Human Reproduction Vol.18, No.5 pp. 1037–1046, 2003 DOI: 10.1093/humrep/deg209

Accuracy of sperm–cervical mucus penetration tests in evaluating sperm motility in semen: a systematic quantitative review

Bolarinde Ola[1,3,4], Masoud Afnan[1], Spyros Papaioannou[1], Khaldoun Sharif[1], Lars Björndahl[1] and Aravinthan Coomarasamy[2]

[1]The Assisted Conception Unit and [2]Education Resource Centre, Birmingham Women's Hospital, Birmingham B15 2TG and [3]Division of Reproductive and Child Health, The University of Birmingham, Birmingham B15 2TH, UK

[4]To whom correspondence should be addressed at: Obstetrics and Gynaecology, The Jessop Wing, Sheffield Teaching Hospital NHS Trust, Tree Root Walk, Sheffield SI0 2SF, UK. E-mail: BolarindeOla@aol.com

BACKGROUND: Our objective was to determine the accuracy of in-vitro sperm penetration into cervical mucus or substitutes in evaluating sperm motility in semen. METHODS: This was a systematic quantitative review of test accuracy studies. The Cochrane library (2000:4), Medline (1966–2001), Embase (1988–2001) and SciSearch (1981–2001) were searched, in addition to manual searches of conference papers and bibliographies of known primary and review articles. Primary studies measuring in-vitro sperm penetration into cervical mucus, or substitutes (i.e. sperm–mucus penetration test, SMPT) and comparing results with sperm motility in semen were included. RESULTS: There were 18 primary diagnostic studies published in 17 papers, involving a total of 2580 samples. Fourteen primary diagnostic tests used vanguard distance as diagnostic criteria ($SMPT_{vd}$) and the pooled likelihood ratio (LR) for positive (LR+) and negative (LR–) tests were 2.29 (1.82–2.87) and 0.52 (0.44–0.63) respectively. Four studies used diagnostic criteria based directly or indirectly on swim-up sperm count per high power field ($SMPT_{sc}$) instead. Their pooled LR+ and LR– were 5.24 (3.36–8.18) and 0.15 (0.06–0.39) respectively. CONCLUSIONS: $SMPT_{vd}$ has a low accuracy in the evaluation of sperm motility in semen. However, $SMPT_{sc}$ was found to be more accurate. This method of using sperm concentration, instead of vanguard distance, as diagnostic criteria of in-vitro SMPT has potential as a useful laboratory-based sperm function test.

Key words: accuracy/hyaluronic acid/likelihood ratio/methyl cellulose/sperm–mucus penetration tests

Introduction

Semen analysis is usually the first and most commonly performed test at infertility consultations. This test, which is currently the World Health Organization (1999) reference test, measures mainly traditional sperm parameters. Some have argued that semen analysis provides only limited prognostic information about fertility (Branigan *et al*., 1999; Tomlinson *et al*., 1999; Guzick *et al*., 2001). The ideal male fertility test should provide some assessment of sperm function as well as measurements of concentration, motility and morphology. However, the numerous assays and tests described so far, except the zona binding assay, have only limited value in assessing sperm function (ESHRE, 1996; Oehninger *et al*., 2000). As a result, many specialist andrology laboratories, in addition to the standard semen analysis, will perform some form of swim-up procedure as a test of sperm function, to further classify normal and abnormal sperm function (Mortimer, 2000). Furthermore, when preparing sperm for assisted conception in the embryology laboratory, it is routine to separate motile sperm from other cells and dead sperm, using either a basic layering and swim-up technique or density gradient centrifugation.

The in-vitro sperm–mucus penetration test (SMPT) is a sperm function test which measures the ability of sperm in the semen to swim up into a column of cervical mucus or substitute. If it can be proven to be as good as semen analysis in assessing progressive sperm motility, then arguably, its additional benefit as a test of functional competence may make it a suitable and cheaper alternative to the present combination of semen analysis and sperm separation procedures (see decision tree in Figure 1). This attraction has led to the re-emergence of research into SMPT for the assessment of semen. Sperm migration into cervical mucus, or any of its substitutes, is based on the same principle as the Kremer (1965) test. Some of the parameters used to reflect the degree of penetration include: the migration or vanguard distance (distance between the foremost sperm in the capillary tube to the end of the tube in the semen reservoir); swim-up sperm count or sperm velocity at 10, 20, 30

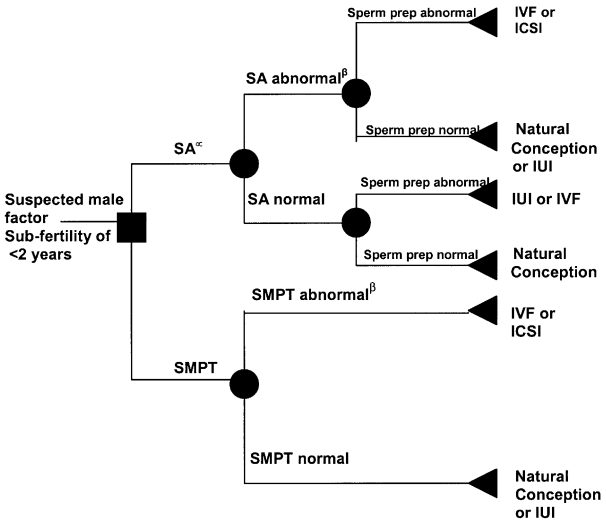

Figure 1. A simple decision tree to illustrate a proposed alternative pathway to advanced semen analysis in the evaluation of male sub-fertility. [a]SA (advanced semen analysis) implies semen analysis and together with one or more sperm procedures. [b]Abnormal SA or sperm–cervical mucus penetration test (SMPT) implies a diagnosis made after a second abnormal test. The fixed (non-consumable laboratory equipment) and variable costs (manpower hourly rates and disposable apparatuses required for SA + sperm preparation) is significantly more than the same costs for SMPT. IUI = intrauterine insemination.

or 40 mm at high power field and migration reduction (the decrease in migration density from 10 to 40 mm). Of these, the most commonly used parameter is the vanguard distance of 30 mm measured after incubation at 37°C for 90 min, although other workers have used different vanguard measurements (Ulstein *et al*., 1972; Matthews *et al*., 1980; Alexander, 1981). Three groups, David *et al*. (1979), Aitken *et al*. (1992) and Ivic *et al*. (2002), however, used swim-up sperm count per high power field at 10, 20 and 30 mm in a flat capillary tube as the diagnostic criteria for SMPT. Amit *et al*. (1982) used average swim-up sperm velocity, which is indirectly related to sperm count per high power field, at 10, 20 and 30 mm after 60 min.

The usefulness of the SMPT in infertility management has hitherto not been assessed rigorously enough for a variety of reasons. For example, many of the primary study designs aimed at measuring the accuracy of SMPT in the past did not facilitate measurements of sensitivity, specificity and likelihood ratios (LR). In others, the choice of a reference or gold standard was either not ideal, had confounding factors or the population studied was narrow, involving only sperm donors (Appendix A). Other diagnostic studies into the usefulness of the SMPT have used pregnancy rate as the gold standard because it is the outcome that infertile couples are ultimately interested in achieving. A disadvantage of using pregnancy rates is that many other co-factors other than semen quality contribute to produce a pregnancy outcome (Ford, 1999). These include numerous male and female factors and other co-variables such as duration of infertility and particularly the concept of time-to-event, which describes the confounding effect of the interval between laboratory tests on the semen and conception (Ducot *et al*., 1988; Dunphy *et al*., 1989; Eimers *et al*., 1994). As a result, there are no validated diagnostic threshold pregnancy or fertilization rates for normal/abnormal semen analysis. Consequently, some diagnostic fertility tests are, in reality, evaluated as prognostic interventions of pregnancy outcomes, but are reported as diagnostic studies. These reasons may explain why the majority of researchers preferred semen analysis as a reference standard in their validation of the diagnostic accuracy of SMPT.

This systematic review was conducted to answer the clinical question: is in-vitro sperm penetration into cervical mucus, or its substitutes, an accurate diagnostic tool in the evaluation of semen? Our aim was to conduct a systematic quantitative review of all primary diagnostic studies that compared SMPT with semen analysis in a manner that allowed meaningful assessment of test quality and usefulness. This objective was to be achieved by identifying the number, scope and quality of primary studies that measured the diagnostic accuracy of SMPT or sperm penetration through known cervical mucus substitutes, by comparing with a reference standard of progressive or total sperm motility in the seminal plasma, as measured manually or by computerized methods. To our knowledge, this is the first such quantitative systematic review of the diagnostic accuracy of SMPT.

Materials and methods

This review was based on a prospective protocol using robust methodology (Deeks, 2001; Khan *et al*., 2001b; Cochrane methods working group, 2000).

Identification of studies

We searched the general bibliographic databases: Medline (1966–2001) and Embase (1988–2001). We also searched the specialist computer databases of the Cochrane library (2000:4) and SciSearch (1981–2001). Furthermore, we conducted manual searching of conference papers of the American Society for Reproductive Medicine (1997–2000), the European Society for Human Reproduction and Embryology (1997–2000) and bibliographies of known primary and review articles.

The search strategy was based on the clinical question: can in-vitro sperm penetration of cervical mucus and mucus substitutes be used to evaluate the quality of sperm motility in the semen? The search terms (sperm or sperm; penetration or interaction or migration; cervical mucus or cervical mucus substitutes; sperm motility or sperm motion or computer-assisted image processing; fertility or infertility) used in all databases were designed to be initially very sensitive (include as many hits as possible) before they were combined using the Boolean 'and'. For example, the following search query was employed for electronic databases: sperm$.mp or exp spermatozoa/;((cerv$ adj mucus) or mucus$).mp or (exp mucus/ and substit$.mp); penetrat$.mp or interact$.mp or migrat$.mp; exp sperm motility/ or motil$.mp. or motility.mp or sperm motion.mp; exp image processing or computer-assisted/; exp fertility/ or exp infertility, male/.

Study selection

The inclusion of primary studies was conducted in two stages. The first involved scanning of titles and abstracts, which were selected provisionally unless clearly not comparing diagnostic accuracies of sperm penetration into cervical mucus or its substitutes with a stated reference standard. Articles that generated doubts regarding accuracy

were retained until the full text documents were retrieved. An account was made of the number of titles identified, retained and excluded in order to ensure reproducibility of this stage. The second stage involved final selection of papers from the list obtained in the first stage. The selection was based on a checklist of inclusion criteria. Studies conducted and reported in non-English language were interpreted and dealt with in the same way as papers written in English.

Inclusion criteria were satisfied if the population comprised fertile and infertile men and if the study tested sperm penetration or migration into cervical mucus or its alternatives. Furthermore, the reference standard required was the World Health Organization reference values for motile, or progressively motile sperm, according to the date of study. Finally, the accuracy measurements based on 2×2 table construction had to be possible either from the text, tables, scattergraph or receiver operator characteristics (ROC) curves. For all included studies, the following were recorded and double-checked: authors, year of publication, study design, setting, population characteristics, details of test and gold standard used. The study by Ivic *et al*. (2002), which has now been published, was included as unpublished data.

Quality assessment

Methodological quality was defined as the confidence with which the study design, conduct, and analyses minimized biases. Based on existing checklists, we performed quality assessment by scrutinizing study designs and relevant features of the population, test and reference standard (Cochrane working group, 2000; Khan *et al*., 2001b). We considered a study to be of good quality if it used a prospective design, consecutive enrolment, adequate test description, and blinding of the test result (Lijmer *et al*., 1999; Khan *et al*., 2001b). Where no explicit information was offered in the paper, this was categorized as unreported.

Diagnostic interventions had to be described in sufficient details to ensure reproducibility. Kremer-type tests using flat or round capillary tubes were considered ideal, but other well-described modifications were also acceptable. Human cervical mucus obtained around the time of natural and drug-induced ovulation, or hormone replacement cycles was ideal, but other well-justified cervical mucus substitutes such as hyaluronic acid or methyl cellulose were considered acceptable.

Despite the problems with interpretation due to its non-dichotomous outcome measures, and inter- and intra-observer errors, semen analysis remains the benchmark test for male infertility (World Health Organization, 1999). The accuracy of the SMPT was assessed in respect of its ability to correctly predict normal or abnormal sperm motility, whether measured by semen analysis or computer-aided sperm analyses (ESHRE, 1996). For the SMPT, any or the entire four outcome measures in common use was considered suitable.

Data abstraction and synthesis

We searched carefully for duplication of data that can result from multiple publications of part or whole research work. For clarity and objectivity, data sheets were used for abstraction, based on study characteristics and quality. To minimize errors, two reviewers (B.O. and A.C.) independently performed the data extraction and measured the degrees of agreement for inclusion and quality criteria. The overall agreement was 93.3%, with a Kappa value (κ) of 0.87 [95% confidence interval (CI) 0.69–1.04]. Kappa statistics assess agreement beyond chance and allow credit for partial accord (Landis and Koch, 1977).

Categorical data were summarized into a standard 2×2 table and where the outcome measures were continuous variables, 2×2 tables were derived from scattergram and on one occasion from the text and ROC plot. Statistical analyses and data synthesis were performed on

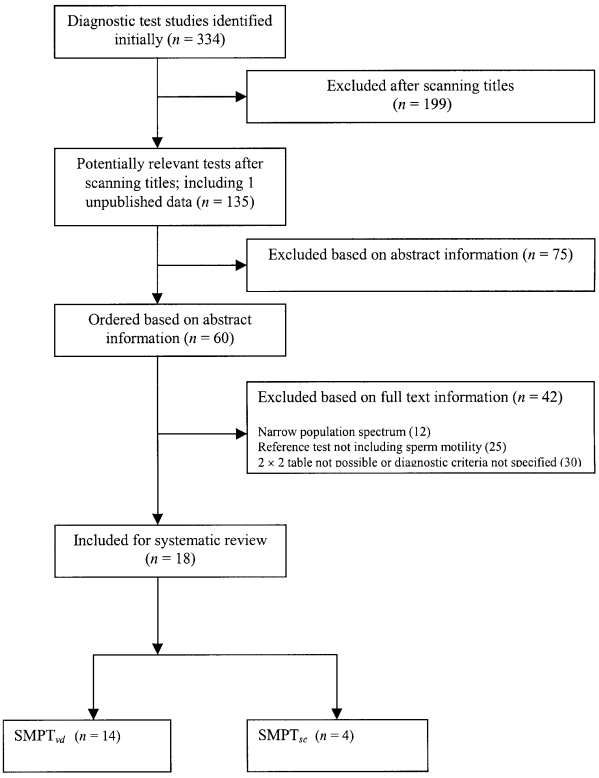

Figure 2. Study selection process showing history of included and excluded primary studies. SMPT$_{vd}$ = sperm–cervical mucus penetration tests where vanguard distance is the diagnostic criterion; SMPT$_{sc}$ = SMPT where sperm count is the diagnostic criterion.

Arcus Quickstat (Biomedical Version 1.0, 1997) and Stata Version 7.0 (Stata Corp., USA). Forest plots and pooled LR+ and LR– statistics were derived using the DerSimonian and Laird (1986) random effects models.

Heterogeneity was tested graphically (Forest plots) and statistically. In our review, although no significant statistical heterogeneity was detected, we found that heterogeneity existed clinically, which resulted in our decision to perform sub-group meta-analysis of the primary studies based on SMPT diagnostic criteria. The reference test had the same end point (sperm motility) but varied slightly in thresholds. We measured effect by pooled LR+ and LR– instead of summary ROC curves because these thresholds had been classified clearly in each primary study and such minor threshold variations do not significantly modify long-term prognosis (Dickey *et al*., 1999; Deeks, 2001).

Results

The results of our literature search are summarized in Figure 2. There were 18 primary diagnostic studies published in 17 papers, involving a total of 2580 samples. The general characteristics of included studies are summarized in Table I under the following categories: author and year of study, country where study was conducted, study design, sample size and whether drawn from a wide or narrow population, type of penetration test and the reference test. Similarly, the methodological qualities of included studies are summarized under the headings shown in Table II.

Table I. Characteristics of studies included in a systematic review of accuracy of sperm–cervical mucus penetration tests (SMPT)

Author and year of study	Country	Study design	Sample size	Population	Type of SMPT	Reference test
Alexander (1981)	USA	Case–control	161	Wide spectrum	S*b*MPT	SA; sperm motility
Aitken *et al*. (1992)	UK	Cohort	77	Wide spectrum	SHAPT (sperm select)	CASA; sperm motility
Amit *et al*. (1982)	Israel	Cohort	197	Wide spectrum	S*h*MPT	SA; sperm motility
David *et al*. (1979)	Israel	Cohort	200	Wide spectrum	S*h*MPT	SA; sperm motility
Eggert-Kruse *et al*. (1989a)	Germany	Cohort	288	Wide spectrum	S*h*MPT	SA; progressive motility
Ivic *et al.* (2002)	UK	Cohort	57	Wide spectrum	SMCPT	SA; progressive motility
Keel *et al*. (1987)	USA	Cohort	161	Wide spectrum	S*b*MPT	SA; sperm motility
Keel *et al*. (1988)	USA	Cohort	226	Wide spectrum	S*b*MPT	SA; sperm motility
Mangione *et al*. (1981)	USA	Cohort	299	Wide spectrum	S*b*MPT	SA; sperm motility
Marshburn *et al*. (1989)	USA	Cohort	86	Wide spectrum	S*b*MPT	SA; sperm motility
Matthews *et al*. (1980)	Australia	Cohort	79	Wide spectrum	S*h*MPT	SA; sperm motility
Moghissi *et al*. (1982)	USA	Case–control	75	Wide spectrum	S*b*MPT	SA; sperm motility
Mortimer *et al*. (1986)	UK	Cohort	100	Wide spectrum	S*h*MPT	SA; progressive motility
Schutte (1987)	Germany	Cohort	89	Wide spectrum	S*b*MPT	SA; progressive motility
Stumpf and Lloyd (1985)	USA	Case–control	15	Wide spectrum	S*b*MPT	SA; sperm motility
Takemoto *et al*. (1985)	USA	Cohort	136	Wide spectrum	S*b*MPT	SA; sperm motility
Ulstein (1972i)	Sweden	Cohort	51	Wide spectrum	S*h*MPT	SA; sperm motility
Ulstein (1972ii)	Sweden	Cohort	283	Wide spectrum	S*h*MPT	SA; sperm motility

Types of SMPT in this review: SMCPT = sperm–methylcellulose penetration test; S*b*MPT = sperm–bovine cervical mucus penetration test; S*h*MPT = sperm–human cervical mucus penetration test; SHAPT = sperm penetration into hyaluronic acid. CASA = computer-assisted sperm analysis.

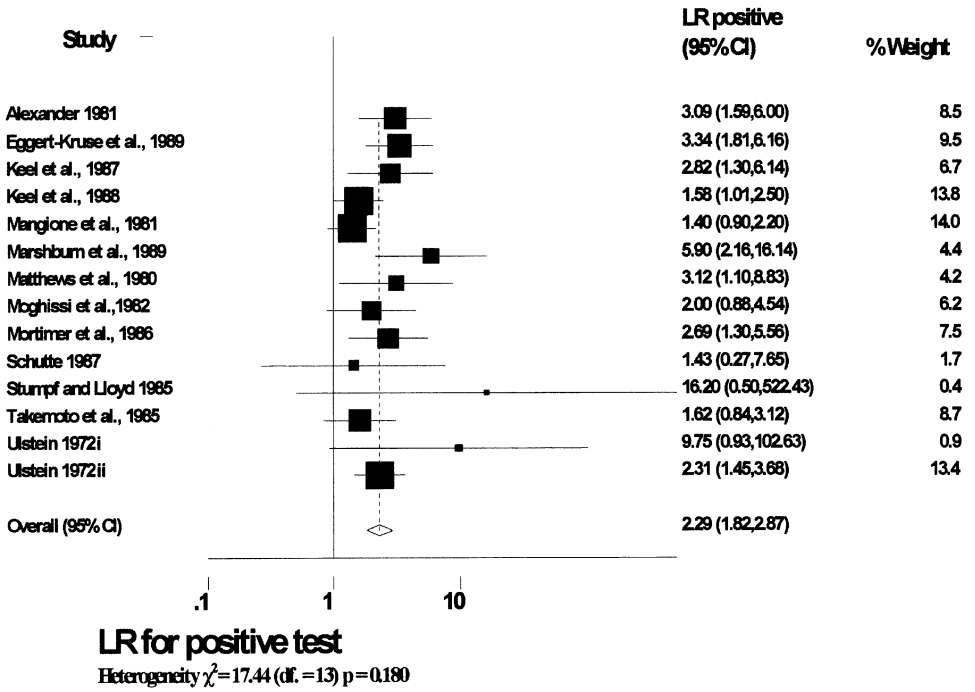

Figure 3. Individual and pooled estimates of likelihood ratio positive (LR+) for sperm–cervical mucus penetration tests where vanguard distance is the diagnostic criterion ($SMPT_{vd}$).

Quantitative data synthesis

Forest plots in Figure 3 and Figure 4 show the individual and overall pooled LR+ and LR– respectively when vanguard distance is used as diagnostic criteria in 14 primary studies ($SMPT_{vd}$). The pooled LR+ and LR– were 2.29 (95% CI 1.82–2.87) and 0.52 (0.44–0.63) respectively. There were four studies, however (David *et al*., 1979; Amit *et al*., 1982; Aitken *et al*., 1992; Ivic *et al*., 2002; see Figures 5 and 6), that used diagnostic criteria based on swim-up sperm count per high power field ($SMPT_{sc}$) instead of vanguard distance as the test threshold. Their pooled LR+ and LR– are 5.24 (3.36–8.18) and 0.15 (0.06–0.39) respectively.

Discussion

The LR+ and LR– of $SMPT_{vd}$ (2.3 and 0.5 respectively) imply that the accuracy of this method is low. However, LR+ and LR– of 5.2 and 0.15 respectively indicate that $SMPT_{sc}$ is moderately accurate in the evaluation of sperm motility in semen.

Table II. Methodological qualities of studies included in a systematic review of accuracy of sperm–cervical mucus penetration tests (SMPT)

Author and year of study	Study design	Patient recruitment	Type of SMPT and normal threshold	Description of reference test threshold	Reference standards used
Alexander (1981)	Case–control	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾15 mm after 90 min	Method described; SA; 50% motile sperm	SA and Pregnancy occurring within 6 months after S*b*MPT
Aitken *et al*. (1992)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	Sperm penetration into HA; count per high power field of swim-up sperm at 10 mm	Method described; CASA; 25% rapid sperm (mean path velocity >25 μm/s)	CASA and hamster oocyte penetration test
Amit *et al*. (1982)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*h*MPT; average swim-up sperm velocity at 10, 20 and 30 mm after 60 min	Method described; SA; 50% motile sperm	One reference standard used
David *et al*. (1979)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*h*MPT; total count per high power field of swim-up sperm at 10, 20 and 30 mm after 90 min	Method described; SA; 50% motile sperm	One reference standard used
Eggert-Kruse *et al*. (1989a)	Cohort	Prospective, random sampling	S*h*MPT; vanguard distance not stated; likely ⩾30 mm	Method not described; SA; 40% progressive motility	SA and HA
Ivic *et al*. (2002)	Cohort	Prospective, convenient sampling	SMCPT; count per high power field of swim-up sperm at 10 mm	Method described; SA; 50% progressive motility	SA and HA
Keel *et al*. (1987)	Cohort	Prospective, convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min	Method described; SA; 40% motile sperm	SA, post-coital test and pregnancy. Time to pregnancy was not stated
Keel *et al*. (1988)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min	Method described; SA; 40% motile sperm	One reference standard used
Mangione *et al*. (1981)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*h*MPT; vanguard distance ⩾20 mm after 30 min	Method not described; poor/good sperm classification	One reference standard used
Marshburn *et al*. (1989)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min	Method described; 60% motile sperm in swim-up supernatant	SA and pregnancy occurring within 6 months after S*b*MPT
Matthew *et al*. (1980)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*h*MPT; vanguard distance ⩾30 mm after 30 min; derived from scatter plot	Method described; SA; 50% motile sperm	One reference standard used
Moghissi *et al*. (1982)	Case–control	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min; derived from scatter plot	Method described; SA; 50% motile sperm	One reference standard used
Mortimer *et al*. (1986)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*h*MPT; vanguard distance; threshold not stated explicitly (normal/abnormal)	Method described; SA; 25×10^6/ml concentration progressive motility	One reference standard used
Schutte *et al*. (1987)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min	Method described; SA; 50% progressive motility	One reference standard used
Stumpf and Lloyd (1985)	Case–control	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min; derived from scatter plot	Method described; SA; 50% motile sperm	One reference standard used
Takemoto *et al*. (1985)	Cohort	Prospective, but cannot tell whether consecutive, random or convenient sampling	S*b*MPT; vanguard distance ⩾30 mm after 90 min	Method described; SA; 50% motile sperm	SA and zona-free hamster egg fertilization
Ulstein (1972i)	Cohort	Prospective, but cannot tell if consecutive, random or convenient sampling	S*h*MPT; vanguard distance ⩾20 mm after 180 min	Method described; SA; 50% motile sperm	One reference standard used
Ulstein (1972ii)	Cohort	Prospective, but cannot tell if consecutive, random or convenient sampling	S*h*MPT; vanguard distance ⩾20 mm after 180 min	Method described; SA; 50% motile sperm	One reference standard used

SA = semen analysis; CASA = computer-assisted sperm analysis; HA = hyaluronic acid.

The use of LR as measures of test usefulness have been well described (Radack *et al*., 1986; Sackett *et al*., 2000; Khan *et al*., 2001a;b). The LR for a positive test quantifies the probability that a truly positive, rather than a falsely positive, test result is to be seen when disease state, as defined by the reference test, is present. Similarly, the likelihood ratio for a negative test quantifies the probability that a false negative rather than a truly negative test result is presented when the reference test is negative. As a general guide to understanding the usefulness of a test, a LR+ >10 and LR– <0.1 generally indicate a highly useful test because they generate large and usually conclusive changes in the pre-test probability of disease. LR+ of 5–10 and LR– 0.1–0.2 indicate a moderately useful test, LR+ of 2–5 and LR– of 0.2–0.5 slightly useful, and LR+ of 1–2 and LR– 0.5–1 are considered indices of a non-valuable test (Sackett *et al*., 2000; Khan *et al*., 2001a). The pooled LR+ and LR– of 2.29

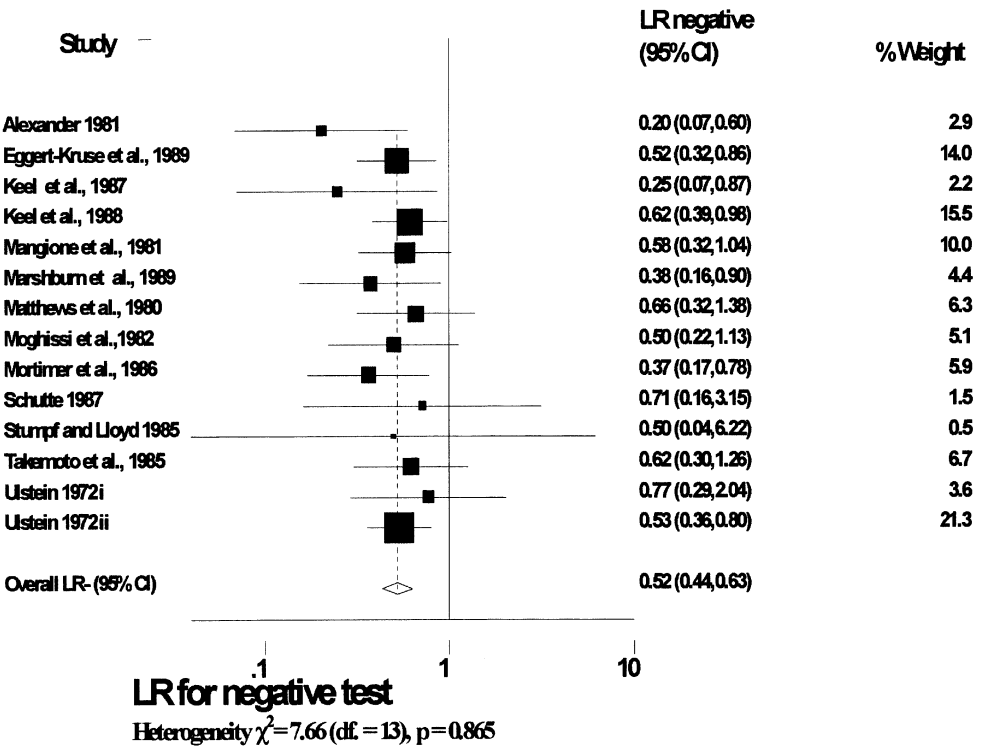

Figure 4. Individual and pooled estimates of likelihood ratio (LR–) negative for sperm–cervical mucus penetration tests where vanguard distance is the diagnostic criterion ($SMPT_{vd}$).

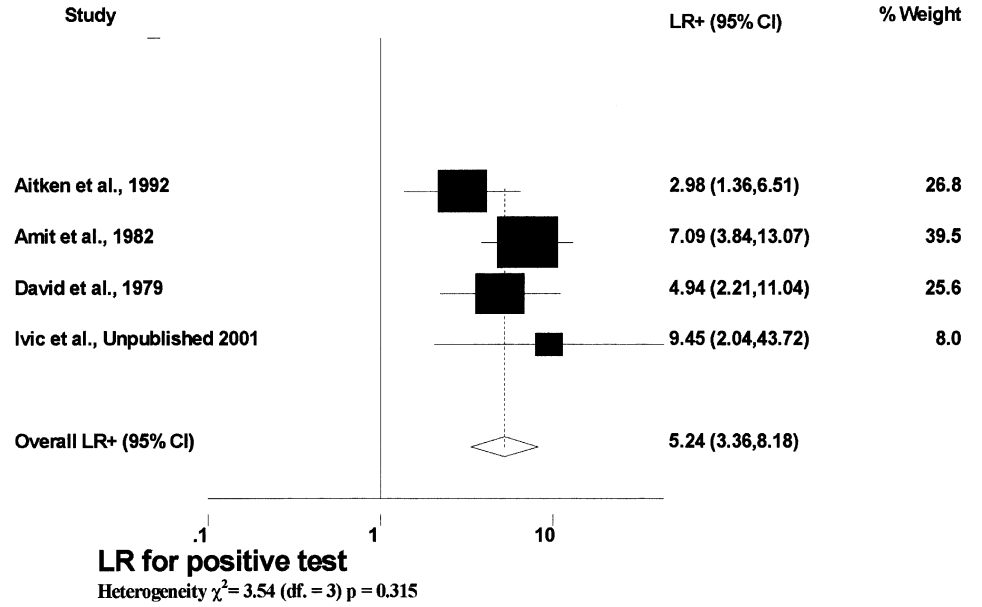

Figure 5. Individual and pooled estimates of likelihood ratio positive (LR+) for sperm–cervical mucus penetration tests where sperm count is the diagnostic criterion ($SMPT_{sc}$).

(1.82–2.87) and 0.52 (0.44–0.63) therefore show that the method of using vanguard distance as diagnostic criterion (i.e. $SMPT_{vd}$) has a low accuracy.

The studies by David *et al*. (1979), Amit *et al*. (1982), Aitken *et al*. (1992) and Ivic *et al*. (2002) are different from all others because they use the number of swim-up sperm per high power field at 10, 20 or 30 mm as the test diagnostic criterion (i.e. $SMPT_{sc}$). There was a substantial difference between the pooled results for these four primary studies; LR+ and LR– respectively are 5.24 (3.36–8.18) and 0.15 (0.06–0.39), and the other 14 studies in which vanguard distance was used. These findings suggest that swim-up sperm count per high power field is a more accurate diagnostic criterion than vanguard distance for SMPT.

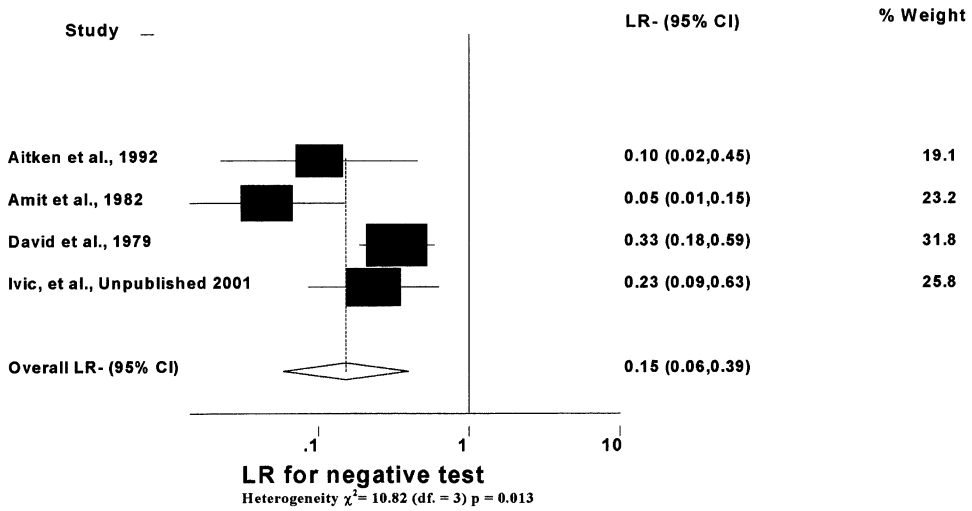

Figure 6. Individual and pooled estimates of likelihood ratio negative (LR–) for sperm mucus penetration tests where sperm count is the diagnostic criterion ($SMPT_{sc}$).

Likelihood ratios from primary studies using slightly different thresholds can be pooled if the end points had been classified clearly and such threshold variations do not significantly modify long-term prognosis (Deeks, 2001). Dickey *et al.* (1999) have shown that variations above the diagnostic thresholds of 5×10^6/ml for total sperm count, and 30% for progressive motility, produced only minimal and insignificant changes in fertility rates. Dickey's group studied 1841 couples undergoing 4056 cycles of intrauterine inseminations and demonstrated that there was very little difference in the effect on fertility rates between normal values of total and progressive motility. They went further to demonstrate that only when initial values of total motile sperm count were $<5\times10^6$/ml, and progressive motility <30%, was fertility significantly affected. In our review, only 6.9% of the samples in our study had a reference threshold of 25% progressive motility, therefore the only valid ground for sub-grouping was based on the SMPT diagnostic criteria (i.e. whether vanguard distance or swim-up sperm concentration).

Cervical mucus plays an important role in selecting motile, mostly morphologically normal sperm for fertilization. For this reason, the SMPT has always held a potential as an important in-vitro sperm function test. However, human cervical mucus (*h*CM), apart from being difficult to obtain in large quantities, possesses wide variations in viscosity (Karni, 1971; World Health Organization, 1987; Eggert-Kruse *et al.*, 1989a;b). As a result, cervical mucus substitutes in many forms are now commonly used in these tests. Bovine cervical mucus (*b*CM) is particularly common because it is easier to obtain in large quantities, has similar rheological and biochemical properties to *h*CM and can be stored in the frozen stage with only minimal changes in rheological properties (Meyer, 1977; Gaddum-Rosse *et al.*, 1980; Lee *et al.*, 1981). However, *b*CM has the disadvantage of between-batch variability in consistency. Hyaluronic acid (HA) has been described as a suitable alternative. It is a muco-polysaccharide with a structure similar to that of cervical mucus (Bothner and Wik, 1987) and a viscosity that can easily be varied to that described for *h*CM (Karni 1971; Ishijima *et al.*, 1986). More recently, however methylcellulose (MC) has been evaluated as an effective sperm swim-up medium, which is comparable with HA in efficacy (Ivic *et al.*, 2002). Most included primary studies tested sperm penetration of either *b*CM or *h*CM, using sperm motility or progressive motility as the reference standard. It was somewhat surprising that not many eligible primary studies were identified that tested artificial substitutes for cervical mucus such as HA and MC.

There were, however, early problems with accurate quantification of swim-up sperm concentration into capillary tubes and micropipettes with circular cross-sections (Kremer, 1965; Ulstein, 1973; Kerin *et al.*, 1976). This problem, well described by Kremer and Kroeks (1975) is the optical effect due to curvature of the circular walls, which rendered sperm underneath grossly distorted. It was partially solved with the subsequent introduction of generations of flat capillary tubes. The most popular for many years had external measurements of either $0.3\times3\times50$ mm (De Geyter *et al.*, 1988) or $0.3\times3\times100$ mm (Kotoulas *et al.*, 1984; Pandya *et al.*, 1986; Morrow *et al.*, 1992). However, loss of accuracy in counting within the sides of the flat tubes had continued to be problematic partly due to distortion and also because sperm tended to swim against the surface of the circular walls. This may have accounted partly for the continued adoption of vanguard distance, which is more easily measured, as SMPT threshold. More recently, the accuracy of quantifying sperm migration into cervical mucus substitutes in flat capillary tubes has been improved by newer designs with central counting chambers, away from the circular walls. One such design (Camlab Ltd, UK) has outer dimensions of $1.2\times4.8\times50$ mm, an inner depth of 0.4 mm, and engraved counting frames at intervals of 10, 20, 30 and 40 mm on the tube. Using this design, it was shown that mean count of 39 sperm per high power field at 10 mm, after 30 min incubation at 37°C, is the optimum threshold criterion (Ivic *et al.*, 2002).

In conclusion, $SMPT_{vd}$ has a low accuracy in the evaluation of semen. However, $SMPT_{sc}$ was found to be moderately

accurate in assessing sperm motility in semen. This method of using sperm concentration, instead of vanguard distance, as diagnostic criteria of in-vitro SMPT has potential as a useful laboratory-based sperm function test.

Acknowledgements

We wish to acknowledge the assistance provided by the following people: Christopher L.R.Barratt Reproductive (Biology & Genetics Group, Department of Medicine); David Moore, Anne Fry-Smith and Fujian Song (Department of Public Health and Epidemiology), the University of Birmingham, UK. This project was also supported by a programme grant #0205 to Professor Christopher L.R.Barratt from the University of Birmingham, UK; and training support to B.O. and A.C. from the West Midlands Development and Evaluation Service, University of Birmingham, UK.

References

Aitken, R.J., Bowie, H., Buckingham, D., Harkiss, D., Richardson, D.W. and West, K.M. (1992) Sperm penetration into a hyaluronic acid polymer as a means of monitoring functional competence. *J. Androl.*, **13**, 44–54.

Alexander, N.J. (1981) Evaluation of male infertility with an in vitro cervical mucus penetration test. *Fertil. Steril.*, **36**, 201–208.

Amit, A., Bergman, A., Yedwab, G., David, M.P., Homonnai, T.Z. and Paz, G. (1982) Human sperm penetration into cervical mucus in vitro: correlation between its quality, penetration capacity and velocity. *Gynecol. Obstet. Invest.*, **14**, 283–291.

Barratt, C.L.R., Osborn, J.C., Harrison, P.E., Monks N., Dunphy, B.C., Lenton, E.A. and Cooke, I.D. (1989) The hypo-osmotic swelling test and sperm mucus penetration test in determining fertilization of the human oocyte. *Hum. Reprod.*, **4**, 430–434.

Barratt, C.L., McLeod, I.D., Dunphy, B.C. and Cook, I.D. (1992) Prognostic value of two putative sperm function tests: hypo-osmotic swelling and bovine sperm mucus penetration test (Penetrak). *Hum. Reprod.*, **7**, 1240–1244.

Bothner, H. and Wik, O. (1987) Rheology of hyaluronate. *Acta Otolaryngol.*, **442** (Suppl.), 25–30.

Branigan, E.F., Estes, M.A. and Muller, C.H. (1999) Advanced semen analysis: a simple screening test to predict intrauterine insemination success. *Fertil. Steril.*, **71**, 547–551.

Cochrane methods working group on systematic reviews of screening and diagnostic tests (2000) Screening and diagnostic tests: recommended methods.

David, M.P., Amit, A., Bergman, A., Yedwab, G., Paz, G.F. and Homonnai, Z.T. (1979) Sperm penetration in vitro: correlations between parameters of sperm quality and the penetration capacity. *Fertil. Steril.*, **32**, 676–680.

Deeks, J.J. (2001) Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests *Br. Med. J.*, **323**, 157–162.

De Geyter, Ch., Bals-Pratsch, M., Doeren, M., Yeung, C.H., Grunert, J.H., Bordt, J., Schneider, H.P.G. and Nieschlag, E. (1988) Human and bovine cervical mucus penetration as a test of sperm function for in-vitro fertilization. *Hum. Reprod.*, **3**, 948–954.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Cont. Clin. Trials*, **7**, 177–188.

Dickey, R.P., Pyrzak, R., Lu, P.Y., Taylor, S.N. and Rye, P.H. (1999) Comparison of the sperm quality necessary for successful intrauterine insemination with World Health Organization threshold values for normal sperm. *Fertil. Steril.*, **71**, 684–689

Ducot, B., Spira, A., Feneux, D. and Jouannet, P. (1988) Male factors and the likelihood of pregnancy in infertile couples. II. Study of clinical characteristics—practical consequences. *Int. J. Androl.*, **11**, 395–404.

Dunphy, B.C., Kay, R., Barratt, C.L.R. and Cooke, I.D. (1989) Female age, the length of involuntary infertility prior to investigation and fertility outcome. *Hum. Reprod.*, **4**, 527–530.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.*, **315**, 629–634.

Eggert-Kruse, W., Gerhard, I., Tilgen, W. and Runnebaum. B. (1989a) Clinical significance of crossed in vitro sperm—cervical mucus penetration tests in infertility investigation. *Fertil. Steril.*, **52**, 1032–1040.

Eggert-Kruse, W., Leinhos, G., Gerhard, I., Tilgen, W. and Runnebaum. B. (1989b) Prognostic value of in vitro sperm penetration into hormonally standardized human cervical mucus. *Fertil. Steril.*, **51**, 317–323.

Eimers, J.M., Tevelde, E.R., Gerritse, R., Vogelzang, E.T., Looman, C.W.N. and Habbema, J.D.F. (1994) The prediction of the chance to conceive in sub-fertile couples. *Fertil. Steril.*, **61**, 44–52.

ESHRE (1996) Consensus workshop on advanced diagnostic andrology techniques, ESHRE Andrology Special Interest Group. *Hum. Reprod.*, **11**, 1463–1479.

Fleiss, J. and Gross, A.J. (1991) Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J. Clin. Epidemiol.*, **44**, 127–139.

Ford, W.C.L. (1999) Prediction of fecundability from semen analysis: problems in providing an accurate prognosis. *Hum. Fertil.*, **2** 25–30.

Gaddum-Rosse, P., Blandau, R.J. and Lee, W.I. (1980) Sperm penetration into cervical mucus in vitro. II. Human spermatozoa in bovine mucus. *Fertil. Steril.*, **33**, 644–648.

Guzick, D.S., Overstreet, J.W., Factor-Litvak, P., Brazil, C.K., Nakajima, S.T., Coutifaris, C., Carson, S.A., Cisneros, P., Steinkampf, M.P., Hill, J.A., Xu, D. *et al*. (2001) Sperm morphology, motility and concentration in fertile and infertile men. *N. Engl. J. Med.*, **345**, 1388–1393.

Ishijima, S., Oshio, S. and Mohri, H. (1986) Flagellar movement of human spermatozoa. *Gamete Res.*, **13**, 185–197.

Ivic, A., Onyeaka, H., Girling, A., Brewis, I.A., Ola, B., Hammadieh, N., Papaioannou, S. and Barratt, C.L.R. (2002) Critical evaluation of methylcellulose as an alternative medium in sperm migration tests. *Hum. Reprod.*, **17**, 143–149.

Karni, Z. (1971) Newtonian viscosity of the human cervical mucus during the menstrual cycle. *Int. J. Fertil.*, **16**, 185–188.

Keel, B.A. and Webster, B.W. (1988) Correlation of human sperm motility characteristics with an in vitro cervical mucus penetration test. *Fertil. Steril.*, **49**, 138–143.

Keel, B.A., Kelly, W., Webster, B.W., Zumbach, K.L. and Roberts, D.K. (1987) Application of a bovine cervical mucus penetration test. *Arch. Androl.*, **19**, 33–41.

Kerin, J.F., Matthews, C.D., Svigos, J.M., Makin, A.E., Symons, R.G. and Smeaton, T.C. (1976) Linear and quantitative migration of stored sperm through cervical mucus during the periovular period. *Fertil. Steril.*, **27**, 1054–1058.

Khan, K.S., Dinnes, J. and Kleijnen, J. (2001a) Systematic reviews to evaluate diagnostic tests. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, **95**, 6–11.

Khan, K. S., ter Riet, G., Popay, J., Glanville, J., Sowden, A.J. and Kleijnen, J. (2001b) Study quality assessment: conducting the review. In *Undertaking Systematic Reviews of Research on Effectiveness*, CRD Report No. 4, Phase 5, Stage II, pp. 1–20.

Kremer, J. and Kroeks, M.V. (1975) Modifications of the in vitro spermatozoal penetration test by means of the sperm penetration meter. *Acta Eur. Fertil.*, **6**, 377–380.

Kotoulas, I.G., Burgos-Briceno, L.A., Arana, J. Balmaceda, J.P. and Asch, R.H. (1984) Human sperm penetration in bovine cervical mucus clinical studies. II. Use of split ejaculates. *Fertil. Steril.*, **42**, 268–273.

Kremer, J. (1965) A simple sperm penetration test. *Int. J. Fertil.*, **10**, 209

Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Lee, W.I., Gaddum-Rosse, P. and Blandau, R.J. (1981) Sperm penetration into cervical mucus in vitro. III. Effect of freezing on estrous bovine cervical mucus. *Fertil. Steril.*, **36**, 209

Lijmer, J.G., Mol, B.W., Heisterkamp, S., Bonsel, G.J., Prins, M.H., van der Meulen, J.H. and Bossuyt, P.M. (1999) Empirical evidence of design-related bias in studies of diagnostic tests. *J. Am. Med. Assoc.*, **282**, 1061–1066.

Mangione, C.M., Medley, N.E. and Menge, A.C. (1981) Studies on the use of estrous bovine cervical mucus in the human sperm–cervical mucus penetration technique. *Int. J. Fertil.*, **26**, 20–24.

Marshburn, P.B., Dodson, W.C. and Haney, A.F. (1989) Bovine cervical mucus penetration by human spermatozoa: lack of association with conception. *Arch. Androl.*, **22**, 29–34.

Matthews, C.D., Makin, A.E. and Woodrow, L. (1980) Experience with in vitro sperm penetration testing in infertile and fertile couples. *Fertil. Steril.*, **33**, 187–192.

Meyer, F.A. (1977) Comparison of structural glycoprotein from mucus of different sources. *Biochim. Biophys. Acta*, **493**, 272–282.

Morrow, A., Drudy, L., Gordon, A. and Harrison, R.F. (1992) Evaluation of bovine cervical mucus penetration as a test of human spermatozoal function for an in vitro fertilization programme. *Andrologia*, **24**, 323–326.

Mortimer, D. (2000) Sperm preparation methods. *J. Androl.*, **21**, 357–366.

Mortimer, D., Pandya, I.J. and Sawers, R.S. (1986) Relationship between human sperm motility characteristics and sperm penetration into human cervical mucus in vitro. *J. Reprod. Fertil.*, **78**, 93–102.

Mortimer, D., Mortimer, S.T., Shu, M.A. and Swart, R. (1990) A simplified approach to sperm–cervical mucus interaction testing using a hyaluronate migration test. *Hum. Reprod.*, **5**, 835–841.

Neuwinger, J., Cooper, T.G., Knuth, U.A. and Nieschlag, E. (1991) Hyaluronic acid as a medium for human sperm migration tests. *Fertil. Steril.*, **58**, 396–400.

Oehninger, S., Franken, D.R., Sayed, E., Barroso, G. and Kolm, P. (2000) Sperm function assays and their predictive value for fertilization outcome in IVF therapy: a meta-analysis. *Hum. Reprod. Update*, **6**, 160–168.

Pandya, I.J., Mortimer, D. and Sawers, R.S. (1986) A standardized approach for evaluating the penetration of human spermatozoa into cervical mucus in vitro. *Fertil. Steril.*, **45**, 357–365.

Radack, K.L., Rouan, G. and Hedges, J. (1986) The likelihood ratio: an improved measure for reporting diagnostic test results. *Arch. Pathol. Lab. Med.*, **110**, 689–693.

Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W. and Haynes, R.B. (2000) Diagnosis and screening. In Sackett, D.L., Straus, S.E., Richardson, W.S. et al. (eds), *Evidence-based Medicine: How to Practice and Teach EBM*, 2nd edn. Churchill Livingstone, Edinburgh.

Sato, T. (1990) Confidence limits for the common odds ratio based on the asymptotic distribution of the Mantel–Haenszel estimator. *Biometrics*, **46**, 71–80.

Schutte, B. (1987) Penetration ability of human spermatozoa into standardized bovine cervical mucus (Penetrak) in patients with normal and pathological semen samples. *Andrologia*, **19**, 217–224.

Stumpf, P.G. and Lloyd, T. (1985) In vitro penetration of human sperm into bovine cervical mucus: effects of sperm washing and exposure to low temperature. *Obstet. Gynecol.*, **65**, 42–45.

Takemoto, F.S., Rogers, B.J., Wiltbank, M.C., Soderdahl, D.W., Vaughn, W.K. and Hale, R.W. (1985) Comparison of the penetration ability of human spermatozoa into bovine cervical mucus and zona-free hamster eggs. *J. Androl.*, **6**, 162–170.

Tomlinson, M.J., Kessopoulou, E. and Barratt, C.L.R. (1999) The diagnostic and prognostic value of traditional semen parameters. *J. Androl.*, **20**, 588–593.

Ulstein, M. (1972) Sperm penetration of cervical mucus as a criterion of male fertility. *Acta Obstet. Gynecol. Scand.*, **51**, 335–340.

Ulstein, M. (1973) In vitro sperm penetration of cervical mucus and male fertility. *Andrologie*, **5**, 189–191.

Urry, R.L., Middleton, R.G. and Mayo, D. (1986) A comparison of the penetration of human sperm into bovine and artificial cervical mucus. *Fertil. Steril.*, **45**, 135–137.

West Midlands Development and Evaluation Service (1999) Handbook Version 2.2. Dept of Public Health and Epidemiology, the University of Birmingham, October 1999. URL: http://www.bham.ac.uk/wmidsdes/. Assessed January 3, 2002.

World Health Organization (1999) *Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, 4th edn. Cambridge University Press, Cambridge.

Submitted on April 26, 2002; resubmitted on January 9, 2003; accepted on January 15, 2003

Appendix A. List of excluded studies

Aitken, R.J., Best, F., Richardson, D.W., Schats, R. and Simm, G. (1983) Influence of caffeine on movement characteristics, fertilizing capacity and ability to penetrate cervical mucus of human spermatozoa. *J. Reprod. Fertil.*, **67**, 19–27.

Amit, A., Bergman, A., Yedwab, G., David, M.P., Homonnai, T.Z. and Paz, G. (1982) Penetration of human ejaculated spermatozoa into human and bovine cervical mucus. II. Pattern and velocity of penetration. *Int. J. Fertil.*, **27**, 160–165.

Bergman, A., Amit, A., David, M.P., Homonnai, T.Z. and Paz, G. (1981) Penetration of human ejaculated spermatozoa into human and bovine cervical mucus. I. Correlation between penetration values. *Fertil. Steril.*, **36**, 363–367.

Biljan, M.M., Taylor, C.T., Manasse, P.R., Joughin, E.C., Kingsland, C.R. and Lewis-Jones, D.I. (1994) Evaluation of different sperm function tests as screening methods for male fertilization potential—the value of the sperm migration test. *Fertil. Steril.*, **62**, 591–598.

Blasco, L., Sokoloski, J.E. and Wolf, D.P. (1979) A practical, objective approach to the evaluation of sperm and cervical mucus in humans. *Fertil. Steril.*, **32**, 55–60.

Clarke, G.N. (1997) A simplified quantitative cervical mucus penetration test. *Hum. Reprod.,* **12**, 1184–1187.

Clarke, G.N., Garrett, C. and Baker, H.W. (1998) Quantitative sperm mucus penetration: modified formulae for calculating penetration efficiency. *Hum. Reprod.*, **13**, 1255–1259.

De Geyter, C., Bals-Pratsch, M., Doeren, M., Yeung, C.H., Grunert, J.H., Bordt. J., Schneider, H.P. and Nieschlag, E. (1988) Human and bovine cervical mucus penetration as a test of sperm function for in-vitro fertilization. *Hum. Reprod.*, **3**, 948–954.

Eggert-Kruse, W., Leinhos, G., Gerhard, I., Tilgen, W. and Runnebaum, B. (1989) Prognostic value of in vitro sperm penetration into hormonally standardized human cervical mucus. *Fertil. Steril.*, **51**, 317–323.

Eggert-Kruse, W., Schwalbach, B., Rohr, G., Klinga, K., Tilgen, W. and Runnebaum, B. (1993) Evaluation of polyacrylamide gel as substitute for human cervical mucus in the sperm penetration test. *Fertil. Steril.*, **60**, 540–549.

Eggert-Kruse, W., Reimann-Andersen, J., Rohr, G. Pohl, S., Tilgen, W. and Runnebaum, B. (1995) Clinical relevance of sperm morphology assessment using strict criteria and relationship with sperm–mucus interaction in vivo and in vitro. *Fertil. Steril.*, **63**, 612–624.

Engel, S. and Petzoldt, R. (1999) Human sperm penetration in different media. *Andrologia*, **31**, 233–239.

Engel, S., Schachschal, G. and Petzoldt, R. (1990) Significance of linear cervix mucus penetration for the evaluation of spermatozoa quality. *Andrologia*, **22**, 347–353.

Esfandiari, N., Rouzrokh, A. and Saremi, A.T. (2000) Effects of cryopreservation on penetration ability of human spermatozoa into bovine cervical mucus. *Arch. Androl.*, **45**, 233–238.

Farhi, J., Valentine, A., Bahadur, G., Shenfield, F., Steele, S.J. and Jacobs, H.S. (1995) In-vitro cervical mucus–sperm penetration tests and outcome of infertility treatments in couples with repeatedly negative post-coital tests. *Hum. Reprod.*, **10**, 85–90.

Ikuma, K., Saito,Y., Takeda, M., Koyama, K. and Isojima, S. (1988) Bovine cervical mucus as a substitute of human cervical mucus. *Nippon Sanka Fujinka Gakkai Zasshi*, **40**, 888–894.

Ikuma, K., Suno, S., Hasegawa, A., Koyama, K. and Isojima, S. (1989) Role of sperm passage through cervical mucus: fertilizing capacity tested by in vitro fertilization with zona-free hamster eggs. *Nippon Sanka Fujinka Gakkai Zasshi*, **41**, 167–172.

Jonsson, B., Eneroth, P., Landgren, B.M. and Wikborn, C. (1986) Evaluation of in vitro sperm penetration testing of 176 infertile couples with the use of ejaculates and cervical mucus from donors. *Fertil. Steril.*, **45**, 353–356.

Jonsson, B., Landgren, B.M. and Eneroth, P. (1989) Repeated mid-cycle tests of in-vitro sperm penetration (Kremer tests) in healthy women during three menstrual cycles. *Hum. Reprod.*, **4**, 670–673.

Katzorke, T., Propping, D. and Kolodziej, F.B. (1985) Crossed sperm-penetration-meter test. Prognostic value in male subfertility. *Andrologia*, **17**, 234–240.

Kerin, J.F., Matthews, C.D., Svigos, J.M., Makin, A.E., Symons, R.G. and Smeaton, T.C (1976) Linear and quantitative migration of stored sperm through cervical mucus during the periovular period. *Fertil. Steril.*, **27**, 1054–1058.

Kolodziej, F.B., Katzorke, T. and Propping, D. (1986) Prognostic value of the sperm-penetration-meter test according to Kremer. *Andrologia*, **18**, 539–544.

Kotoulas, I.G., Burgos-Briceno, L.A., Arana, J., Balmaceda, J.P. and Asch, R.H. (1984) Human sperm penetration in bovine cervical mucus clinical studies. II. Use of split ejaculates. *Fertil. Steril*., **42**, 268–273.

Kremer, J. and Kroeks, M.V. (1975) Modifications of the in vitro spermatozoal penetration test by means of the sperm penetration meter. *Acta Eur. Fertil*., **6**, 377–380.

Lee, W.I., Gaddum-Rosse, P. and Blandau, R.J. (1981) Sperm penetration into cervical mucus in vitro. III. Effect of freezing on estrous bovine cervical mucus. *Fertil. Steril*., **36**, 209–213.

Morrow, A., Drudy, L., Gordon, A. and Harrison, R.F. (1992) Evaluation of bovine cervical mucus penetration as a test of human spermatozoal function for an in vitro fertilization programme. *Andrologia*, **24**, 323–326.

Neuwinger, J., Cooper, T.G., Knuth, U.A. and Nieschlag, E. (1991) Hyaluronic acid as a medium for human sperm migration tests. *Hum. Reprod*., **6**, 396–400.

Niederberger, C.S., Lamb, D.J., Glinz, M., Lipshultz, L.I. and Scully, N.F. (1993) Tests of sperm function for evaluation of the male: Penetrak and Tru-Trax. *Fertil. Steril*., **60**, 319–323.

Pandya, I.J., Mortimer, D. and Sawers, R.S. (1986) A standardized approach for evaluating the penetration of human spermatozoa into cervical mucus in vitro. *Fertil. Steril*., **45**, 357–365.

Randall, J.M. and Templeton, A. (1991) Cervical mucus score and in vitro sperm mucus interaction in spontaneous and clomiphene citrate cycles. *Fertil. Steril*., **56**, 465–468.

Sharara, F.I., Beatse, S.N., Bailey, S.A., Neal G.S., Coddington, C.C. 3rd and Scott, R.T. Jr (1994) Characterization of Tru-Trax in-vitro penetration testing of cervical mucus. *Hum. Reprod*., **9**, 2027–2031.

Sharara, F.I., Illions, E.H., Coddington, C.C. 3rd and Scott, R.T. Jr (1995) Evaluation of the Tru-Trax cervical mucus penetration test in predicting fertilization and pregnancy rates in in-vitro fertilization. *Hum. Reprod*., **10**, 1481–1485.

Shibahara, H., Naito, S., Hasegawa, A., Mitsuo, M., Shigeta, M. and Koyama, K. (1997) Evaluation of sperm fertilizing ability using the Sperm Quality Analyzer. *Int. J. Androl*., **20**, 112–117.

Sokol, R.Z., Madding, C.I., Handelsman, D.J. and Swerdloff, R.S. (1986) The split ejaculate: assessment of fertility potential using two in vitro test systems. *Andrologia*, **18**, 380–386.

Tang, S., Garrett, C. and Baker, H.W. (1999) Comparison of human cervical mucus and artificial sperm penetration media. *Hum. Reprod*., **14**, 2812–2817.

Ulstein, M. (1972) Sperm penetration of cervical mucus as a criterion of male fertility. *Acta Obstet. Gynecol. Scand*., **51**, 335–340.

Ulstein, M. (1973a) In vitro sperm penetration of cervical mucus and male fertility. *Andrologie*, **5**, 189–191.

Ulstein, M. (1973b) Fertility, motility and penetration in cervical mucus of freeze-preserved human spermatozoa. *Acta Obstet. Gynecol. Scand*., **52**, 205–210.

Urry, R.L., Carrell, D.T., Hull, D.B., Middleton, R.G. and Wiltbank, M.C. (1983) Penetration of zona-free hamster ova and bovine cervical mucus by fresh and frozen human spermatozoa. *Fertil. Steril*., **39**, 690–694.

Urry, R.L., Middleton, R.G. and Mayo, D. (1986) A comparison of the penetration of human sperm into bovine and artificial cervical mucus. *Fertil. Steril*., **45**, 135–137.

Zhu, J.J., Barratt, C.L.R. and Cooke, I.D. (1992) Effect of human cervical mucus on human sperm motion and hyperactivation in vitro. *Hum. Reprod*., **7**, 1402–1406.

Appendix B. Definitions of some commonly used terms in diagnostic accuracy studies using examples from a 2×2 table

	Reference test or gold standard		
	Disease present	Disease absent	Totals
Diagnostic test			
Test positive	a (true test positive)	b (false test positive)	a + b
Test negative	c (false test negative)	d (true test negative)	c + d
Totals	a + c	b + d	a + b + c + d

Sensitivity is the truly positive test rate or the probability of a positive test result, if disease is present
= a ÷ (a + c)
Specificity is the truly negative test rate or the probability of a negative test result, if disease is absent
= d ÷ (b + d)
Positive predictive value (PPV) is the probability of disease, if the test is positive
= a ÷ (a + b)
Negative predictive value (NPV) is the probability of 'no disease', if test is negative
= d ÷ (c + d)
Likelihood ratio of a positive test (LR+) is the ratio of a true positive test rate to a false positive test rate when disease is present
= sensitivity ÷ (1 – specificity)
Likelihood ratio of a negative test (LR–) is the ratio of a false negative test rate to a true negative test rate when disease is absent
= (1 – sensitivity) ÷ specificity