Inhibin B and anti-Müllerian hormone as markers of persistent spermatogenesis in men with non-obstructive azoospermia: a meta-analysis of diagnostic accuracy studies

Konstantinos A. Toulis*, Paschalia K. Iliadou, Christos A. Venetis, Christos Tsametis, Basil C. Tarlatzis, Ioannis Papadimas, and Dimitrios G. Goulis

Unit of Reproductive Endocrinology, First Department of Obstetrics and Gynecology, Aristotle University of Thessaloniki, “Papageorgiou” General Hospital, Ring Road, Nea Efkarpia, 54603 Thessaloniki, Greece

*Correspondence address. Tel: +30-2310-693131; Fax: +30-2310-991510, E-mail: touliskos@gmail.com

Submitted on December 19, 2009; resubmitted on April 19, 2010; accepted on May 21, 2010

Introduction:
A non-invasive test, which could predict the presence of sperm during a testicular sperm extraction (TESE) procedure in men with non-obstructive azoospermia (NOA), would be of profound clinical importance. Inhibin B (Inh-B) and anti-Müllerian hormone (AMH) have been proposed as direct markers of Sertoli cell function and indirect markers of spermatogenesis.

Methods:
A search was conducted in the electronic databases MEDLINE, EMBASE and Cochrane Central Register of Controlled Trials from inception through June 2009. Thirty-six different studies reported data on the predictive value of one or more index markers (serum Inh-B: 32 studies, seminal Inh-B: 5 studies, serum AMH: 2 studies, seminal AMH: 4 studies) and were included in the systematic review. Nine studies, which had serum Inh-B as index marker, met the predefined criteria and were included in the meta-analysis.

Results:
Serum Inh-B demonstrated a sensitivity of 0.65 (95% confidence interval [CI]: 0.56–0.74) and a specificity of 0.83 (CI: 0.64–0.93) for the prediction of the presence of sperm in TESE. When the pre-test probability of 41% was incorporated in a Fagan’s nomogram, resulted in a positive post-test probability of 73% and a negative post-test probability of 23% for the presence of sperm in TESE.

Conclusions:
Serum Inh-B cannot serve as a stand-alone marker of persistent spermatogenesis in men with NOA. Although limited, evidence on serum AMH and serum/seminal AMH do not support their diagnostic value in men with NOA.
**Introduction**

The introduction to clinical practice of assisted reproduction techniques such as intra-cytoplasmic sperm injection (ICSI) made fatherhood possible for men with non-obstructive azoospermia (NOA). The extraction of sperm from the testis through fine-needle aspiration (FNA), testicular sperm extraction (TESE) or open biopsy can result in a favorable reproductive outcome. Nevertheless, all these procedures are invasive in nature with variable possibility of successful sperm extraction. As a consequence, a non-invasive test that could predict the presence of sperm in men with NOA would be of profound clinical importance. Until recently, follicular stimulating hormone (FSH) and testicular volume constituted the best options to make such a prediction. Inhibin B (Inh-B) and anti-Müllerian hormone (AMH), two glycoproteins of the transforming growth factor-β superfamily, are produced almost exclusively by the Sertoli cells and have been proposed as direct markers of their function and indirect markers of spermatogenesis.

Inh-B, the only inhibin form present in males (Illingworth et al., 1996), was named for its ability to negatively regulate FSH. Serum Inh-B levels have been found to be significantly lower in men with testicular dysfunction compared with controls, as well as positively correlated with classical markers of spermatogenesis, such as sperm concentration, total sperm count and testicular volume (Anawalt et al., 1996; Jensen et al., 1997; Pierik et al., 1998; Pierik et al., 2003). AMH was named for its effect of causing regression of the Müllerian ducts, the anlagen of the uterus, fallopian tubes and upper vagina (Lee and Donohoe, 1993). AMH is secreted bi-directionally by Sertoli cells: apically into the seminiferous tubules, and basally toward the interstitium and the circulation. Serum concentrations of AMH are maintained at a high level until puberty, when they decrease dramatically, remaining very low during adulthood (Lee et al., 1996). After puberty, AMH is secreted preferentially by the apical pole of the Sertoli cell, resulting in higher concentrations of AMH in the seminal plasma than in serum (Fenichel et al., 1999). Seminal AMH levels have been found to be significantly lower (Fallat et al., 1996; Fenichel et al., 1999; Fujisawa et al., 2002; Mostafa et al., 2007; Duvilla et al., 2008) or equal (Al-Qahtani et al., 2005) in subfertile men compared with fertile controls. In a similar way, serum AMH levels have been found to be lower (Al-Qahtani et al., 2005; Muttukrishna et al., 2007; Goulis et al., 2008) or equal (Isikoglu et al., 2006; Appasamy et al., 2007; Tuttelmann et al., 2009) in subfertile men compared with controls.

During the last decade, a series of studies that attempted to estimate the predictive value of serum or seminal Inh-B or AMH as non-invasive markers of persistent spermatogenesis in men with NOA have reported contradictory results. The aim of the present study was to estimate this predictive value through a systematic review of diagnostic accuracy studies published in the literature and a meta-analysis of the best evidence available. To preserve the two-dimensional nature of the diagnostic performance data while incorporating the effect of either (explicit or implicit) clinical or methodological differences among studies (Reitsma et al., 2005), a bivariate model for the meta-analysis of diagnostic accuracy studies was used.

**Methods**

**Search strategy**

To identify eligible studies, the main search was conducted in the electronic databases MEDLINE, EMBASE and Cochrane Central Register of Controlled Trials from inception through June 2009 (last search update, April 2010), using various combinations of the Medical Subject Headings (MeSH) terms ‘Anti-Müllerian Hormone’, ‘Inhibins’, Inhibin-beta Subunits and ‘Sperm Retrieval’, ‘Spermatozoa’, ‘Spermatozoids’, without language restriction. The reference sections of all relevant studies and reviews were used for search completion. The main search, as well as screening of titles, abstracts and full-text articles, was completed independently by two reviewers and any discrepancy was solved by consensus.

**Eligibility of relevant studies**

Studies which provided data on the accuracy of serum or seminal AMH and/or serum or seminal Inh-B (index tests) for the prediction of the presence of spermatozoa in sperm retrieval techniques in men with azoospermia were considered potentially eligible. Studies were included in the meta-analysis only if TESE (reference standard) was performed in men with NOA. TESE was used as the reference standard based on the findings of a relevant systematic review, which reported a higher sperm retrieval rate in men with NOA undergoing TESE as compared with FNA (Donoso et al., 2007). Studies were excluded from the analysis if (1) sperm retrieval techniques other than TESE had been used as the reference test or if (2) data allowing the construction of 2 × 2 contingency tables were not available, even after communication with the primary investigators. Reviews or letters to the editor were not considered eligible.

**Data extraction**

Information from each study was extracted independently by two reviewers using a standardized data extraction form, which had been constructed on the basis of the STARD checklist (Bossuyt et al., 2003). The general characteristics of the studies, characteristics of the group with azoospermia, methodology of the index and the reference tests and outcomes in the form of true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) results were extracted, where available, and double checked. Data from patients with NOA were only considered for the meta-analysis. Where appropriate, data set was completed through communication with the authors. The QUADAS tool was used independently by two reviewers for the assessment of the included studies (Whiting et al., 2003). Any disagreement was resolved by consensus.

**Outcomes**

To assess the accuracy of an index test (serum/semenal AMH and Inh-B) for the prediction of the presence of spermatozoa in men with azoospermia, summary statistics, namely sensitivity, specificity, positive likelihood ratio (LR), negative LR and diagnostic odds ratio (DOR) were calculated, when possible. Positive LR is defined as the ratio of the true positive to false positive rate of a test [i.e. how many times more likely is to find Inh-B concentrations above a specific threshold in men with successful sperm retrieval in TESE compared with those with unsuccessful results]. Negative LR is the ratio of the false negative to true negative rate of a test (Grimes and Schulz, 2003; Broekmans et al., 2006). Finally, diagnostic
odds ratio is the ratio of the odds of sensitivity to the inverse of the odds of specificity. Positive LRs above 10 and negative LRs below 0.1 are considered as indicators of an adequate diagnostic test, while values between 5 and 10 and between 0.2 and 0.1, respectively, are considered to indicate a moderate test. A DOR greater than 80 was considered to identify a diagnostic test with an excellent performance, whereas that between 10 and 80 was considered as an indicator of an adequate test (Deeks, 2001; Reitsma et al., 2005). As an alternative indicator of the overall discriminatory capacity of the index tests, area under the summary ROC (receiver operator characteristic) curve (sROC-AUC) was calculated with a value of 0.5 to be considered as a guess (Swets, 1988). To gain a measure of the relative performance of the index test, summary statistics and sROC-AUC were calculated for follicle stimulating hormone (FSH) from the same population. This comparative analysis included only studies that have directly compared the tests in the same patients, in the same setting in an effort to limit confounding (Leeflang et al., 2008). This approach was also considered as an alternative method to gain insight into the complementary diagnostic strength of Inh-B to established markers of spermatogenesis, such as FSH, since individual patient data that could have allowed a reliable combined analysis of the two hormones were not available. Finally, a Fagan nomogram, a two-dimensional graphical tool for estimating how much the result of a diagnostic test changes the probability that a patient has a disease, was designed to estimate the clinical value of the index test (Fagan, 1975). After logarithmic transformation of the basic formula of the Bayes’ theorem, post-test log-odds are linear functions of the pre-test log-odds and the log LRs. Using labeling in terms of prevalence and spacing at the log-odds, Fagan nomogram provides the visual depiction of this relationship using three axes (pre-test log-odds, log-LR and post-test log-odds). This approach bypasses the double conversion from pre-test prevalence to pre-test odds and post-test odds to post-test prevalence and permits individualization of diagnostic evidence, since LRs could be generated from any target population (Jaeschke et al., 1994; Deeks and Altman, 2004).

Statistical analysis

Data synthesis was performed within the bivariate mixed-effects logistic regression modeling framework. This is a two-level model fitting a bivariate normal model for the logit transformations of sensitivity and specificity between studies and independent binomial distributions for true positives and true negatives in each study conditional on the sensitivity and specificity in each study (Reitsma et al., 2005). This model incorporates the possible correlation between logit sensitivity and specificity within studies and allows for heterogeneity beyond chance due to clinical or methodological differences between studies. A standard correction of adding 0.5 to all cells of the $2 \times 2$ table was applied when either sensitivity or specificity was 100%. An SROC was constructed using the bivariate model to produce a 95% confidence contour within ROC space. Heterogeneity between the results of different studies was examined by $I^2$ test, which can be interpreted as the percentage of total variation across several studies due to heterogeneity (Higgins et al., 2003). All analyses were implemented using Stata/MP 10.0 for Windows (StataCorp LP, 4905 College Station, TX 77845, USA). Statistical calculations were performed using the MIDAS routine (Dwamena, 2007).

Results

Search results

From the 1550 studies initially identified, 1514 studies were excluded on a title/abstract basis as irrelevant, reviews or duplicates. Thirty-six different studies reported data on the efficacy of one or more index tests (serum Inh-B: 32 studies, seminal Inh-B: five studies, serum AMH: two studies, seminal AMH: four studies) for the presence of spermatozoa in men with azoospermia undergoing sperm retrieval techniques. These studies were included in the systematic review and considered as potentially eligible for the meta-analysis. Eleven of them were excluded from the meta-analysis since sperm retrieval techniques other than TESE had been performed (Pierik et al., 1998; Bohring and Krause, 1999; Foresta et al., 1999; Yalti et al., 2002; Ramos et al., 2004; Halder et al., 2005; Smits et al., 2007; Zhang et al., 2007; Gouli et al., 2008; Nowroozi et al., 2008; Gouli et al., 2009). It is worth mentioning that, in most of them, data were not available in an extractable format. Among these studies, men with suspected obstructive azoospermia had been recruited in two (Ramos et al., 2004; Smits et al., 2007), while FNA had been performed in four (Foresta et al., 1999; Halder et al., 2005; Gouli et al., 2008; Gouli et al., 2009). Moreover, 12 studies were excluded from the meta-analysis because data were not available in an extractable format (in many of them, only mean values were reported) (Bohring et al., 2002; Westlander et al., 2003; Tsujimura et al., 2004; Betteilla et al., 2005; Tsujimura A et al., 2005; Dong et al., 2006; Fei et al., 2006; Liu et al., 2006; Tunc et al., 2006; Zitzmann et al., 2006; Koga et al., 2007; Ferhi et al., 2009). Finally, only nine studies which had Inh-B as index test, met the predefined criteria (von Eckardstein et al., 1999; Ballesca et al., 2000; Brugo-Olmedo et al., 2001; Vernaeeve et al., 2002; Bailly et al., 2003; Nagata et al., 2005; Ziaee et al., 2006; Duvilla et al., 2008; Mitchell et al., 2010) and were included in the meta-analysis.

Systematic review

Main data of the systematic review are summarized in the Table I. The majority of the trials were held in Europe and published between 1998 and 2010. Serum Inh-B was the most commonly investigated among the index tests to predict the presence of spermatozoa. Among these studies, three had recruited only men with Klinefelter syndrome (Westlander et al., 2003; Koga et al., 2007; Ferhi et al., 2009). In many of the trials, results were not extractable in the format of TP, FP, FN and TN or not presented as sensitivity or specificity for a given threshold, since original investigators preferred to investigate the potential value of the index tests by comparing their mean values among groups. In most of the studies, thresholds applied were based on ROC analyses; however, in some of them, assay detection limit or evidence from previous literature were used for the threshold selection. Despite that, threshold selection across studies did not seem to correlate with the index test performance. In all of the studies, enzyme-linked immunosorbent assays (ELISA) were used for the measurements. With the exception of serum Inh-B, the small number of studies that evaluated seminal Inh-B, serum AMH and seminal AMH did not allow data synthesis and were not included in the meta-analysis.

Seminal AMH (number of studies: 4)

These studies evaluated the performance of seminal AMH as a non-invasive marker of persistent spermatogenesis in men with NOA. Despite the first encouraging results (Fenichel et al., 1999), the predictive value of seminal AMH for TESE outcome was not confirmed in subsequent studies (Mostafa et al., 2007; Duvilla et al., 2008; Mitchell et al., 2010).
<table>
<thead>
<tr>
<th>First author, year, country</th>
<th>Description of design</th>
<th>Study groups</th>
<th>Cause of azoospermia</th>
<th>Index test assay: inhibin B (ng/l) or AMH (pmol/l or ng/ml)</th>
<th>Reference test in patients with azoospermia</th>
<th>Results</th>
<th>QUADAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitchell, 2010, France</td>
<td>Prospective, single center</td>
<td>NOA (n = 139)</td>
<td>Idiopathic (n = 128), KS (n = 7), Yq microdeletion (n = 4)</td>
<td>ELISA, detection limit = 7.8, CVs by manufacturer, threshold applied &gt; 27.5 based on ROC analysis</td>
<td>TESE (n = 139)</td>
<td>TP = 36, FP = 17, FN = 24, TN = 62</td>
<td>8</td>
</tr>
<tr>
<td>Goulis, 2009, Greece</td>
<td>Cross-sectional, single center</td>
<td>NOA/OA (n = 51), control (n = 31, fertile)</td>
<td>Idiopathic (n = 34), cryptorchidism (n = 4), varicocele (n = 3), KS (n = 3), OA (n = 3)</td>
<td>ELISA, detection limit = 7, intra-assay CV = 3.5%, inter-assay CV = 6.2%, threshold applied &gt; 14.5 based on ROC analysis</td>
<td>FNA (n = 45)</td>
<td>TP = 21, FP = 17, FN = 0, TN = 7</td>
<td>8</td>
</tr>
<tr>
<td>Fehri, 2008, France</td>
<td>Retrospective, single center</td>
<td>NOA (n = 27)</td>
<td>Non-mosaic KS (n = 27)</td>
<td>Not reported</td>
<td>TESE (n = 23), TESA (n = 4)</td>
<td>Not extractable</td>
<td>5</td>
</tr>
<tr>
<td>Duvilla, 2008, France</td>
<td>Prospective, single center</td>
<td>NOA (n = 49), OA (n = 18), OAT (n = 28), control (n = 27, fertile)</td>
<td>Idiopathic (n = 27), chemo/radiotherapy (n = 11), KS (n = 8), Yq microdeletion (n = 3)</td>
<td>ELISA, detection limit and CVs not reported, threshold applied &gt; 15 based on ROC analysis</td>
<td>TESE (n = 26)</td>
<td>TP = 7, FP = 3, FN = 4, TN = 12</td>
<td>6</td>
</tr>
<tr>
<td>Goulis, 2008, Greece</td>
<td>Prospective, single center</td>
<td>NOA/OA (n = 51), control (n = 31, fertile)</td>
<td>Idiopathic (n = 34), cryptorchidism (n = 4), varicocele (n = 3), KS (n = 3), OA (n = 3)</td>
<td>ELISA, detection limit = 7, intra-assay CV = 3.5%, inter-assay CV = 6.2%, threshold applied &gt; 14.5 based on ROC analysis</td>
<td>FNA (n = 45)</td>
<td>TP = 21, FP = 17, FN = 0, TN = 7</td>
<td>8</td>
</tr>
<tr>
<td>Nowroozi, 2008, Iran</td>
<td>Cross-sectional, multi-center</td>
<td>NOA (n = 70)</td>
<td>Varicocele: 19%, orchiopexy: 6%, hernioraphy: 3%, chemotherapy: 2%, urinary tract infection: 2%, orchitis: 3%</td>
<td>Not reported</td>
<td>Biopsy (n = 49)</td>
<td>Not extractable</td>
<td>6</td>
</tr>
<tr>
<td>Koga, 2007, Japan</td>
<td>Retrospective, multi-center</td>
<td>NOA (n = 26)</td>
<td>Non-mosaic KS (n = 26)</td>
<td>Not reported</td>
<td>Micro-dissection TESE (n = 26)</td>
<td>Not extractable</td>
<td>5</td>
</tr>
<tr>
<td>Smit, 2007, Netherlands</td>
<td>Retrospective, single center</td>
<td>Suspected OA (n = 43)</td>
<td>OA (n = 43), patients with vasectomy or CBVAD excluded</td>
<td>Assay not reported, threshold applied &gt; 150 based on previous study (lower normal value)</td>
<td>MESA (n = 43), biopsy (n = 37)</td>
<td>TP = 22, FP = 11, FN = 5, TN = 3</td>
<td>5</td>
</tr>
<tr>
<td>Zhang, 2007, China</td>
<td>Non-extractable</td>
<td>NOA (n = 33), OA (n = 37), control: (n = 25, fertile)</td>
<td>Non-extractable</td>
<td>Non-extractable</td>
<td>Biopsy (n = 70)</td>
<td>Non-extractable</td>
<td>–</td>
</tr>
<tr>
<td>Dong, 2006, China</td>
<td>Non-extractable</td>
<td>NOA/OA (n = 83)</td>
<td>Non-extractable</td>
<td>Non-extractable</td>
<td>Biopsy (n = 83)</td>
<td>Non-extractable</td>
<td>–</td>
</tr>
<tr>
<td>Fei, 2006, China</td>
<td>Non-extractable</td>
<td>NOA/OA/control (n = 63)</td>
<td>Non-extractable</td>
<td>Non-extractable</td>
<td>TESE (n = Non-extractable)</td>
<td>Non-extractable</td>
<td>–</td>
</tr>
<tr>
<td>Liu, 2006, China</td>
<td>Non-extractable</td>
<td>NOA (n = 40), OA (n = 20), control: (n = 10, fertile)</td>
<td>Non-extractable</td>
<td>Non-extractable</td>
<td>TESE (n = 40)</td>
<td>Non-extractable</td>
<td>–</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Sample Size</td>
<td>Patients Excluded</td>
<td>Test Methods</td>
<td>TESE (n = 52)</td>
<td>TP = 28, FP = 23, FN = 3, TN = 4</td>
<td>5</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------------</td>
<td>-------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>---</td>
</tr>
<tr>
<td>Tunk, 2006, Turkey</td>
<td>Retrospective, single center</td>
<td>NOA (n = 52)</td>
<td>Yq microdeletions, CBAVD and chromosomal abnormalities excluded</td>
<td>ELISA, detection limit = 15, CVs by manufacturer, threshold applied &gt; 6.25 based on ROC analysis</td>
<td>TESE (n = 52)</td>
<td>TP = 28, FP = 23, FN = 3, TN = 4</td>
<td></td>
</tr>
<tr>
<td>Ziaee, 2006, Iran</td>
<td>Design not clearly stated, single center</td>
<td>NOA (n = 85)</td>
<td>Patients with KS excluded</td>
<td>ELISA, detection limit and CVs not reported, threshold applied &gt; 39.8 based on ROC analysis</td>
<td>TESE (n = 85)</td>
<td>TP = 13, FP = 3, FN = 5, TN = 64</td>
<td></td>
</tr>
<tr>
<td>Zitzmann, 2006, Germany</td>
<td>Retrospective cohort, single center</td>
<td>NOA/OA (n = 203)</td>
<td>Idiopathic (n = 179), CBAVD (n = 7), Yq microdeletions (n = 5), ejaculation disorders (n = 6), post-vasectomy (n = 4), KS (n = 2)</td>
<td>ELISA, detection limit = 7.8, intra-assay CV &lt; 4%, inter-assay CV &lt; 20%</td>
<td>TESE (n = 203)</td>
<td>Not extractable</td>
<td></td>
</tr>
<tr>
<td>Bettella, 2005, Italy</td>
<td>Retrospective, single center</td>
<td>NOA (n = 125)</td>
<td>Idiopathic (n = 49), cryptorchidism (n = 28), varicocele (n = 18), post-mumps orchitis (n = 13), chemo/radiotherapy (n = 8), testicular trauma (n = 6), testicular torsion (n = 3)</td>
<td>Not reported</td>
<td>FNA (n = 125), TESE (n = 125)</td>
<td>Not extractable</td>
<td></td>
</tr>
<tr>
<td>Halder, 2005, India</td>
<td>Retrospective, single center</td>
<td>NOA (n = 46), control (n = 5, fertile)</td>
<td>Idiopathic (n = 46), chromosomal abnormalities not evaluated</td>
<td>ELISA, detection limit = 15, intra-assay and inter-assay CV &lt; 7%, threshold applied not reported</td>
<td>FNA or biopsy (n = 13)</td>
<td>Not extractable</td>
<td></td>
</tr>
<tr>
<td>Nagata, 2005, Japan</td>
<td>Design not clearly stated, single center</td>
<td>NOA (n = 62)</td>
<td>Iatrogenic NOA excluded</td>
<td>ELISA, detection limit = 15, intra-assay CV = 6.4%, inter-assay CV = 7.2%, threshold applied &gt; 34 based on ROC analysis</td>
<td>TESE (n = 62)</td>
<td>TP = 12, FP = 2, FN = 5, TN = 43</td>
<td></td>
</tr>
<tr>
<td>Tsujimura, 2004, Japan</td>
<td>Retrospective, multi-center</td>
<td>NOA (n = 100)</td>
<td>Chromosomal abnormalities excluded</td>
<td>Not reported</td>
<td>Micro-dissection TESE (n = 100)</td>
<td>Not extractable</td>
<td></td>
</tr>
<tr>
<td>Ramos, 2004, Netherlands</td>
<td>Retrospective, single center</td>
<td>NOA/OA (n = 147)</td>
<td>Post-vasectomy (n = 47), CBAVD (n = 38), idiopathic (n = 41), other (n = 21), patients with testis &lt; 12 ml, high FSH, Yq microdeletions excluded</td>
<td>ELISA, detection limit = 15, CVs not reported, threshold applied &gt; 80 based on previous studies</td>
<td>PESA (n = 147), biopsy (n = 147)</td>
<td>TP = 113, FP = 8, FN = 16, TN = 10</td>
<td></td>
</tr>
<tr>
<td>Tsujimura, 2004, Japan</td>
<td>Retrospective, multi-center</td>
<td>NOA (n = 100)</td>
<td>Chromosomal abnormalities excluded</td>
<td>Not reported</td>
<td>Micro-dissection TESE (n = 100)</td>
<td>Not extractable</td>
<td></td>
</tr>
<tr>
<td>Bailly, 2003, France</td>
<td>Retrospective, single center</td>
<td>NOA (n = 75), OA (n = 39)</td>
<td>Not reported</td>
<td>ELISA, detection limit = 15, intra-assay and inter-assay CVs &lt; 7%, threshold applied &gt; 15 based on detection limit</td>
<td>TESE (n = 75)</td>
<td>TP = 17, FP = 16, FN = 9, TN = 33</td>
<td></td>
</tr>
<tr>
<td>Westlander, 2003, Sweden</td>
<td>Prospective, single center</td>
<td>NOA (n = 18), controls (n = 18, fertile)</td>
<td>Non-mosaic KS (n = 18)</td>
<td>ELISA, detection limit = 5, intra-assay CV &lt; 9%, inter-assay CV = 15%, threshold applied not reported</td>
<td>TESE (n = 18)</td>
<td>Not extractable</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author, year, country</th>
<th>Description of design</th>
<th>Study groups</th>
<th>Cause of azoospermia</th>
<th>Index test assay: inhibin B (ng/l) or AMH (pmol/l or ng/ml)</th>
<th>Reference test in patients with azoospermia</th>
<th>Results</th>
<th>QUADAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bohring, 2002, Germany</td>
<td>Retrospective, multi-center</td>
<td>NOA/OA (n = 52)</td>
<td>Not reported</td>
<td>ELISA, detection limit not reported, intra-assay CV = 5.7%, threshold applied not reported</td>
<td>Biopsy (n = 52), TESE (n = 43)</td>
<td>Not extractable</td>
<td>5</td>
</tr>
<tr>
<td>Vernaeve, 2002, Belgium</td>
<td>Retrospective, single center</td>
<td>NOA (n = 185)</td>
<td>Patients with KS or central hypogonadism excluded</td>
<td>ELISA, detection limit = 8, inter-assay CV = 7.5%, threshold applied &gt;13.7 based on ROC analysis</td>
<td>TESE (n = 185)</td>
<td>TP = 41, FP = 34, FN = 51, TN = 59</td>
<td>7</td>
</tr>
<tr>
<td>Yalti, 2002, Turkey</td>
<td>Prospective, single center</td>
<td>NOA/OAT (n = 52), control (n = 20, normal spermiogram)</td>
<td>Germ cell failure (n = 15)</td>
<td>ELISA, detection limit = 10, intra-assay CV = 3.3%, inter-assay CV = 18%, threshold applied not reported</td>
<td>Biopsy (n = 15)</td>
<td>Not extractable</td>
<td>5</td>
</tr>
<tr>
<td>Brugo-Olmedo, 2001, Argentina</td>
<td>Prospective, single center</td>
<td>NOA (n = 78), OA (n = 15), control (n = 10, fertile)</td>
<td>Idiopathic (n = 72), KS (n = 5), chemotherapy (n = 1)</td>
<td>ELISA, detection limit = 15, intra-assay and inter-assay CVs &lt;7%, threshold applied &gt;53 based on ROC analysis</td>
<td>TESE (n = 72)</td>
<td>TP = 20, FP = 4, FN = 10, TN = 38</td>
<td>7</td>
</tr>
<tr>
<td>Ballesca, 2000, Spain</td>
<td>Prospective, single center</td>
<td>NOA (n = 17), OA (n = 22), control (n = 29, normal spermiogram)</td>
<td>Idiopathic (n = 12), cryptorchidism (n = 3), chemotherapy (n = 2)</td>
<td>ELISA, detection limit = 15, intra-assay CV &lt;11%, inter-assay CV = 15%, threshold applied &gt;40 based on ROC analysis</td>
<td>TESE (n = 17)</td>
<td>TP = 9, FP = 0, FN = 1, TN = 7</td>
<td>7</td>
</tr>
<tr>
<td>Bohring, 1999, Germany</td>
<td>Design not clearly stated, single center</td>
<td>OAT (n = 148)</td>
<td>Not reported</td>
<td>ELISA, intra-assay CV = 5.7%, threshold applied not reported</td>
<td>Biopsy (n = 20)</td>
<td>Not extractable</td>
<td>4</td>
</tr>
<tr>
<td>Foresta, 1999, Italy</td>
<td>Design not clearly stated, single center</td>
<td>NOA/OA (n = 89), control (n = 30, normal spermiogram)</td>
<td>Idiopathic (n = 47), cryptorchidism (n = 15), Yq microdeletion (n = 7), post-mumps orchitis (n = 6), varicocele (n = 4), iatrogenic (n = 6), cystic fibrosis (n = 3), trauma (n = 1)</td>
<td>ELISA, detection limit = 15, intra-assay CV = 6.8%, inter-assay CV = 6.4%, threshold applied not reported</td>
<td>FNA (n = 89)</td>
<td>Not extractable</td>
<td>8</td>
</tr>
<tr>
<td>von Eckardstein, 1999, Germany*</td>
<td>Retrospective, single center</td>
<td>NOA/OA/OAT (n = 91), control (n = 84, fertile)</td>
<td>Idiopathic (n = 27), cryptorchidism (n = 29), OA (n = 24),orchitis (n = 4), chemotherapy (n = 5)</td>
<td>ELISA, detection limit = 7.8, intra-assay CV = 3.3%, inter-assay CV = 18%, thresholds applied 94 and 139, based on the lower limit of normal curve and previous results, respectively</td>
<td>Biopsy (n = 91), TESE (n = 52)</td>
<td>Threshold 1 (94 ng/l): TP = 16, FP = 6, FN = 18, TN = 12, Threshold 2 (139 ng/l): TP = 9, FP = 2, FN = 25, TN = 16</td>
<td>8</td>
</tr>
<tr>
<td>Pierik, 1998, Netherlands</td>
<td>Prospective, single center</td>
<td>OAT (n = 218)</td>
<td>Idiopathic moderate OAT (n = 69), idiopathic severe OAT (n = 58), idiopathic NOA (n = 15), cryptorchidism (n = 17), OA (n = 6), KS (n = 4)</td>
<td>ELISA, detection limit = 5, intra-assay CV &lt;9%, inter-assay CV &lt;15%, threshold applied &gt;139 based on ROC analysis</td>
<td>Biopsy (n = 22)</td>
<td>TP = 7, FP = 2, FN = 2, TN = 11</td>
<td>7</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Center</td>
<td>N</td>
<td>Disease</td>
<td>Control</td>
<td>Assay</td>
<td>Detection Limit</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>-------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Mitchell, 2010, France</td>
<td>Prospective, single center</td>
<td>NOA (n = 139)</td>
<td>139</td>
<td>NOA</td>
<td>Idiopathic (n = 128), KS (n = 7), Yq microdeletion (n = 4)</td>
<td>Non-extractable 8</td>
<td>ELISA, detection limit: 7.8 pg/ml, intra-assay CV = 12%, inter-assay CV = 17%, threshold applied not reported</td>
</tr>
<tr>
<td>Duvilla, 2008, France</td>
<td>Prospective, single center</td>
<td>NOA (n = 49), OA (n = 18), OAT (n = 28), control (n = 20, normal spermiogram)</td>
<td>139</td>
<td>NOA</td>
<td>Idiopathic (n = 27), chemotherapy radiotherapy (n = 11), KS (n = 8), Yq microdeletion (n = 3)</td>
<td>Non-extractable 8</td>
<td>ELISA, detection limit and CVs not reported greater than 30 based on ROC analysis</td>
</tr>
<tr>
<td>Zhang, 2007, China</td>
<td>Non-extractable</td>
<td>NOA (n = 33), OA (n = 37), control: (n = 25, fertile)</td>
<td>139</td>
<td>NOA</td>
<td>Non-extractable</td>
<td>Non-extractable</td>
<td>Biopsy (n = 70)</td>
</tr>
<tr>
<td>Nagata, 2005, Japan</td>
<td>Design not clearly stated, single center</td>
<td>NOA (n = 62)</td>
<td>62</td>
<td>NOA</td>
<td>Iatrogenic NOA excluded</td>
<td>Non-extractable</td>
<td>Biopsy (n = 70)</td>
</tr>
<tr>
<td>El Garem, 2002, Egypt and Belgium</td>
<td>Design not clearly stated, single center</td>
<td>NOA/OA (n = 50), control (n = 10, normal spermiogram), control (n = 10, vasectomy)</td>
<td>50</td>
<td>NOA</td>
<td>Not reported</td>
<td>Non-extractable</td>
<td>Biopsy (n = 70)</td>
</tr>
<tr>
<td>Goulis, 2009, Greece</td>
<td>Cross-sectional, single center</td>
<td>NOA/ OA (n = 51), control (n = 31, fertile)</td>
<td>51</td>
<td>NOA</td>
<td>Idiopathic (n = 34), cryptorchidism (n = 4), varicocele (n = 3), KS (n = 3), OA (n = 3)</td>
<td>ELISA, detection limit: 0.017 ng/ml, intra-assay CV = 4.8%, inter-assay CV = 5.9%, threshold applied greater than 3.6 based on ROC analysis</td>
<td>FNA (n = 45)</td>
</tr>
<tr>
<td>Isikoglu, 2006, Turkey</td>
<td>Prospective, single center</td>
<td>NOA (n = 24), control (n = 23, normal spermiogram)</td>
<td>24</td>
<td>NOA</td>
<td>Varicocele (n = 8), cryptorchidism (n = 4), chromosomal abnormalities excluded</td>
<td>Non-extractable</td>
<td>Tease or TESE (n = 24)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitchell, 2010, France</td>
<td>Prospective, single center</td>
<td>NOA (n = 139)</td>
<td>139</td>
<td>NOA</td>
<td>Idiopathic (n = 128), KS (n = 7), Yq microdeletion (n = 4)</td>
<td>Non-extractable 8</td>
<td>ELISA, detection limit: 2.3 pmol/l, intra-assay CV = 5.9%, inter-assay CV = 10.3%, threshold not reported</td>
</tr>
<tr>
<td>Duvilla, 2008, France</td>
<td>Prospective, single center</td>
<td>NOA (n = 49), OA (n = 18), OAT (n = 28), control (n = 27, normal spermiogram)</td>
<td>139</td>
<td>NOA</td>
<td>Idiopathic (n = 27), chemotherapy radiotherapy (n = 11), KS (n = 8), Yq microdeletion (n = 3)</td>
<td>Non-extractable 8</td>
<td>ELISA, intra-assay CV = 12.3%, inter-assay CV = 14.2%, threshold applied greater than 3.6 pmol/l based on ROC analysis</td>
</tr>
</tbody>
</table>
Serum AMH (number of studies: 2)
These studies evaluated the performance of serum AMH as a non-invasive marker of persistent spermatogenesis in patients with NOA, using either testicular FNA and ROC analysis (Goulis et al., 2009) or TESE and mean values comparison between NOA and fertile control groups (Isikoglu et al., 2006). Both of them failed to provide any evidence of potential utility for serum AMH.

Seminal Inh-b (number of studies: 5)
These studies evaluated the performance of seminal Inh-B as a non-invasive marker of persistent spermatogenesis in men with NOA. Whereas, seminal Inh-B did not differ significantly between azoospermic with positive and negative TESE in an initial report (El Garem et al., 2002), it was found to be a useful predictor for the presence of spermatozoa in subsequent studies, either alone (Nagata et al., 2005) or combination with other parameters (Zhang et al., 2007; Duvilla et al., 2008). The most recent evidence suggests that there is no value in seminal plasma levels of inhibin B as criterion for sperm extraction in men with NOA (Mitchell et al., 2010). Thus, the potential role of seminal Inh-B for the prediction of persistent spermatogenesis in men with NOA undergoing TESE remains debatable and should be further investigated.

Serum Inh-b (number of studies: 32)
Numerous studies evaluated the performance of serum Inh-B as a non-invasive marker of persistent spermatogenesis in men with NOA. To synthesize evidence from a similar clinical setting and secure reliable interpretation, only extractable data regarding patients with NOA from studies in which TESE had been performed were considered suitable for the meta-analysis (von Eckardstein et al., 1999; Ballesca et al., 2000; Brugo-Olmedo et al., 2001; Vernaeve et al., 2002; Bailly et al., 2003; Nagata et al., 2005; Ziaee et al., 2006; Duvilla et al., 2008; Mitchell et al., 2010). For those studies that were excluded from the meta-analysis, a descriptive approach was followed as an alternative to the complete loss of evidence. Four of them suggested that serum Inh-B may be a useful marker of the presence of spermatozoa, mainly in combination with other markers, such as FSH (Pierik et al., 1998; Bohring et al., 2002; Yalti et al., 2002; Tsujimura et al., 2004). On the other hand, nine publications reported a non-satisfactory diagnostic performance of serum Inh-B, after assessing it by means of regression models, ROC analysis or direct comparison of its levels between successful and failed sperm retrieval groups. In two studies on patients with obstructive azoospermia (Ramos et al., 2004; Smit et al., 2007) and in three studies on patients with Klinefelter’s syndrome (Westlander et al., 2003; Koga et al., 2007; Ferhi et al., 2009), serum Inh-B did not predict sperm retrieval. The assessment of four studies from China was proven unfeasible, whereas one study had a sample overlap (Goulis et al., 2009).

Meta-analysis
After combining data regarding men with NOA from nine studies (a total of 726 men), serum Inh-B demonstrated a sensitivity of 0.65 (95% confidence interval [CI]: 0.56–0.74) and a specificity of 0.83 (CI: 0.64–0.93) for the prediction of the presence of spermatozoa in TESE (Fig. 1). No sign of publication bias was detected (P = 0.1), yet a significant amount of heterogeneity was detected (I² = 93%).
The sROC-AUC was calculated at 0.73 (CI: 0.69–0.77). The sROC curve, along with 95% confidence contour, is presented in Fig. 2.

Pooled DOR was calculated at 9 (CI: 3–30), pooled positive LR at 3.8 (CI: 1.6–9.4) and negative LR at 0.42 (CI: 0.30–0.6). The prevalence of the presence of spermatozoa in TESE (pre-test probability) in this group was estimated at 41%. Incorporating this evidence in a Fagan’s nomogram, it appears that the positive post-test probability is 73% and the negative post-test probability 23% (Fig. 3). This finding could be interpreted in terms of clinical practice as follows: a man from that population has 73% chance of a successful TESE if he has a serum Inh-B concentration above the predefined threshold, whereas his chances would drop to 23% if Inh-B concentration is below this threshold.

In the comparative analysis of FSH and Inh-B, eight studies, in which data on diagnostic performance of FSH were available, were included (von Eckardstein et al., 1999; Ballesca et al., 2000; Brugo-Olmedo et al., 2001; Vernaeve et al., 2002; Nagata et al., 2005; Ziaee et al., 2006; Duvilla et al., 2008; Mitchell et al., 2010). FSH demonstrated a sensitivity of 0.71 (CI: 0.48–0.87), a specificity of 0.62 (CI: 0.48–0.75), a positive LR of 1.9 (CI: 1.2–3.1), a negative LR of 0.46 (CI: 0.22–1.00), a DOR of 4 (CI: 1–14) and a sROC-AUC of 0.70 (CI: 0.66–0.74). In the same setting, serum Inh-B demonstrated a sensitivity of 0.65 (CI: 0.55–0.75), a specificity of 0.85 (CI: 0.64–0.95), a positive LR of 4.3 (CI: 1.5–12.3), a negative LR of 0.41 (CI: 0.28–0.60), a DOR of 11 (CI: 3–42) and a sROC -AUC of 0.74 (CI: 0.70–0.78).
recent systematic review (La Marca et al., 2010) preferred to investigate their research hypothesis by simply comparing levels of Inh-B between patients with NOA and controls; in case a significant difference was not detected, they did not proceed to a binary classification test, that could provide extractable evidence and, therefore, they were excluded from the present meta-analysis. Since there are plausible reasons to believe that this subset of studies were not a random sample of the total, but, predominantly, those with non-significant results, the possibility of an overestimation in the diagnostic accuracy of Inh-B, as assessed by the present meta-analysis, should be taken into consideration. In fact, a descriptive analysis of the studies excluded from the meta-analysis suggests that serum Inh-B is a rather non-satisfactory diagnostic marker, in line with the main findings of the meta-analysis. In any case, serum Inh-B demonstrates a moderate performance in confirming and/or excluding the presence of spermatozoa in patients with NOA undergoing TESE; thus, it cannot serve as a stand-alone marker of persistent spermatogenesis.

In an attempt to investigate whether Inh-B is superior to FSH as a predictor of sperm retrieval, we performed a comparative analysis of the two hormones, in eight studies, where data were available. FSH demonstrated a sensitivity of 0.71 and a specificity of 0.62 as compared with 0.65 and 0.85, respectively, for Inh-B. Thus, despite the similar overall diagnostic performance of the two hormones (sROC-AUC 0.70 versus 0.74 for both of them), Inh-B might be better in correct identification of men with NOA and no spermatogenesis (higher specificity) whereas FSH is better in correct identification of men with NOA and foci of spermatogenesis (higher sensitivity). In fact, a single sROC-AUC value may not be a suitable basis for clinical interpretation and extrapolation (Walter, 2005). Further studies are needed aiming to specifically investigate whether the combined application of FSH and Inh-B serum values are better in predicting sperm retrieval than each hormone alone.

Even in the case that all available parameters (testicular size, FSH, Inh-B) suggest low probability of successful sperm retrieval, the majority of men, at least in our practice, will decide to undergo a TESE procedure in an attempt to perform ICSI using their own genetic material. For this reason, we have applied a Fagan nomogram in order to counsel these men according to the principles of evidence-based medicine. In the studies included in the present meta-analysis the pre-test probability of successful sperm retrieval was 41%. In case of favorable (i.e. high) serum Inh-B levels, this probability increases to 73% whereas, in the opposite case, it decreases to 23%. In case of the later, an alternative option, such as donor sperm, could be available.

In conclusion, serum Inh-B cannot serve as a stand-alone marker of persistent spermatogenesis in men with NOA. Better prediction could be achieved from a multivariate model that will include not only all established parameters but also other hormones of testicular microenvironment that do not currently constitute part of the routine investigation of subfertile men.

**Discussion**

The aim of the present study was to estimate the predictive value of serum and seminal Inh-B and AMH as non-invasive markers of persistent spermatogenesis in men with NOA through a systematic review of diagnostic accuracy studies published in the literature and a meta-analysis of the best evidence available. The main finding was that serum Inh-B has a sensitivity of 0.65 and a specificity of 0.83 for the prediction of the presence of sperm in TESE. No clear conclusions can be deduced for seminal Inh-B and serum/seminal AMH due to the small number of relative studies. Although limited, evidence does not support the diagnostic value of neither seminal Inh-B, nor serum/seminal AMH. The latter is consistent with the findings of a recent systematic review (La Marca et al., 2010).

The bivariate model was used for the meta-analysis of pairs of sensitivity and specificity. This approach was preferred over the separate pooling of sensitivity and specificity or the summary ROC approach to avoid the limitations of these two methods. The former analysis ignores the negative correlation between sensitivity and specificity (Rutter and Gatsonis, 2001). The latter, using a single estimator of diagnostic accuracy, that is the diagnostic odds ratio, fails to distinguish between the ability of identifying the patient (sensitivity) and the healthy (specificity), which is important in determining the optimal use of a test in clinical practice (Reitsma et al., 2005). The results of the present meta-analysis have to be interpreted with caution. Many studies (Westlander et al., 2003; Bettella et al., 2005; Koga et al., 2007; Nowroozi et al., 2008; Ferhi et al., 2009) preferred to investigate their research hypothesis by simply comparing levels of Inh-B between patients with NOA and controls; in case a significant difference was not detected, they did not proceed to a binary classification test, that could provide extractable evidence and, therefore, they were excluded from the present meta-analysis. Since there are plausible reasons to believe that this subset of studies were not a random sample of the total, but, predominantly, those with non-significant results, the possibility of an overestimation in the diagnostic accuracy of Inh-B, as assessed by the present meta-analysis, should be taken into consideration. In fact, a descriptive analysis of the studies excluded from the meta-analysis suggests that serum Inh-B is a rather non-satisfactory diagnostic marker, in line with the main findings of the meta-analysis. In any case, serum Inh-B demonstrates a moderate performance in confirming and/or excluding the presence of spermatozoa in patients with NOA undergoing TESE; thus, it cannot serve as a stand-alone marker of persistent spermatogenesis.

![Figure 3](image-url) Fagan’s (Bayesian) nomogram of serum inhibin B for the prediction of the presence of spermatozoa in patients with non-obstructive azoospermia.

**Acknowledgements**

We are very grateful to Prof. Dr. E. Nieschlag and Dr. S. von Eckardstein (Institute of Reproductive Medicine of the University, Münster, Germany) and Dr. V. Mitchell (Hospital A. Calmette, Lille, ...
France) who kindly provided additional data requested on their publication. We, also, thank Prof. Dr. C. Foresta (University of Padova, Italy), Prof. Dr. F.H. Comhaire (Gent University Hospital, Belgium) and Dr. Y.F. El Garem (Alexandria University, Egypt) and Dr. M. Isikoglu (Antalya IVF Center, Turkey) for their willingness to provide further information on their publications.

References


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;78:1195–1198.


Jensen TK, Andersson AM, Hjollund NH, Schelke T, Kolstad H, Giwercman A, Henriksen TB, Ernst E, Bonde JP, Olsen J et al. Inhibin B as a serum marker of spermatogenesis: correlation to differences in...


von Eckardstein S, Simoni M, Bergmann M, Weinbauer GF, Gassner P, Scheipers AG, Nieschlag E. Serum inhibin B in combination with serum follicle-stimulating hormone (FSH) is a more sensitive marker than serum FSH alone for impaired spermatogenesis in men, but cannot predict the presence of sperm in testicular tissue samples. J Clin Endocrinol Metab 1999;84:2496–2501.


