	31	11	40	100	36.4
Vernaeve *et al.*	2002	Belgium	2	68	52	24	41	185	35.6
Nagata *et al.*	2005	Japan	3	10	7	7	38	62	35
Fei *et al.*	2006	China	2	8	7	1	12	28	29.6
Tunc *et al.*	2006	Turkey	2	28	17	3	4	52	34.5
Mostafa *et al.*	2007	Eygpt	2	15	10	6	9	40	35.5
Ma *et al.*	2011	China	2	62	18	26	40	146	31.8
Boitrelle *et al.*	2011	France	2	89	57	60	74	280	33.2
Ghalayini *et al.*	2011	Jordan	2	53	10	41	31	135	35.1
Huang *et al.*	2012	China	2	114	43	23	125	305	29
Total				474	256	203	417	1350	

TP: true positive; TN: true negative; FP: false positive; FN: false negative

Heterogeneity analysis

We performed meta regression and stratified analyses to identify sources of heterogeneity among these studies. **Table 2** presents two factors that appeared most strongly associated with the observed heterogeneity. Studies in region 1 produced DOR estimates nearly 4 times higher than studies in other regions, and the former showed an AUC >0.7. Studies with patients' average age under 33 produced DOR estimates nearly 4 times higher than studies with patients' average age above 33, and the former showed an AUC >0.7.

DISCUSSION

This is, to the best of our knowledge, the first systematic review with a meta-analysis of the diagnostic value of noninvasive parameters for SRR in patients with NOA before TESE/MESE.

Our pooled analysis for FSH in predicating SRR in patients with NOA showed that the AUC of FSH's diagnostic value was 0.72 ± 0.04. As far as is known, AUC < 0.7, 0.7–0.9, and > 0.9 mean little, moderate and high diagnosis value, respectively.[28,29] This meta-analysis indicated that FSH had a dubitable moderate diagnostic value in predicating SRR.

High sensitivity means low specificity, and DOR makes a balance of both. Heterogeneity of DOR showed statistical significance ($P < 0.01$, $I^2 = 71.3\%$). Then, meta regression and stratified analyses showed that region and average patients' age were two factors that appeared most strongly associated with the observed heterogeneity. In East Asia or with younger patients, FSH showed a more clear diagnostic value.

Region, interestingly, had an influence on the diagnostic value of FSH according to this meta-analysis, indicating that other factors affecting spermatogenic function might have less effects in East Asia. One factor draw our attention was serum and seminal leptin level. People in region 1 have lower body mass index than other regions,[30,31] and leptin, which impacts spermatogenic,[17] is associated with this.[32] Thus patients in region 2 or 3 might have a leptin level around the threshold, interfering FSH's diagnostic value.

Aging is a clear factor that impact spermatogenic function, and meanwhile increase FSH.[33] Our results suggested that age had a greater influence on the former. In fact, the increase of FSH is a side-effect of decrease of androgen with aging, and deficiency of androgen is also an etiology of dyszoospermia.

However, FSH alone is still quite not enough (AUC < 0.9). Recent studies[17,25,33] have payed more attentions on models of combinations of different noninvasive parameters, for example inhibin B FSH ratio, and adorable AUC has been produced. Thus similar studies on other

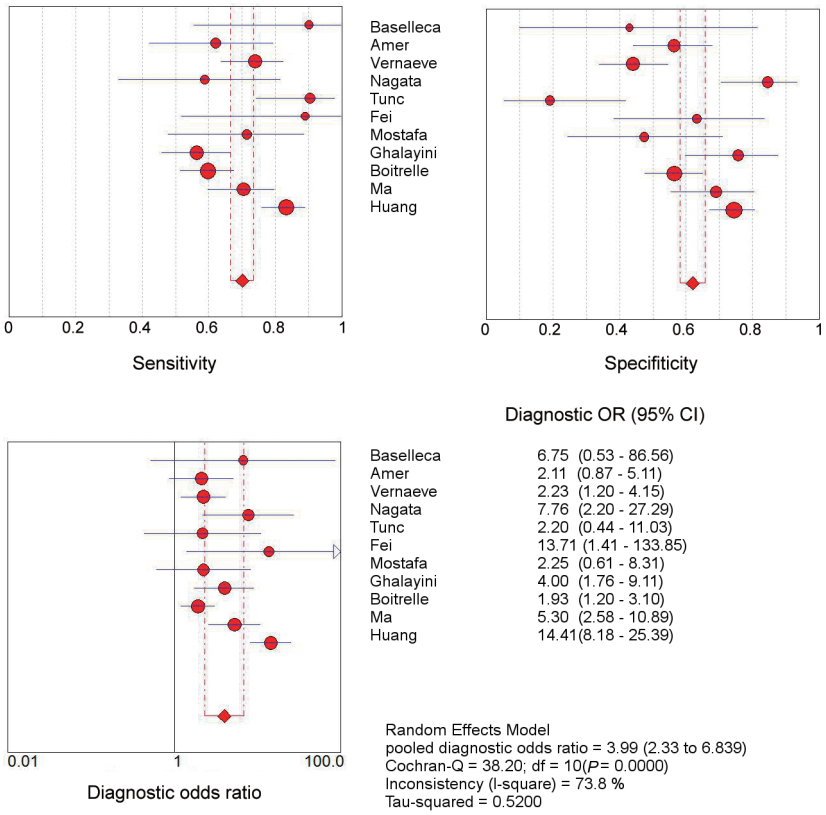

Figure 1: Sensitivity, specificity and diagnosis OR estimates from each of the 11 studies. OR: odds ratio; CI: confidence interval; df: degree of freedom.

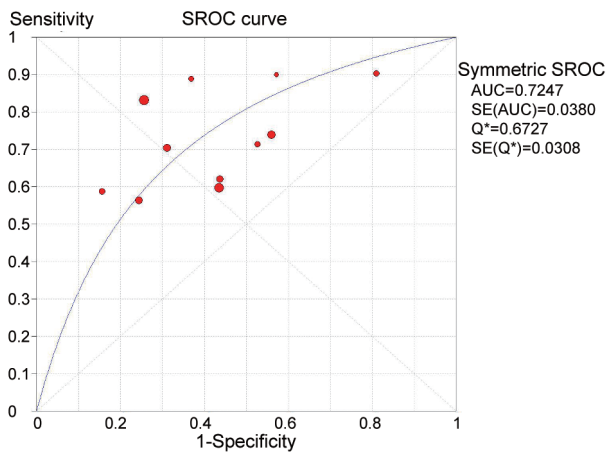

Figure 2: SROC curve from each of the 11 studies. SROC: summary receiver operation characteristics; AUC: area under curve; SE: standard error; Q*: Q-index.

noninvasive parameters, such as inhibin B, testis volume, leptin and on models are of great value.

Our review has some limitations. Diagnosis criteria for NOA were different among authors. Thus only studies with the term NOA were included. Our analysis lacked data on FSH level for each patient. Then, a threshold could not be calculated. However, different threshold means different specificity and sensitivity, which resulting in the SROC. And public bias could be evidenced in the forest plot, luckily subgroup analyses showed little heterogeneity.

Table 2: Stratified analyses for the evaluation of heterogeneity in studies

Subgroup	*n*[a]	*Summary OR (95% CI)*	*P*[b]	I^2	*AUC*[c]
Region					
1	4	9.17 (5.06–16.63)	0.19	0.374	0.81±0.04
2+3	7	2.30 (1.70–3.11)	0.81	0	0.64±0.03
Age					
≤33	4	9.28 (4.79–17.98)	0.19	0.367	0.81±0.06
>33	7	2.42 (1.80–3.26)	0.45	0	0.66±0.02

[a]Number of studies; [b]The heterogeneity *P* value; [c]Area under the SROC; Region – 1: China+Japan for East Asia; 2: Eygpt+Turkey+Jordan for the Middle East; 3: Spain+Belgium+UK+France for Europe. SROC: susmmary receiver operation characteristics; OR: odds ratio; CI: confidence interval; AUC: area under curve

CONCLUSIONS

Follicle-stimulating hormone had moderate diagnostic value as an independent predictor for SRR in patients with NOA. Region and patients' age might influence its diagnostic value. FSH showed more diagnostic value in East Asia and with younger patients. The threshold was still unclear, thus, more detailed diagnosis tests should be anticipated in the future to confirm the diagnostic value of other noninvasive parameters and models of combinations of them.

AUTHOR CONTRIBUTIONS

QY reviewed articles, analyzed data, and drafted the manuscript; YPH reviewed articles, analyzed data and revised the manuscript critically; HXW participated in as the third reviewer and drafting the manuscript; KH participated in data analyzing and revised the manuscript; YXW participated in its design and helped to draft the manuscript; YRH

supervised the project and revised manuscript; BC conceived of the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (No. 81270741).

REFERENCES

1 Willott GM. Frequency of azoospermia. *Forensic Sci Int* 1982; 20: 9–10.
2 Jarow JP, Espeland MA, Lipshultz LI. Evaluation of the azoospermic patient. *J Urol* 1989; 142: 62–5.
3 Devroey P, Liu J, Nagy Z, Tournaye H, Silber SJ, *et al*. Normal fertilization of human oocytes after testicular sperm extraction and intracytoplasmic sperm injection. *Fertil Steril* 1994; 62: 639–41.
4 Ishikawa T. Surgical recovery of sperm in non-obstructive azoospermia. *Asian J Androl* 2012; 14: 109–15.
5 Chan PT, Schlegel PN. Nonobstructive azoospermia. *Curr Opin Urol* 2000; 10: 617–24.
6 Glander HJ, Horn LC, Dorschner W, Paasch U, Kratzsch J. Probability to retrieve testicular spermatozoa in azoospermic patients. *Asian J Androl* 2000; 2: 199–205.
7 Chen CS, Chu SH, Lai YM, Wang ML, Chan PR. Reconsideration of testicular biopsy and follicle-stimulating hormone measurement in the era of intracytoplasmic sperm injection for non-obstructive azoospermia? *Hum Reprod* 1996; 11: 2176–9.
8 Ezeh UI, Moore HD, Cooke ID. A prospective study of multiple needle biopsies versus a single open biopsy for testicular sperm extraction in men with non-obstructive azoospermia. *Hum Reprod* 1998; 13: 3075–80.
9 Jarvi K, Lo K, Fischer A, Grantmyre J, Zini A, *et al*. CUA Guideline: the workup of azoospermic males. *Can Urol Assoc J* 2010; 4: 163–7.
10 Ballescá JL, Balasch J, Calafell JM, Alvarez R, Fábregues F, *et al*. Serum inhibin B determination is predictive of successful testicular sperm extraction in men with non-obstructive azoospermia. *Hum Reprod* 2000; 15: 1734–8.
11 Amer M, Abd Elnasser T, El Haggar S, Mostafa T, Abdel-Malak G, *et al*. May-Grünwald-Giemsa stain for detection of spermatogenic cells in the ejaculate: a simple predictive parameter for successful testicular sperm retrieval. *Hum Reprod* 2001; 16: 1427–32.
12 Vernaeve V, Tournaye H, Schiettecatte J, Verheyen G, Van Steirteghem A, *et al*. Serum inhibin B cannot predict testicular sperm retrieval in patients with non-obstructive azoospermia. *Hum Reprod* 2002; 17: 971–6.
13 Nagata Y, Fujita K, Banzai J, Kojima Y, Kasima K, *et al*. Seminal plasma inhibin-B level is a useful predictor of the success of conventional testicular sperm extraction in patients with non-obstructive azoospermia. *J Obstet Gynaecol Res* 2005; 31: 384–8.
14 Fei Qianjin HX, Zhang Liya, Li Chengdi. The predictive value of serum inhibin B for successful testicular sperm extraction in patients with non-obstructive azoospermia. *Chin J Androl* 2006; 20: 25–7.
15 Tunc L, Kirac M, Gurocak S, Yucel A, Kupeli B, *et al*. Can serum inhibin B and FSH levels, testicular histology and volume predict the outcome of testicular sperm extraction in patients with non-obstructive azoospermia? *Int Urol Nephrol* 2006; 38: 629–35.
16 Mostafa T, Amer MK, Abdel-Malak G, Nsser TA, Zohdy W, *et al*. Seminal plasma anti-Müllerian hormone level correlates with semen parameters but does not predict success of testicular sperm extraction (TESE). *Asian J Androl* 2007; 9: 265–70.
17 Ma Y, Chen B, Wang H, Hu K, Huang Y. Prediction of sperm retrieval in men with non-obstructive azoospermia using artificial neural networks: leptin is a good assistant diagnostic marker. *Hum Reprod* 2011; 26: 294–8.
18 Boitrelle F, Robin G, Marcelli F, Albert M, Leroy-Martin B, *et al*. A predictive score for testicular sperm extraction quality and surgical ICSI outcome in non-obstructive azoospermia: a retrospective study. *Hum Reprod* 2011; 26: 3215–21.
19 Ghalayini IF, Al-Ghazo MA, Hani OB, Al-Azab R, Bani-Hani I, *et al*. Clinical comparison of conventional testicular sperm extraction and microdissection techniques for non-obstructive azoospermia. *J Clin Med Res* 2011; 3: 124–31.
20 Huang X, Bai Q, Yan LY, Zhang QF, Geng L, *et al*. Combination of serum inhibin B and follicle-stimulating hormone levels cannot improve the diagnostic accuracy on testicular sperm extraction outcomes in Chinese non-obstructive azoospermic men. *Chin Med J (Engl)* 2012; 125: 2885–9.
21 Brugo-Olmedo S, De Vincentiis S, Calamera JC, Urrutia F, Nodar F, *et al*. Serum inhibin B may be a reliable marker of the presence of testicular spermatozoa in patients with nonobstructive azoospermia. *Fertil Steril* 2001; 76: 1124–9.
22 Huang X, Zheng J, Wu X. Surgical retrieval of testicular spermatozoa and intracytoplasmic sperm injection in non-obstructive azoospermic patients. *Chin J Urol* 2002; 11: 002.
23 Bohring C, Schroeder-Printzen I, Weidner W, Krause W. Serum levels of inhibin B and follicle-stimulating hormone may predict successful sperm retrieval in men with azoospermia who are undergoing testicular sperm extraction. *Fertil Steril* 2002; 78: 1195–8.
24 Colpi GM, Colpi EM, Piediferro G, Giacchetta D, Gazzano G, *et al*. Microsurgical TESE versus conventional TESE for ICSI in non-obstructive azoospermia: a randomized controlled study. *Reprod Biomed Online* 2009; 18: 315–9.
25 Ramasamy R, Padilla WO, Osterberg EC, Srivastava A, Reifsnyder JE, *et al*. A comparison of models for predicting sperm retrieval before microdissection testicular sperm extraction in men with nonobstructive azoospermia. *J Urol* 2013; 189: 638–42.
26 Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, *et al*. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; 120: 667–76.
27 Devillé WL, Buntinx F, Bouter LM, Montori VM, de Vet HC, *et al*. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; 2: 9.
28 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–60.
29 Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000; 45: 23–41.
30 Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ *Tech Rep Ser* 2000; 894: i-xii, 1–253.
31 Dongfeng G, Guangyong H, Xigui W, Xiufang D. Relationship between body mass index and major cardiovascular diseases in Chinese population. *Natl Med J China* 2002; 82: 13–6.
32 Cabler S, Agarwal A, Flint M, du Plessis SS. Obesity: modern man's fertility nemesis. *Asian J Androl* 2010; 12: 480–9.
33 Grunewald S, Glander HJ, Paasch U, Kratzsch J. Age-dependent inhibin B concentration in relation to FSH and semen sample qualities: a study in 2448 men. *Reproduction* 2013; 145: 237–44.