Sperm function assays and their predictive value for fertilization outcome in IVF therapy: a meta-analysis

Sergio Oehninger1,*, Daniel R.Franken4, Ellen Sayed2, Gerardo Barroso1 and Paul Kolm3

1The Jones Institute for Reproductive Medicine, Department of Obstetrics and Gynecology, 601 Colley Avenue, Norfolk, VA 23507, USA, 2Moorman Memorial Library, 3Bio-statistics, Eastern Virginia Medical School, and 4Reproductive Biology Unit, Department of Obstetrics and Gynecology, University of Stellenbosch, Tygerberg, Republic of South Africa

The prevalence of male infertility and the availability of new, highly successful therapeutic options make the testing of sperm functional competence mandatory. An objective, outcome-based examination of the validity of the currently available assays was performed based upon the results obtained from 2906 subjects evaluated in 34 prospectively designed, controlled studies. The aim was carried out through a meta-analytical approach that examined the predictive value of four categories of sperm functional assays: computer-aided sperm motion analysis (CASA); induced-acrosome reaction testing; sperm penetration assay (SPA); and sperm–zona pellucida binding assays for IVF outcome. Results demonstrated a high predictive power of the sperm–zona pellucida binding and the induced-acrosome reaction assays for fertilization outcome. On the other hand, the findings indicated a poor clinical value of the SPA as predictor of fertilization and a real need for standardization and further investigation of the potential clinical utility of CASA systems. This analysis points out to limitations of the current tests and the need for standardization of methodologies and provides objective evidence on which clinical management and future research can be based.

Key words: acrosome reaction/CASA/meta-analysis/sperm penetration assay/sperm–zona pellucida binding assays

TABLE OF CONTENTS

Introduction 160
Meta-analysis 161
Computer-aided sperm motion analysis 162
Acrosome reaction 163
Sperm penetration assay 163
Sperm–zona pellucida binding assays 164
Comparative predictive power of sperm function assays 164
References 166

Introduction

The clinical (andrological) investigation of the male partner in an infertile couple relies on a thorough history and physical examination. In addition, urological and endocrinological investigations should be implemented as needed. Nonetheless, the semen analysis still remains the cornerstone of the diagnostic management. We and others have been promoters of a sequential, multi-step diagnostic approach for the evaluation of the various structural, dynamic and functional sperm characteristics when abnormalities are found in the initial evaluation (Oehninger et al., 1991, 1997a; Amman and Hammerstedt, 1993). This diagnostic scheme should include: (i) assessment of the ‘basic’ semen analysis; and (ii) functional sperm testing. The ‘basic’ semen analysis performed by the infertility specialist should include the evaluation of physical semen characteristics (volume, pH, agglutination and viscosity), the assessment of sperm concentration, progressive motility and normal morphology, the examination of sperm vitality, presence of leukospermia and immature sperm cells, detection of antisperm antibodies and a bacteriological investigation (World Health Organization, 1992). The investigation of the total motile sperm fraction following a separation procedure should also be implemented in this step (Oehninger, 1995).

If abnormalities are found during the basic investigation, the work-up should progress to the examination of specific sperm functions. The two questions that immediately arise are: firstly, which sperm functions should be examined? and secondly, what validated tests are available?

Numerous assays have been proposed to assess the various sperm functions. The categories of assays that are usually considered include: (i) tests that examine defective sperm functions indirectly through the use of biochemical tests (i.e. measurement of the generation of reactive oxygen species or evidence of peroxidative damage, measurement of enzyme activities such as creatine phosphokinase and others); (ii) bioassays of gamete interaction (i.e. the heterologous zona-free hamster oocyte test and the homologous sperm–zona pellucida

*To whom correspondence should be addressed at The Jones Institute for Reproductive Medicine, Department of Obstetrics and Gynecology, 601 Colley Avenue, Norfolk, VA 23507, USA. Tel: +1 757 446 7116; Fax: +1 757 446 8998; e-mail: sergio@jones1.evms.edu
binding assays) and induced-acrosome reaction scoring; and (iii) computer-aided semen analysis (CASA) for the evaluation of sperm motion characteristics (Yanagimachi et al., 1976; Cross et al., 1986; Burkman et al., 1988; Liu et al., 1988; Aitken et al., 1989a,b, 1991; Mortimer, 1990; Mortimer et al., 1990; World Health Organization, 1992; Huszar et al., 1992; Huszar and Vigue, 1994; Consensus Workshop, ESHRE Andrology Special Interest Group, 1996). Different laboratories have highlighted the diagnostic power of a variety of such assays particularly related to the outcome of assisted reproductive technologies (ART). IVF and embryo transfer therapy, singularly, offers the advantage of making use of fertilization as a main outcome measure of successful sperm functional capacities. Embryo implantation potential and pregnancy are obviously the ultimate desired outcome but their multi-factorial nature makes it more difficult to use them as specific end points.

What is the overall significance of sperm functional testing and which test(s) should be chosen? The truth of the matter is that the ‘basic’ semen evaluation allows the diagnosis of ‘male infertility,’ usually without providing evidence for an aetiological or physiopathological origin (except in inflammatory-infectious or immunological cases). Furthermore, if no specific therapy is indicated (i.e. urological, hormonal, pharmacological or other) or these treatments have failed, if the diagnosis is ‘idiopathic' infertility, or if the degree of sperm abnormalities is severe enough to refer the couple to ART, there is a real need to determine the sperm functional capacity in order to direct treatment to intrauterine insemination (IUI), IVF or intracytoplasmic sperm injection (ICSI). The answer then is that there is a real need to assess sperm functional competence in the extended evaluation of the infertile man.

How often do clinicians involved in reproductive medicine care confront this situation? Infertility in general and the proportion of cases where a male factor can be identified as the cause (alone or in combination with female factors) are prevalent conditions (Hull et al., 1985; Mosher and Pratt, 1990; Wilcox and Mosher, 1993; Society for Assisted Reproductive technology/American Society for Reproductive Medicine, 1995; Centers for Disease Control and Prevention, 1997). Although diagnostic problems make it difficult to establish the extent of the male partner’s contribution with certainty, a number of studies suggest that male problems represent the commonest single defined cause of infertility (Irvine, 1998). In addition, currently available therapeutic options (both urological and ART) are extremely successful in aiding couples achieve conception. The answer to the question, therefore, is that sperm function examination using validated tests would be extremely valuable for an improved clinical management in reproductive medicine.

Evidence-based medicine has been defined as the judicious and conscientious use of current best evidence from medical care research for making medical decisions (Swets, 1988; Evidence-Based Medicine Working Group, 1992; Jaeschke et al., 1994; Sackett, 1995; Sackett et al., 1996; Collins, 1997; Schleselman, 1997). Evidence-based medicine has now been extensively used in gynaecology and in some fields of ART, but the discipline of andrology has been late in accepting the need for controlled trials (Evers, 1997; Nieschlag, 1997). Although various reports have recently assessed the value of applying evidence-based medicine to therapeutic modalities, to date, no studies have examined in depth the clinical and research evidence supporting the use of the most widely used sperm function tests. Moreover, only 11 meta-analyses comparing a diagnostic test against a concurrent reference standard were identified by extensive searching of the general medicine literature (Irwig et al., 1995). Consequently, evidence-based medicine and meta-analysis of diagnostic sperm function testing may have the potential to assist clinical-decision making in reproductive medicine.

The objective of the present study was to perform a meta-analysis to determine the diagnostic test accuracy and predictive value of various sperm function assays for IVF outcome. There are numerous tests presently being used to assess different sperm functions. Here, we elected to define the accuracy of four categories of such tests using a meta-analytic approach as a guide to the routine use of these techniques in the ART discipline. The tests scrutinized were CASA, acrosome reaction testing, the zona-free hamster egg penetration test or sperm penetration assay (SPA) and sperm–zona pellucida binding assays. The selection of these assays for the meta-analysis does not negate the potential significance of the evaluation of other sperm functional capacities.

Meta-analysis

To find the relevant studies we conducted a computerized search using MEDLINE on Silver Platter CD-ROM. We searched the English language literature from January 1983 to December 1997 using various MeSH (medical subject) headings. The search process evolved in three steps: (i) the initial search was conducted using primarily controlled vocabulary; (ii) the search strategy was modified to a combination of text words and controlled vocabulary; and (iii) the researcher then conducted a manual review of bibliographies based on his knowledge of the subject area (expanded search) (Schleselman, 1997; Oxman et al., 1993, 1994). This approach brought up several references not retrieved in the initial MEDLINE search (Dickerson et al., 1994; Council, 1997).

A variety of exploded MeSH subjects were used. For test identification they were: for CASA, computer-aided sperm motion analysis, hyperactivated motility and cross-headings; for acrosome reaction: acrosome reaction, spontaneous or induced, and cross-headings; for SPA: zona-free hamster egg penetration assay, sperm penetration assay and cross-headings; and for sperm–zona pellucida binding assays: sperm–zona pellucida binding test and hemizona assay (HZA). Sensitivity, specificity, regression analysis, likelihood ratios and probability were the headings used for selection of studies addressing the predictive value of the tests. IVF was the heading used to identify the main outcome measure.

The statistical methods employed in the primary studies included Spearman’s rank order correlation between IVF and test parameters, logistic regression of fertilization (defined as ‘good’ or ‘poor’ based on various cut-off values) on tests parameters, or multiple linear regression if fertilization was analysed as continuous. In addition, predictive statistics or sensitivity, specificity and predictive values (positive = PPV and negative = NPV) were reported in some studies as well as odds ratios (ORs). In many cases, even if a study did not report the statistics the other
studies did, an adequate analysis could be calculated from data given.

Here, the meta-analysis was carried out combining correlation coefficients (Altman, 1991; Fleiss, 1993; Collins, 1997) and computing summary ORs with 95% confidence intervals (CI), and summary receiving operating characteristics (ROC) curves (Irwig et al., 1995). Computation of partial areas under the ROC curve followed a previously described method (McClish, 1989).

Table I shows the number of references addressing the predictive value of the four sperm function tests for IVF outcome as retrieved from the MEDLINE and ‘expanded’ searches and the number of references that were included in the meta-analysis. There were 47 and 75 references obtained from the MEDLINE and ‘expanded’ searches respectively. Adequate data could be retrieved from 34 references (half of them retrieved by the ‘expanded’ search) in order to perform the meta-analysis (i.e. raw data on test outcome and fertilization results were directly provided or could be abstracted from 2 × 2 tables). Studies not providing those data could not be combined for further analysis and were not included in the analysis. Altogether, we were able to analyse the predictive value of the individual tests for IVF outcome in a total of 2906 cycles.

Computer-aided sperm motion analysis

There were only four studies that addressed the relationship between IVF rates and data from CASA systems (Check et al., 1990; Liu et al., 1991; Sukcharoen et al., 1995, 1996). These studies included 289 patients (IVF cycles). Overall, there were inconsistencies among the studies, including use of various sperm preparation techniques (i.e. use of liquefied semen or separation of the motile sperm fraction), incubation time, CASA equipment and parameter settings of the computers, and analysis of fertilization cut-off levels.

In the initial analysis, the four studies were combined and examined with a Spearman rank-order correlation of fertilization with the various sperm motion parameters. Table II presents the correlations reported in the studies or those that were calculated

### Table I. Number of references retrieved from searches and used in the meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of references addressing the predictive value for IVF outcome</th>
<th>No. of references included in the meta-analysis</th>
<th>No. of patients analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline search</td>
<td>Expanded search</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASA</td>
<td>18</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Acrosome reaction</td>
<td>9</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>SPA</td>
<td>16</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>Sperm–zona pellucida binding tests</td>
<td>4</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

CASA = computer-aided sperm motion analysis; SPA = sperm penetration assay.

### Table II. Computer-aided sperm motion analysis (CASA) and prediction of IVF outcome

<table>
<thead>
<tr>
<th>Study</th>
<th>Liu et al., 1991</th>
<th>Suukcharoen et al., 1996</th>
<th>Suukcharoen et al., 1995</th>
<th>Check et al., 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sperm parameter</td>
<td>r</td>
<td>OR</td>
<td>r</td>
<td>OR</td>
</tr>
<tr>
<td>VCL</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>VSL</td>
<td>0.22</td>
<td>1.0</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>VAP</td>
<td>NR</td>
<td>NS</td>
<td>NS</td>
<td>1.1</td>
</tr>
<tr>
<td>LIN</td>
<td>0.37</td>
<td>1.0</td>
<td>–0.26</td>
<td>0.9</td>
</tr>
<tr>
<td>ALH</td>
<td>NS</td>
<td>NS</td>
<td>0.24</td>
<td>NS</td>
</tr>
<tr>
<td>BCF</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>HA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

NS = correlation was computed but was not statistically different from zero; NR = correlation was probably performed in the study but not reported; – = correlation was not done in the particular study; r = Spearman’s correlation coefficient; OR = odds ratio; VCL = curvilinear velocity; VSL = straight line velocity; VAP = average path velocity; LIN = linearity; ALH = amplitude of lateral head displacement; BCF = beat-cross frequency; HA = hyperactive motility.
from other information given (i.e. a contingency table of rates and parameter values).

Subsequently, we calculated an OR for the influence of the particular variable motion parameter (adjusted for all other variables in the equation) on fertilization. Three of the primary studies performed multiple logistic regression for data analysis. Exponential plots of the regression coefficient were, therefore, used to yield the estimated OR for the influence of the particular variable on fertilization. For the fourth primary study, which did not perform multiple regression analysis, 2 tables were constructed and used to compute the OR. The combined results of all studies demonstrated a large degree of variability indicating a poor predictive power for sperm parameters assessed by CASA and IVF results (Table II). Predictive statistics demonstrated low specificity and sensitivity and a high rate of false positives (data not shown).

**Acrosome reaction**

There were eight studies that addressed the relationship between the induced-acrosome reaction results and IVF outcome and sperm–zona pellucida binding assays for IVF outcome. Five of the primary studies presented results in the form of a Spearman’s rank order or Pearson correlation of acrosome reaction and fertilization rates. Table III shows the correlations for each of the five studies that could be analysed directly. The estimated overall correlation was 0.458 (95% CI = 0.455–0.462, P < 0.0001). However, the test for homogeneity of correlations across studies was significant (P < 0.0001), indicating that correlations were not homogeneous across studies.

Predictive statistics could be analysed for seven of the eight primary studies (see Table III). With one exception, PPV were ≥75%; the NPV tended to be more variable. Those seven studies could be combined to construct a summary ROC curve (Figure 1). The partial area under the curve was 32 and 82% when scaled to the total area. A sensitivity of 80% was achieved with a little >20% false positive rate. The slope of the regression line of log OR on the sum of the logits of true and false positive rates was 0.0487 (P = 0.895) suggesting that test accuracy could be summarized by a common OR. The common OR was 13.97 (95% CI = 2.91–67.14). The ORs of the studies ranged from 3 to 451.

**Sperm penetration assay**

There were 12 studies that addressed the relationship between the SPA results and IVF outcome in a total of 842 subjects (Margalioth et al., 1983, 1986; Wolf et al., 1983; Foreman et al., 1984; Aitken et al., 1985; Aitken et al., 1987; Kruger et al., 1988; Coetzee et al., 1989; Ibrahim et al., 1989; Nahhas and Blumenfeld, 1989; McClure et al., 1990; Soffer et al., 1992). There were some inconsistencies among studies regarding capacitation conditions, gamete co-incubation times, use or not of conditions to induce acrosome reaction and cut-off levels selected for data analysis. The above-mentioned reports either provided adequate data or the counts of 2 × 2 tables for calculation of diagnostic accuracy statistics. Of those studies not reporting individual data, three used 20% (or a close value of 17%) as the SPA cut-off value. All other studies used 10% as the cut-off value. For those studies reporting individual data, the necessary cell counts for 2 × 2 tables could be determined for either cut-off. In addition, ‘good’ fertilization was defined differently. Some studies defined fertilization as any value >0% but others defined it as ≥50%. Two additional studies provided breakdowns to define fertilization either way.

Table IV presents correlations and diagnostic statistics for all the 12 primary studies. The estimated overall correlation for the studies was 0.396 (95% CI = 0.393–0.398, P < 0.0001). However, the test for homogeneity of correlations was significant (P < 0.0001) indicating that correlations were not homogeneous across studies. Further analysis indicated two homogeneous subsets, the largest one composed of nine studies (r = 0.458, 95% CI = 0.456–0.461, P < 0.0001) and the other one composed of three studies (r = 0.079, 95% CI = 0.066–0.092, P < 0.0001). With one exception, the PPV were all >70% (range = 50–96%). However, the false negative rate was generally high (low specificity). The slope of the regression line of the log of the ORs on the sum of the logits of true and false positive rates was 0.0954 (P = 0.712) suggesting that test accuracy could be summarized by a common OR. The common OR for all studies was 7.61 (95% CI = 2.86–20.26). The ORs of the studies ranged from 1 to 42. The summary ROC curve for all studies is plotted in Figure 1. The partial area under the curve was 53% and 56% when scaled to the

![Figure 1](image-url)
A pre-subsets. The largest subset included seven HZA studies (across studies. Further analysis indicated three homogeneous (the test for homogeneity of correlations was significant \(r\) studies carried out with the sperm–zona binding assay (≥Table V), PPV were

For eight studies that could be combined for predictive statistics of a single HZA study with an unusually high correlation of 0.96. \(P\) CI = 0.465–0.470, generally >70%. Importantly, the false negative rate was

such a way to construct a 2 \(×\) such two reports (out of the 10 studies) that did not analyse their data in (categorized as ‘good’ or ‘poor’) could be constructed. There were

meta-analysis, there were eight from which a 2 \(×\) evaluated a total of 978 male patients (587 in the HZA and 323 in
care for the sperm–zona binding test) (Liu et al., 1988, 1989; Oehninger et al., 1989, 1992a,b; Franken et al., 1993a,b; Coddington et al., 1994; Gamzu et al., 1994; Liu and Baker, 1994). Those studies prospectively

evaluated a total of 978 male patients (587 in the HZA and 323 in the sperm–zona binding test). Of the 10 studies selected for the meta-analysis, there were eight from which a 2 \(×\) 2 table of sperm–zona binding cut-off values (either hemizona index for the HZA or binding ratio for the sperm–zona binding test) and fertilization (categorized as ‘good’ or ‘poor’) could be constructed. There were two reports (out of the 10 studies) that did not analyse their data in such a way to construct a 2 \(×\) 2 table. The latter two studies correlated the sperm–zona binding ratio with fertilization rate by the Spearman’s rank order correlation. A Spearman correlation can be calculated from a 2 \(×\) 2 table and, therefore, a meta-analysis of this statistic could also be performed for all 10 studies.

Table V presents the correlations and diagnostic statistics for all the 10 primary studies. The estimated overall correlation for the studies was 0.641 (95% CI = 0.64–0.642, \(P < 0.0001\)). However, the test for homogeneity of correlations was significant \((P < 0.0001)\) indicating that correlations were not homogeneous across studies. Further analysis indicated three homogeneous subsets. The largest subset included seven HZA studies \((r = 0.643, CI = 0.64–0.645, P < 0.00001)\); the second one included two studies carried out with the sperm–zona binding assay \((r = 0.470, CI = 0.465–0.470, P < 0.0001)\); and the third one was composed of a single HZA study with an unusually high correlation of 0.96. For eight studies that could be combined for predictive statistics (see Table V), PPV were ≥80% (range = 79–95%) and NPV were generally >70%. Importantly, the false negative rate was consistently low, ranging from 2 to 25%. The slope of the regression line of the log of the ORs on the sum of the logits of true and false positive rates was 0.6647 \((P = 0.192)\). The common OR of the eight studies was 23.68 although with a large 95% CI (4.83–115.99). A summary ROC curve could be constructed for those eight studies and is plotted in Figure 1. The partial area under the curve was 25 and 85% when scaled to the total area.

### Sperm–zona pellucida binding assays

There were 10 studies that investigated the relationship between sperm–zona pellucida binding assays and fertilization \(in vitro\) (seven of them using the HZA and three of them the sperm–zona binding test) (Liu et al., 1988, 1989; Oehninger et al., 1989, 1992a,b; Franken et al., 1993a,b; Coddington et al., 1994; Gamzu et al., 1994; Liu and Baker, 1994). Those studies prospectively evaluated a total of 978 male patients (587 in the HZA and 323 in the sperm–zona binding test). Of the 10 studies selected for the meta-analysis, there were eight from which a 2 \(×\) 2 table of sperm–zona binding cut-off values (either hemizona index for the HZA or binding ratio for the sperm–zona binding test) and fertilization (categorized as ‘good’ or ‘poor’) could be constructed. There were two reports (out of the 10 studies) that did not analyse their data in such a way to construct a 2 \(×\) 2 table. The latter two studies correlated the sperm–zona binding ratio with fertilization rate by the Spearman’s rank order correlation. A Spearman correlation can be calculated from a 2 \(×\) 2 table and, therefore, a meta-analysis of this statistic could also be performed for all 10 studies.

Table V presents the correlations and diagnostic statistics for all the 10 primary studies. The estimated overall correlation for the studies was 0.641 (95% CI = 0.64–0.642, \(P < 0.0001\)). However, the test for homogeneity of correlations was significant \((P < 0.0001)\) indicating that correlations were not homogeneous across studies. Further analysis indicated three homogeneous subsets. The largest subset included seven HZA studies \((r = 0.643, CI = 0.64–0.645, P < 0.00001)\); the second one included two studies carried out with the sperm–zona binding assay \((r = 0.470, CI = 0.465–0.470, P < 0.0001)\); and the third one was composed of a single HZA study with an unusually high correlation of 0.96. For eight studies that could be combined for predictive statistics (see Table V), PPV were ≥80% (range = 79–95%) and NPV were generally >70%. Importantly, the false negative rate was consistently low, ranging from 2 to 25%. The slope of the regression line of the log of the ORs on the sum of the logits of true and false positive rates was 0.6647 \((P = 0.192)\). The common OR of the eight studies was 23.68 although with a large 95% CI (4.83–115.99). A summary ROC curve could be constructed for those eight studies and is plotted in Figure 1. The partial area under the curve was 25 and 85% when scaled to the total area.

### Comparative predictive power of sperm function assays

The present meta-analysis was based on the results of 34 different studies independently performed by 25 centres in various geographical locations throughout the world in a total of 2906 subjects. These were all prospectively performed studies, with appropriate internal controls, addressing the predictive power of a defined functional test for fertilization outcome (outcome-based clinical research). Overall correlations and/or predictive statistics were performed for all the 34 primary studies. It must be acknowledged, however, that there were some inconsistencies in the primary studies as related to methodologies, patient populations analysed and clinical thresholds selected for data analysis. Some of these factors may introduce biases into the meta-analysis. Because of lack of consistency and/or data presentation among some reports, subsets of studies were reanalysed to obtain more accurate ORs and ROC curves (for a total of 27 studies out of the 30 primary studies exclusive of CASA analysis). Although the meta-analytical approach may have inherent imperfections (e.g. potential problems with the selected primary studies, the issue of combinability of data sets currently available, lack of literature reporting on negative results, discrepancies between meta-analyses and subsequent large randomized, controlled trials), it is still a very useful method to gain objective evidence on the predictive power of diagnostic assays and efficacy of therapeutic trials (Collins, 1997; L’Lorier, 1997; Comhaire, 1998). In spite of these limitations, several important conclusions can be drawn from the present meta-analysis.
Results clearly demonstrate the high predictive power of the sperm–zona pellucida binding and induced-acrosome reaction assays for IVF outcome. The sperm–zona pellucida binding assays and the acrosome reaction testing were both comparatively better predictors of fertilization than the SPA. The areas under the summary ROC curves scaled to the total curve were 85 and 82%, (OR of 23.6 and 13.97) for the zona binding and acrosome reaction assays, respectively. This compared to only 56% for the SPA (OR of 7.61) which was significantly lower than the area under the curve for the former assays \((P < 0.05)\) thus demonstrating the poor clinical value of the SPA as predictor of fertilization (unacceptably high false positive rate).

The interaction between spermatozoa and the zona pellucida is a critical event leading to fertilization and reflects multiple sperm functions (i.e. completion of capacitation as manifested by the ability to bind to the zona pellucida and to undergo ligand-induced acrosomal exocytosis) (Liu and Baker, 1992; Oehninger et al., 1987). Although different in their methodologies, they both use assessment of tight sperm binding to the zona as the primary endpoint in an independent comparison within an internally controlled assay. As proven here, such tests can therefore be immediately applied to the clinical management of infertile patients within the assisted reproduction setting. The high positive and negative predictive values, but more importantly, the low false negative rate (i.e. robust power to identify patients at high risk for fertilization failure) underscore the predictive ability of these tests. The only present limitation to the use of the sperm–zona pellucida binding assays is the need for a constant supply of human oocytes. Because the induced-acrosome reaction assays appear to be equally predictive of fertilization outcome and are simpler in their methodologies, the former tests could be favoured. However, prospective studies should be carried out to compare their predictive abilities in the same group of patients.

The induced-acrosome reaction testing, coupled with sperm vitality assessment, demonstrated an equal predictive power (similar area under the ROC curve) to that of the sperm–zona binding assays. It also has to be acknowledged that the acrosome reaction conditions and inducing agents varied for the studies (i.e. calcium ionophore, follicular fluid, temperature, capacitation conditions, etc). Consequently, more studies are needed for optimization and validation of such assays. The calcium ionophore induced-acrosome reaction is at the present time the most widely used methodology (Tesarik, 1989; Mortimer et al., 1990; Cummins et al., 1991). Nevertheless, the implementation of assays using small volumes of human solubilized zonae pellucidae (Franken et al., 1996), biologically active recombinant human ZP3 (Chapman and Barratt, 1996) or active, synthetic ZP3 peptides (or analogues) (Hinsch et al., 1998) will probably allow for the design of improved, physiologically-oriented acrosome reaction assays. The use of such agents combined with a better understanding of the biochemistry of the carbohydrate–protein interactions that take place during gamete recognition, binding and induction of acrosomal exocytosis will undoubtedly help in their elaboration.

The SPA, although a very valuable research tool, is proven to offer little help in the clinical setting due to its low predictive power. Another meta-analysis of the predictive value of this assay also reached a similar conclusion (Mol et al., 1998). Claims that the SPA predictability can be improved by modified versions of the assay (Aitken et al., 1987; Johnson et al., 1991) need to be corroborated by more studies. Of interest, the SPA results have been positively correlated with the outcome of spontaneous reproduction setting. The high positive and negative predictive values, but more importantly, the low false negative rate (i.e. robust power to identify patients at high risk for fertilization failure) underscore the predictive ability of these tests. The only present limitation to the use of the sperm–zona pellucida binding assays is the need for a constant supply of human oocytes. Because the induced-acrosome reaction assays appear to be equally predictive of fertilization outcome and are simpler in their methodologies, the former tests could be favoured. However, prospective studies should be carried out to compare their predictive abilities in the same group of patients.

The induced-acrosome reaction testing, coupled with sperm vitality assessment, demonstrated an equal predictive power (similar area under the ROC curve) to that of the sperm–zona binding assays. It also has to be acknowledged that the acrosome reaction conditions and inducing agents varied for the studies (i.e. calcium ionophore, follicular fluid, temperature, capacitation conditions, etc). Consequently, more studies are needed for optimization and validation of such assays. The calcium ionophore induced-acrosome reaction is at the present time the most widely used methodology (Tesarik, 1989; Mortimer et al., 1990; Cummins et al., 1991). Nevertheless, the implementation of assays using small volumes of human solubilized zonae pellucidae (Franken et al., 1996), biologically active recombinant human ZP3 (Chapman and Barratt, 1996) or active, synthetic ZP3 peptides (or analogues) (Hinsch et al., 1998) will probably allow for the design of improved, physiologically-oriented acrosome reaction assays. The use of such agents combined with a better understanding of the biochemistry of the carbohydrate–protein interactions that take place during gamete recognition, binding and induction of acrosomal exocytosis will undoubtedly help in their elaboration.

The SPA, although a very valuable research tool, is proven to offer little help in the clinical setting due to its low predictive power. Another meta-analysis of the predictive value of this assay also reached a similar conclusion (Mol et al., 1998). Claims that the SPA predictability can be improved by modified versions of the assay (Aitken et al., 1987; Johnson et al., 1991) need to be corroborated by more studies. Of interest, the SPA results have been positively correlated with the outcome of spontaneous

---

**Table IV. Sperm penetration assay (SPA) and IVF outcome**

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>r</th>
<th>Fertilization rate cut-off (%)</th>
<th>SPA cut-off (%)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolf et al., 1983</td>
<td>27</td>
<td>0.25</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>85</td>
<td>14</td>
<td>74</td>
<td>25</td>
</tr>
<tr>
<td>Margalioth et al., 1983</td>
<td>20</td>
<td>-0.08</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>100</td>
<td>0</td>
<td>50</td>
<td>–</td>
</tr>
<tr>
<td>Margalioth et al., 1986</td>
<td>134</td>
<td>0.56</td>
<td>&gt;0</td>
<td>&gt;20</td>
<td>94</td>
<td>57</td>
<td>85</td>
<td>78</td>
</tr>
<tr>
<td>Foreman et al., 1984</td>
<td>37</td>
<td>0.47</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>86</td>
<td>53</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Ausmanas et al., 1985</td>
<td>54</td>
<td>-0.03</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>84</td>
<td>33</td>
<td>96</td>
<td>11</td>
</tr>
<tr>
<td>Kruger et al., 1988</td>
<td>84</td>
<td>0.18</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>59</td>
<td>62</td>
<td>82</td>
<td>33</td>
</tr>
<tr>
<td>Coetzee et al., 1989</td>
<td>71</td>
<td>0.37</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>65</td>
<td>85</td>
<td>95</td>
<td>35</td>
</tr>
<tr>
<td>Ibrahim et al., 1989</td>
<td>59</td>
<td>0.54</td>
<td>&gt;0</td>
<td>&gt;17</td>
<td>84</td>
<td>73</td>
<td>90</td>
<td>61</td>
</tr>
<tr>
<td>McClure et al., 1990</td>
<td>19</td>
<td>0.68</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>93</td>
<td>75</td>
<td>93</td>
<td>75</td>
</tr>
<tr>
<td>Soffer et al., 1992</td>
<td>241</td>
<td>0.44</td>
<td>&gt;0</td>
<td>&gt;20</td>
<td>96</td>
<td>38</td>
<td>82</td>
<td>74</td>
</tr>
<tr>
<td>Nahhas et al., 1989</td>
<td>31</td>
<td>0.45</td>
<td>&gt;0</td>
<td>&gt;10</td>
<td>100</td>
<td>22</td>
<td>76</td>
<td>100</td>
</tr>
<tr>
<td>Aitken et al., 1987</td>
<td>65</td>
<td>0.28</td>
<td>&gt;50</td>
<td>&gt;10</td>
<td>85</td>
<td>41</td>
<td>80</td>
<td>50</td>
</tr>
</tbody>
</table>

\(n =\) number of cases; \(r =\) Spearman rank order or Pearson correlation; PPV = positive predictive value; NPV = negative predictive value.
pregnancy and conceptions following other interventions (Corson et al., 1988; Gwatkin et al., 1990).

More work is needed to define and validate the use of CASA systems for sperm motion analysis. Firstly, there were too few studies related to the use of sperm motion analysis and prediction of IVF outcome to reach general conclusions. Secondly, results demonstrated the lack of uniform criteria applied by different laboratories. Recently, guidelines have been proposed in order to standardize these methodologies (ESHRE Andrology Special Interest Group, 1998). We adhere to such recommendations as related to equipment, computer parameter settings, semen preparation techniques and overall criteria established for clinical use. It is expected that new studies will soon be available reporting on the clinical application of these systems.

Ideally, sperm function assays should sequentially examine the various dynamic properties of the spermatozoon. These include: (i) maturation and capacitation status; (ii) interaction with the female tract components; (iii) interaction with the oocyte vestments; and (iv) interaction with the ooplasm, oocyte activation and contribution to early embryogenesis. (Sharpe, 1992; Oehninger et al., 1992a, 1997a; Amman and Hammerstedt, 1993; Fraser, 1995; Aitken, 1997). Obviously, no single test will be able to assess those different and complex properties. The sequential, multi-step analysis of the main sperm functions examined by a combination of tests may prove to be clinically applicable (Oehninger et al., 1992a; Amman and Hammerstedt, 1993). For a more universal application of such tests, we need to have well defined methodologies, establish a defined outcome measure (i.e. fertilization in vitro or pregnancy in vivo, use in fertility evaluation, contraception or reproductive toxicology studies), have a common definition of male infertility and establish whether the tests will be applied to whole semen or to a selected sperm population (Jeyendran and Zaneveld, 1993; Cummins and Jequier, 1994; Mortimer, 1994).

It is expected that imminent advances in the cellular and molecular aspects of human gamete physiology and in biotechnology may help us develop rational, accurate and more predictive assays. At present, clinicians are usually forced to make decisions based on non-standardized techniques and imperfect studies. Or even worse, patients may be directed to efficient therapies such as ICSI without any knowledge about a physiopathological diagnosis. The present meta-analysis provides objective evidence in which clinical management and recommendations can be based. It also points out to the limitations of the current tests and can serve as an important guide to direct future research.

Table V. Sperm–zona pellucida binding assays and IVF outcome

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>r</th>
<th>Fertilization rate cut-off (%)</th>
<th>HZA cut-off (%)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oehninger et al., 1997</td>
<td>196</td>
<td>0.69</td>
<td>&gt;50</td>
<td>30</td>
<td>93</td>
<td>73</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td>Franken et al., 1993a</td>
<td>112</td>
<td>0.55</td>
<td>&gt;50</td>
<td>30</td>
<td>84</td>
<td>72</td>
<td>85</td>
<td>70</td>
</tr>
<tr>
<td>Franken et al., 1993b</td>
<td>48</td>
<td>0.50</td>
<td>&gt;50</td>
<td>30</td>
<td>75</td>
<td>68</td>
<td>81</td>
<td>68</td>
</tr>
<tr>
<td>Oehninger et al., 1989</td>
<td>28</td>
<td>0.79</td>
<td>&gt;65</td>
<td>36</td>
<td>95</td>
<td>83</td>
<td>95</td>
<td>83</td>
</tr>
<tr>
<td>Oehninger et al., 1992</td>
<td>44</td>
<td>0.70</td>
<td>&gt;65</td>
<td>36</td>
<td>100</td>
<td>61</td>
<td>79</td>
<td>100</td>
</tr>
<tr>
<td>Gamzu et al., 1994</td>
<td>133</td>
<td>0.96</td>
<td>&gt;0</td>
<td>23</td>
<td>100</td>
<td>94</td>
<td>85</td>
<td>100</td>
</tr>
<tr>
<td>Coddington et al., 1994</td>
<td>94</td>
<td>0.59</td>
<td>&gt;0</td>
<td>15</td>
<td>82</td>
<td>78</td>
<td>88</td>
<td>69</td>
</tr>
<tr>
<td>Liu et al., 1988</td>
<td>20</td>
<td>0.80</td>
<td>50</td>
<td>–</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Liu et al., 1989</td>
<td>106</td>
<td>0.50</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Liu and Baker, 1994</td>
<td>197</td>
<td>0.45</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

n = number of cases; r = Spearman rank order or Pearson correlation; PPV = positive predictive value; NPV = negative predictive value.

References


