



www.sciencedirect.com  
www.rbmonline.com



## REVIEW

# Sperm chromatin structure assay and classical semen parameters: systematic review

JA Castilla <sup>a,b,\*</sup>, S Zamora <sup>a</sup>, MC Gonzalvo <sup>a</sup>, JD Luna del Castillo <sup>c</sup>,  
JA Roldan-Nofuentes <sup>c</sup>, A Clavero <sup>a</sup>, L Björndahl <sup>d</sup>, L Martínez <sup>a</sup>


<sup>a</sup> Reproduction Unit, Hospital 'Virgen de las Nieves', E-18014 Granada, Spain; <sup>b</sup> Sperm Bank CEIFER, Granada, Spain;

<sup>c</sup> Biostatistics, Faculty of Medicine, University of Granada, Granada, Spain; <sup>d</sup> Centre for Andrology and Sexual Medicine, Clinic of Endocrinology, Department of Medicine, Huddinge, Karolinska Institutet, Sweden

\* Corresponding author. E-mail address: josea.castilla.sspa@juntadeandalucia.es (JA Castilla).



Jose Castilla graduated in medicine in 1985, completed his PhD in 1986 (Granada University) and then specialised in clinical analyses. He has been Director of the Andrology and Embryology, Virgen de las Nieves, Granada, Spain since 1991 and founded the CEIFER Sperm Bank in 1993. He has been co-ordinator of the Spanish External Quality Control Programme for Semen Analysis since 1999 and for the Assisted Reproduction Laboratory since 2003. A founder member of the Spanish Association of Clinical Embryologists (ASEBIR), and on its executive committee from 1993 to 2000, he is currently on the executive committee of the Spanish Fertility Society and deputy co-ordinator of ESHRE's Special Interest Group in Andrology.

**Abstract** The present study is based on a PubMed search and compares the clinical validity of classical semen parameters (CSP) and the sperm chromatin structure assay (SCSA) in different clinical contexts. The PubMed database was searched using keywords on the sperm diagnostic test for pregnancy in three clinical scenarios: (i) couples attempting to conceive; (ii) couples who had been attempting to conceive for 12 months without success; and (iii) couples treated with intrauterine insemination (IUI). There was a considerable heterogeneity among the studies included. For couples attempting to conceive following a SCSA that produced an abnormal result, the likelihood of male factor infertility ranged from a pre-test value of 7.5% to a post-test value of 32.1% [95% confidence interval (CI) 15.7–54.5], while after CSP with an abnormal result, the post-test probability was 17.3% (95% CI 11.8–24.5). For a pre-test prevalence of male factor infertility of 50%, the post-test probability of male factor infertility after an abnormal test is very similar for both SCSA and CSP. In couples treated with IUI, the clinical validity of SCSA is higher than that of sperm morphology alone, but not enough to introduce SCSA as a test in male infertility work-up. 

© 2009, Reproductive Healthcare Ltd. Published by Elsevier Ltd. All rights reserved.

**KEYWORDS:** clinical validity, semen parameters, sperm DNA fragmentation

## Introduction

Traditionally, the definition of a male factor has been based on the concentration, motility and morphology of spermatozoa in semen, which are referred to as the classical sperm

parameters (CSP) (Kvist and Björndahl, 2002; World Health Organization [WHO], 1999). Nevertheless, the clinical value of CSP has been questioned (Alvarez et al., 2003; Bonde et al., 1998; Glazener et al., 1987; Lewis, 2007; Nallella et al., 2006; Swan, 2006). In consequence, other functional

tests have been proposed for a more direct assessment of male fertility potential. Among such tests are those based on an analysis of DNA stainability to estimate DNA damage or fragmentation. The introduction of the sperm chromatin structure assay (SCSA), first described in 1980, enabled the level of DNA breaks to be quantified by means of the DNA fragmentation index (DFI) using a flow-cytometric technique (Evenson et al., 1999).

Sperm DNA fragmentation assessment has been recommended as a complementary test in male infertility work-up by some authors (Evenson and Wixon, 2008; Saleh et al., 2002, 2003), especially in order to choose the most appropriate assisted reproduction treatment (Erenpreiss et al., 2006). However, other authors have suggested that this investigation should be introduced as a routine test in infertile men, based on the observation that a significant proportion of men with normal CSP have high levels of altered DNA stainability, which was interpreted as DNA damage (Erenpreiss et al., 2008; Nicopoullos et al., 2008). However, before introducing any diagnostic test into clinical practice, its clinical validity needs to be compared with that of existing methods (Bossuyt et al., 2000).

One way to assess the usefulness of a test is to calculate the ratio of the probability of a given test proving positive for patients with a specific condition (e.g. subfertility caused by a male factor) to the probability of the same test producing a positive result for patients without this condition (e.g. fertile men). This is termed the likelihood ratio (LR) and it represents the magnitude of change from a clinician's initial suspicion of the presence of disease (pre-test probability) to the likelihood of disease after the test result (post-test probability). The likelihood ratio is of enormous practical value, and it is becoming the preferred way of expressing and comparing the usefulness of different tests (Greenhalgh, 1997; Sackett et al., 1991). It is also necessary to determine the pre-test probability of disease (clinical scenario) in order to define the clinical validity of a test.

A test used in the investigation of male infertility can be put into practice in different clinical scenarios with different pre-test probabilities. First, when a couple decide to try to conceive, men with a normal result may be encouraged to try longer, but a man with an abnormal result is not likely to cause a spontaneous pregnancy and the couple is best helped if an infertility investigation with semen analysis according to international standards is initiated without further delay (Lefièvre et al., 2007). The prevalence of male factor subfertility in this scenario has been estimated at 7.5% (15% of couples are subfertile and male factors are believed to be involved in half of them). Second, a male factor test may be performed when a couple has been trying to conceive for 12 months without success (subfertile patients), and so a diagnostic test is required to determine the cause of this infertility and to predict which couples are likely to conceive spontaneously during the forthcoming months, and which are not. The prevalence of a male factor among subfertile couples attending a tertiary referral centre is 50%. Third, male factor subfertility is tested when assisted reproduction treatment is required. In this case, recent meta-analyses have shown that DNA fragmentation does not predict clinical pregnancy when IVF or intracytoplasmic sperm injection is performed (Collins et al., 2008; Zini et al., 2008). Therefore, the only use of the DNA frag-

mentation test in relation to assisted reproduction treatment is to decide whether or not to carry out intrauterine insemination (IUI) or to pass directly to IVF.

This paper, based upon a PubMed search, compares the clinical validity of CSP and SCSA in different clinical scenarios in order to determine whether the introduction of DNA fragmentation assessment would give more information than that obtained from CSP.

## Materials and methods

In order to compare the clinical validity of SCSA and CSP in diagnosing male factor infertility, the results of the systematic reviews or meta-analyses that have been published regarding each clinical situation were compared, and when none was found, a systematic search was performed.

Systematic reviews and meta-analyses were only found for IUI (Cocuzza et al., 2007; Evenson and Wixon, 2006; Van Waart et al., 2001) (Table 1). For other clinical situations, a systematic review of PubMed was performed (from 1993 to August 2008). The systematic search for papers relevant to this study produced only limited results, which were supplemented by consulting other information sources via citations and an extensive review of the literature. For the systematic search, the terms 'semen', 'pregnancy' and 'human' for humans were combined and initially obtained 1592 hits. After scanning of citations and abstracts, 1563 of these were excluded. Thus, a total of 29 possible relevant publications were analysed. After reviewing these papers, a further six were excluded because they compared normozoospermic with oligoastenozoospermic patients, and a further eight because they were based on the time to pregnancy. Of the 15 potentially useful articles, the study only included the ones that contained (or enabled the reconstruction of)  $2 \times 2$  tables (true positive, false positive, true negative and false negative), such that the diagnostic efficiency statistics could be derived. According to Standards for Reporting of Diagnostic Accuracy (STARD) an abnormal result is denominated positive (Bossuyt et al., 2003). Therefore, the number of publications included was further reduced to 11, of which three were included in the group of Natural Relations (two SCSA and one CSP), and eight in the group of fertile versus infertile (two SCSA and six CSP). These studies are described in Table 1.

In order to compare the clinical validity of SCSA and CSP, a statistical analysis was carried out as described below. The following diagnostic efficiency statistics were used for this comparison: sensitivity, specificity, positive LR (LR+) and negative LR (LR-). In addition, whenever possible, the positive predictive value (PPV) and the negative predictive value (NPV) were calculated. To investigate all studies standardized for different levels of prevalence, predictive values with fixing values of prevalence were calculated for a prevalence of 7.5% (cf. couples attempting to conceive; PPV-7.5; NPV-7.5), 50% (cf. fertile versus subfertile couples; PPV-50; NPV-50) and 85% (cf. couples receiving IUI; PPV-85; NPV-85), and unconditional PV (uPPV; uNPV) were calculated (Jialiang et al., 2007).

In all cases, the point estimations of the diagnostic efficiency statistics and the asymptotic confidence intervals were calculated. For the case of uPPV and uNPV, bootstrap

**Table 1** Studies included in systematic review.

<i>Study</i>	<i>Clinical scenario</i>	<i>Spectrum of patient</i>	<i>Test</i>	<i>Seminal parameter</i>	<i>TP</i>	<i>FN</i>	<i>FP</i>	<i>TN</i>
Evenson et al. (1999)	Attempting to conceive	Couples attempting to conceive	SCSA	DFI	6	25	4	109
Spanò et al. (2000)	Attempting to conceive	First pregnancy planners with no previous knowledge of their fertility capability	SCSA	DFI	25	84	4	102
Bonde et al. (1998)	Attempting to conceive	Couples attempting to conceive	CSP	Sperm concentration	49	120	28	221
Chohan et al. (2006)	Fertile versus subfertile patients	Fertile donors, subfertile men	SCSA	DFI	12	48	0	7
Saleh et al. (2002)	Fertile versus subfertile patients	Fertile donors, subfertile men	SCSA CSP	DFI Concentration, motility, morphology	53 21	39 71	0 1	15 15
Bartoov et al. (1993)	Fertile versus subfertile patients	Subfertile males, fertile males	CSP	Semen analysis index	43	64	3	98
Guzick et al. (2001)	Fertile versus subfertile patients	Subfertile males men with proven fertility	CSP	Sperm concentration	113	652	27	669
Menkveld et al. (2001)	Fertile versus subfertile patients	Oligozoospermic subfertile males, fertile men	CSP	Morphology	77	26	24	83
Nallella et al. (2006)	Fertile versus subfertile patients	Subfertile males, men with proven fertility	CSP	Percentage motility	123	43	6	50
Ombelet et al. (1997a)	Fertile versus subfertile patients	Subfertile couples, fertile men	CSP	Normal count	85	51	27	117
Bungum et al. (2004)	Intrauterine insemination	Unexplained subfertility	SCSA	DFI	22	86	1	22
Bungum et al. (2007)	Intrauterine insemination	Unexplained subfertility	SCSA	DFI	64	245	2	76
Karabinus and Gelety (1997)	Intrauterine insemination	Whole subfertile population	CSP	Morphology	50	441	3	44
Lindheim et al. (1996)	Intrauterine insemination	Whole subfertile population	CSP	Morphology	98	62	1	15
Matorras et al. (1995)	Intrauterine insemination	Whole subfertile population	CSP	Morphology	154	89	18	10
Montanaro Gauci et al. (2001)	Intrauterine insemination	Whole subfertile population	CSP	Morphology	37	239	1	35
Ombelet et al. (1997b)	Intrauterine insemination	Whole subfertile population	CSP	Morphology	295	384	40	76
Toner et al. (1995)	Intrauterine insemination	Whole subfertile population	CSP	Morphology	80	274	6	35

CSP, classical semen parameters; DFI, DNA fragmentation index; FN, false negative; FP, false positive; TN, true negative; TP, true positive; SCSA, sperm chromatin structure assay.

intervals were computed because explicit expressions for standard errors of estimates were not available (Jialiang et al., 2007). When the  $2 \times 2$  tables contained zero cells, reasonable estimates of some parameters (likelihood ratios, odds ratio, etc.) were not possible. In order to avoid these problems, 0.5 was added to all cells of the table (Zhou

et al., 2002). Post-test probability was calculated using a likelihood ratio nomogram (Fagan, 1975).

It was anticipated that, in accordance with other diagnostic reviews (Collins et al., 2008), there would be considerable heterogeneity of results among the different studies. For each of the tests in the different clinical scenarios, the

heterogeneity of the diagnostic test properties was assessed by Cochran's  $Q$  test (Greenland, 1987). In addition, the heterogeneity was quantified by the  $I^2$  value, i.e. the proportion of variability across studies that is due to heterogeneity rather than chance (Higgins and Thompson, 2002). Very high values in this respect (above 0.5) reflect a high degree of heterogeneity and suggest the need for a more detailed study of the subgroups included. In this report, the small number of studies did not allow for a detailed exploration of the reasons for heterogeneity using meta-regression techniques. Finally, a pooled estimation of the diagnostic efficiency statistics for each test was performed and compared SCSA and CSP using the method proposed by Dersimonian and Laird (1987), which is affected only to a minor degree by heterogeneity among the studies. To calculate the pooled diagnostic odds ratio and LR, a correction factor of 0.5 was added to all four cells in the  $2 \times 2$  table and logs were used in accordance with the recommendations of Gart and Zweifl (1967). The data for the different studies were analysed using STATA version 10.1 software (StataCorp LP, College Station, TX, USA).

## Results

In all the scenarios analysed, the pooled specificity, PPV, PPV calculated with a fixing value of prevalence and uPPV were higher than the pooled sensitivity, NPV, NPV calculated with a fixing value of prevalence and uNPV, respectively. In the same way, the increase in the probability of male factor subfertility or of non-pregnancy with IUI after a positive SCSA or CSP result (LR+; rule in disease)

was higher than the increase in the probability of a couple not presenting male factor subfertility or pregnancy with IUI after a negative result (LR-; rule out disease).

For couples seeking to conceive, only pooled specificity showed a significantly higher value for SCSA than for CSP, while the other pooled diagnostic efficiency statistics obtained were similar in SCSA and CSP (Table 2). In this scenario, after a SCSA with an abnormal result, the likelihood of male factor subfertility would increase from the pre-test probability of 7.5% to a post-test probability of 32.1% (95% CI 15.7–54.5), and after a CSP analysis with an abnormal result, it would increase to a post-test probability of 17.3% (95% CI 11.8–24.5) (Figure 1). These differences in LR are not statistically significant (Table 2).

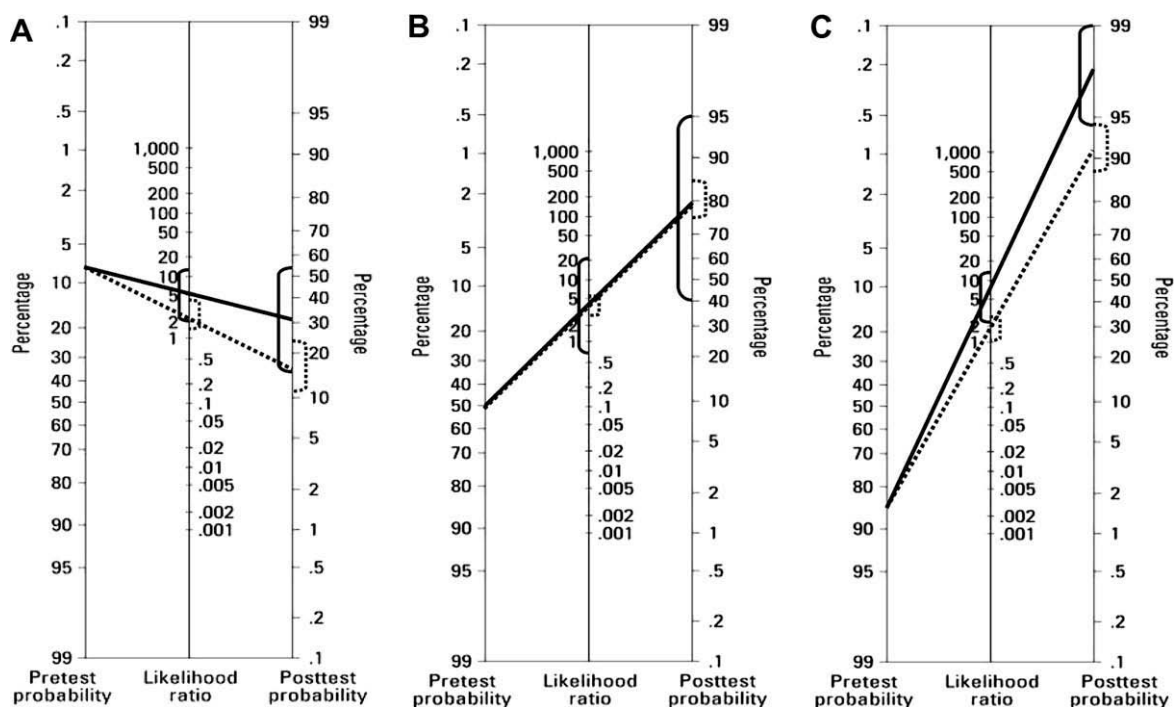
In the fertile versus subfertile scenario, almost all the sensitivity and specificity values were heterogeneous (i.e. the  $I^2$  values were higher than 0.50) (Table 3). Statistically significant differences ( $P < 0.05$ ) were only observed in PPV and NPV, with a higher PPV for SCSA and a higher NPV for CSP. Nevertheless, these differences disappear with PPV-50 and uPPV (Table 3). For a pre-test prevalence of male factor infertility of 50%, the post-test probability of male factor infertility after an abnormal test is very similar for SCSA and CSP (Figure 1).

A high degree of heterogeneity was observed among the studies included in the meta-analysis of sperm morphology in IUI (Table 4). In this scenario, the clinical validity of SCSA was higher than that of sperm morphology, with a LR+ value for SCSA of 6.1 (2.6–14.6) and of 1.9 (1.1–3.0) for sperm morphology ( $P < 0.05$ ). For a pre-test probability of non-pregnancy after IUI of 85%, the post-test probability of non-pregnancy after an abnormal SCSA was 97.2% and after

**Table 2** Diagnostic efficiency statistics not related to prevalence for sperm chromatin structure assay and classical semen parameters when couples are seeking to conceive.

	Study	Sensitivity	Specificity	Estimated PPV according to 7.5% prevalence	Unconditional positive predictive value	Positive likelihood ratio	Negative likelihood ratio	Diagnostic odds ratio
SCSA	Evenson et al. (1999)	0.19 (0.70–0.36)	0.96 (0.92–0.99)	0.31 (0.10–0.68)	0.76 (0.55–0.91)	5.47 (1.36–26.60)	0.84 (0.66–0.97)	6.20 (1.53–24.76)
	Spanò et al. (2000)	0.23 (0.15–0.32)	0.96 (0.92–0.99)	0.33 (0.17–0.70)	0.77 (0.65–0.91)	6.08 (2.57–28.11)	0.80 (0.71–0.89)	6.87 (2.78–25.69)
	Pooled	0.22 (0.15–0.30)	0.96 (0.94–0.99)	0.32 (0.13–0.52)	0.77 (0.66–0.87)	5.83 (2.30–14.81)	0.81 (0.73–0.89)	6.60 (2.77–15.73)
	$I^2$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CSP	Bonde et al. (1998)	0.29 (0.22–0.36)	0.89 (0.85–0.93)	0.17 (0.13–0.26)	0.65 (0.60–0.73)	2.58 (1.78–4.28)	0.80 (0.72–0.88)	3.19 (1.98–5.56)
	Pooled	0.29 (0.22–0.36)	0.89 (0.85–0.93)	0.17 (0.11–0.24)	0.65 (0.59–0.72)	2.58 (1.66–4.00)	0.80 (0.71–0.89)	3.19 (1.91–5.36)
	$I^2$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$P^a$	NS	0.001	NS	0.070	NS	NS	NS

Values in parentheses are 95% CI; CSP, classical semen parameters; NS, not statistically significant; SCSA, sperm chromatin structure assay. <sup>a</sup> $P$ -value obtained from the comparison of the pooled estimation of diagnostic efficiency statistics of SCSA versus CSP.



**Figure 1** Fagan nomogram using likelihood ratio and pre-test probability for sperm chromatin structure assay (SCSA) and classical semen parameters (CSP): (A) couples attempting to conceive; (B) couples who have been trying for 12 months to conceive without success; and (C) couples treated with intrauterine insemination. Solid lines are SCSA test and dotted lines are CSP. Confidence intervals are in brackets. To use this tool, the probability or prevalence of the disease and the likelihood ratio for the diagnostic test has to be known. With this information, a line connecting the pre-test probability and the likelihood ratio is drawn and extended until it intersects with the post-test probability. The point of intersection is the new estimate of the probability that the patient has this disease.

an abnormal sperm morphology result, it was 91.3% (Figure 1).

## Discussion

Depending on the disease and the use made of the test (diagnostic or screening) the clinician needs certain attributes of a test to be more accentuated than others (e.g. more sensitivity or more specificity). In the case of the diagnostic test of male infertility, the test should have the lowest possible number of false positives (fertile men testing as abnormal). This is for two reasons; first, due to the 'physiopathological reason' described by Jeyendran and Zaneveld (1993). False negatives in diagnostic testing of male infertility are not important because when a single test gives normal results this does not mean that other defects, which may render the spermatozoa infertile, are not present. The characteristic evaluated by the test is only one among a multitude of factors determining fertility. For this reason, diagnostic tests of male infertility normally present low levels of sensitivity. However, a false positive in a diagnostic test of male infertility indicates that the parameter analysed is not really important because, even when it is abnormal, fertility can occur. The second reason could be named the 'clinical reason'. From a clinical point of view, it seems more acceptable to consider diagnosing subfertile males as fertile rather than diagnosing fertile males as subfertile. This approach would prevent the over-treatment of

potentially fertile males, for instance, by referring couples for intracytoplasmic sperm injection when another, less aggressive treatment could have been employed (Sergerie et al., 2005).

The false positive rate is inversely related with PPV. Therefore, for the reasons mentioned above, a diagnostic test of male infertility should have a high PPV. However, PPV is crucially dependent on the population chosen and on the prevalence of the disease or disorder. The test (PPV) performs less well with lower rates of prevalence. This means that differences among studies that analyse the clinical utility of diagnostic tests may be caused by differences in prevalence rates, and not by true differences in test performance. For an appropriate comparison of PPV rates, various strategies were used, including deriving the predictive value calculated with a fixing value of prevalence and calculating uPPV and uNPV. Neither in the case of couples seeking to conceive nor in the comparison of fertile versus subfertile men did either of these strategies reveal significant differences concerning the predictive value of SCSA and CSP.

Unlike PPV, LR is independent of disease prevalence, and the larger the LR+, the greater the likelihood of male infertility. In accordance with the above observations, an ideal diagnostic test of male subfertility should be capable of finding male subfertility, i.e. it should have a high LR+. The DNA fragmentation test has been presented as fulfilling this condition of being a test of 'infertility' (Evenson and Wixon, 2006). Evenson and Wixon (2008), basing their

**Table 3** Diagnostic efficiency statistics not related to prevalence for sperm chromatin structure assay and classical semen parameters in fertile patients versus subfertile patients.

	<i>Study</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Estimated PPV according to 50% prevalence</i>	<i>Unconditional positive predictive value</i>	<i>Positive likelihood ratio</i>	<i>Negative likelihood ratio</i>	<i>Diagnostic odds ratio</i>
SCSA	Chohan et al. (2006)	0.20 (0.10–0.31)	0.87 (0.60–0.96)	0.62 (0.30–0.85)	0.58 (0.36–0.76)	1.60 (0.43–5.69)	0.91 (0.75–1.35)	1.29 (0.25–6.84)
	Saleh et al. (2002)	0.58 (0.48–0.67)	0.94 (0.78–0.98)	0.90 (0.71–0.96)	0.82 (0.64–0.90)	9.22 (2.43–24.07)	0.45 (0.34–0.59)	14.00 (3.60–58.97)
	Pooled	0.39 (0.02–0.76)	0.92 (0.84–1.01)	0.79 (0.51–1.06)	0.71 (0.48–0.95)	3.94 (0.71–21.91)	0.64 (0.32–1.28)	4.43 (0.43–45.79)
	$I^2$	96.3	0.0	71.7	74.7	74.4	91.6	78.6
CSP	Bartoov et al. (1993)	0.40 (0.31–0.49)	0.97 (0.93–0.99)	0.93 (0.85–0.97)	0.85 (0.76–0.95)	13.53 (5.53–70.00)	0.62 (0.52–0.71)	18.98 (7.37–89.57)
	Guzick et al. (2001)	0.15 (0.12–0.17)	0.96 (0.95–0.97)	0.79 (0.72–0.86)	0.71 (0.65–0.77)	3.81 (2.56–5.92)	0.89 (0.86–0.92)	4.23 (2.79–6.71)
	Menkveld et al. (2001)	0.75 (0.66–0.83)	0.78 (0.7–0.85)	0.77 (0.71–0.84)	0.69 (0.64–0.75)	3.33 (2.42–5.16)	0.32 (0.23–0.45)	9.97 (5.49–19.51)
	Nallella et al. (2006)	0.74 (0.67–0.8)	0.89 (0.8–0.96)	0.87 (0.78–0.95)	0.79 (0.70–0.88)	6.92 (3.59–18.66)	0.29 (0.22–0.37)	22.06 (9.82–56.20)
	Ombelet et al. (1997a)	0.62 (0.55–0.7)	0.81 (0.75–0.87)	0.77 (0.71–0.84)	0.69 (0.64–0.75)	3.33 (2.41–5.13)	0.46 (0.36–0.57)	7.09 (4.27–13.02)
	Saleh et al. (2002)	0.23 (0.15–0.33)	0.94 (0.78–0.98)	0.79 (0.47–0.91)	0.70 (0.48–0.83)	3.65 (0.89–10.35)	0.82 (0.71–0.97)	3.11 (0.71–13.19)
	Pooled	0.48 (0.22–0.74)	0.90 (0.84–0.95)	0.82 (0.76–0.88)	0.73 (0.68–0.78)	3.92 (3.00–5.13)	0.53 (0.38–0.74)	8.37 (4.80–14.59)
	$I^2$	99.0	88.5	70.3	58.6	26.4	96.2	69.8

Values in parentheses are 95% CI; CSP, classical semen parameters; SCSA, sperm chromatin structure assay; none of the *P*-values obtained from the comparison of the pooled estimation of diagnostic efficiency statistics of SCSA versus CSP were statistically significant.

**Table 4** Diagnostic efficiency statistics not related to prevalence for sperm chromatin structure assay and classical semen parameters in intrauterine insemination.

	<i>Study</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Estimated PPV according to 85% prevalence</i>	<i>Unconditional positive predictive value</i>	<i>Positive likelihood ratio</i>	<i>Negative likelihood ratio</i>	<i>Diagnostic odds ratio</i>
SCSA	Bungum et al. (2004)	0.20 (0.13–2.29)	0.96 (0.85–0.98)	0.96 (0.86–0.97)	0.74 (0.52–0.85)	4.68 (1.12–12.75)	0.83 (0.73–0.95)	3.90 (0.99–16.49)
	Bungum et al. (2007)	0.21 (0.16–0.26)	0.97 (0.93–0.99)	0.98 (0.94–0.99)	0.80 (0.67–0.92)	8.08 (2.86–33.83)	0.81 (0.76–0.87)	8.04 (3.03–42.51)
	Pooled	0.21 (0.17–0.25)	0.97 (0.94–1.00)	0.98 (0.95–1.00)	0.78 (0.68–0.88)	6.13 (2.57–14.59)	0.82 (0.77–0.87)	5.73 (2.19–15.01)
	$I^2$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CSP	Karabinus and Gelety (1997)	0.10 (0.08–0.13)	0.94 (0.85–0.99)	0.90 (0.79–0.98)	0.58 (0.43–0.81)	1.60 (0.66–8.57)	0.96 (0.90–1.05)	1.45 (0.59–9.46)
	Lindheim et al. (1996)	0.61 (0.54–0.68)	0.94 (0.78–0.98)	0.98 (0.94–0.99)	0.83 (0.66–0.91)	9.80 (2.74–26.04)	0.41 (0.33–0.51)	16.29 (4.50–67.12)
	Matorras et al. (1995)	0.84 (0.77–0.91)	0.19 (0.05–0.43)	0.86 (0.83–0.89)	0.51 (0.48–0.56)	1.04 (0.86–1.47)	0.84 (0.33–3.64)	1.36 (0.25–4.20)
	Montanaro Gauci et al. (2001)	0.13 (0.10–0.18)	0.97 (0.90–100)	0.97 (0.88–0.99)	0.74 (0.54–0.85)	4.83 (1.25–12.36)	0.89 (0.84–0.96)	3.71 (1.12–14.26)
	Ombelet et al. (1997b)	0.43 (0.40–0.47)	0.66 (0.57–0.74)	0.88 (0.85–0.91)	0.54 (0.50–0.60)	1.26 (0.99–1.73)	0.86 (0.74–1.01)	1.45 (0.98–2.29)
	Toner et al. (1995)	0.23 (0.18–0.28)	0.85 (0.73–0.95)	0.90 (0.82–0.96)	0.57 (0.46–0.73)	1.54 (0.80–4.44)	0.91 (0.80–1.07)	1.60 (0.73–4.44)
	Pooled	0.39 (0.19–0.59)	0.78 (0.63–0.93)	0.91 (0.86–0.97)	0.62 (0.53–0.70)	1.85 (1.14–3.02)	0.79 (0.67–0.94)	2.32 (1.19–4.53)
	$I^2$	99.2	94.7	89.7	83.3	75.5	90.0	60.8
	$P^a$	NS	0.012	0.033	0.014	0.018	NS	0.131

Values in parentheses are 95% CI; values are CSP, classical semen parameters; NS, not significant; SCSA, sperm chromatin structure assay.

<sup>a</sup>P-value obtained from the comparison of the pooled estimation of diagnostic efficiency statistics of SCSA versus CSP.

conclusions on calculations of the value and statistical significance of the odds ratio (OR) of two studies (Evenson et al., 1999; Spanò et al., 2000) proposed the SCSA as a prognostic pregnancy test using natural relations. It is believed that, for the reasons mentioned above, this decision should be based on the calculation and analysis of the LR+ and not the OR, because the latter may be distorted by the LR-, giving rise to a high OR, produced by a low LR- rather than a high LR+. Niederberger (2003) and Sackett et al. (1996) recommended against its use, while many authors have preferred to use the LR rather than the OR in order to determine the clinical validity of a test (Biggerstaff, 2000; Dujardin et al., 1994; Roldan-Nofuentes and Luna del Castillo, 2007).

The pooled LR+ obtained with SCSA in the three scenarios described ranged from 3.53 to 6.13, which means that the increased probability of male factor infertility being present or of non-pregnancy after IUI can only be classified as small or moderate. A LR+ is considered to represent a large and often conclusive increase in the likelihood of disease when its value is more than 10 (Jaescheke et al., 2002). This result leads us to disagree with Evenson and Wixon (2008), and not to recommend the use of SCSA as a routine test in subfertile men, this viewpoint being in accordance with the guidelines of the ASRM (Practice Committee of American Society for Reproductive Medicine, 2006).

Studies designed to test the clinical validity of a test for couples attempting to conceive need to be prospective; in these studies, the prevalence of non-pregnancy will depend on the follow-up time, with a range of 20–40% at 6 months and 10–15% at 12 months. The influence of these differing rates of prevalence on the rate of false positives has been discussed by Van der Steeg et al. (2005), who showed that bringing forward sperm tests can lead to a diagnosis of subfertility being made among a high percentage of couples that, nevertheless, achieve pregnancy in the next few months. Thus, unnecessary assisted reproduction treatment is provided to these couples. This conclusion is confirmed by the fact that the spontaneous pregnancy rate has been found to be higher for subfertile couples referred by a general practitioner to a secondary centre than for couples referred by a gynaecologist to a tertiary centre (Snick et al., 1997; Wouts et al., 1987). The studies included in the pooled analysis to compare the clinical validity of SCSA and CSP in couples trying to conceive reported different rates of prevalence, and so PPV and NPV were calculated with a fixing value of prevalence according to the theoretical prevalence of male infertility in this clinical situation (7.5%). Under these circumstances, no significant differences were observed between SCSA and CSP.

Tests are more useful when the pre-test probability is 50%. Numerical changes in the post-test column of the nomogram are greater when the starting point in the pre-test column is at 50% than elsewhere. Many semen tests have been analysed for prevalence rates of around 50%, at which the performance of the test is maximized (case-control studies in subfertile couples). However, in this analysis of case-control studies in which this prevalence was applied to determine the predictive value of SCSA and CSP, not even in this favourable clinical scenario was there found to be any greater clinical benefit from SCSA than from CSP. The case-control studies referred to above commonly suf-

fer from selection bias, known as verification, work-up or referential bias, as the reproductive capacity of the couples included is more extreme (there is a higher level of reproductive capacity among the controls and a lower one among the cases) than among the couples included in prospective studies designed to analyse the validity of a sperm test among subfertile couples who conceive spontaneously and those who do not. The effect of this bias has been estimated to represent an increase of at least three times in the power of a test (Horvath and Pewsner, 2004; Moons and Grobbee, 2002).

In order to analyse the clinical validity of a test for couples given IUI treatment, it is necessary to determine the magnitude of change from a clinician's initial suspicion of non-pregnancy using IUI to the likelihood of non-pregnancy using IUI after the test result (post-test probability). In the IUI programme, the pre-test probability of non-pregnancy in the total subfertile population is around 85% (Nyboe et al., 2008; Ombelet et al., 2003). This high prevalence means that even when uPPV, PPV-85 and LR+ are higher for SCSA than for sperm morphology, the differences in post-test probabilities of non-pregnancy using IUI are not very large, in clinical terms, amounting to around 97% after an abnormal SCSA result and 91% after an abnormal sperm morphology result. The populations included in the SCSA and artificial insemination studies were unexplained infertile couples, while those in the sperm morphology and IUI studies were whole population. These populations are different and not comparable. However, as far as is known, there have been no studies focusing solely on CSP for couples treated with IUI for unexplained infertility. While it is obviously true that due to pre-treatment sperm screening and the exclusion of couples with male factor sterility, CSP is unlikely to be predictive of IUI success in couples with unexplained sterility, it is not known whether other variables such as sperm concentration after preparation or the number of inseminated motile spermatozoa may influence unexplained infertility, as other authors have observed with respect to infertile couples treated with IUI (Branigan et al., 1999; Merviel et al., 2008; Royère, 2004; Van Voorhis et al., 2001). Before recommending the use of SCSA for couples who are to receive IUI treatment for unexplained infertility, it is important to compare its clinical validity with that of other, more economical sperm function tests, such as the hypo-osmotic swelling test (Check, 2005; Misro and Chaki, 2008) or 24-h survival test (Branigan et al., 1999).

Various methods may be applied to increase the diagnostic utility of a test before replacing it with another one of unproven utility. First, by modifying the reference values of the variable being analysed by the diagnostic test: in the case of CSP, by decreasing the value of the lower reference limits below which a result is considered abnormal. Thus, when a test result is abnormal, it is more likely that a sperm alteration is responsible for the infertility, and therefore requires assisted reproduction treatment. Different authors have suggested that the lower reference limits established by the WHO (1999) should be revised downwards (Guzick et al., 2001; Haugen et al., 2006; Nallella et al., 2006). This would necessarily improve the LR+ of the CSP. Second, the higher the pre-test probability of the disease, the higher the post-test

probability, irrespective of the test result: therefore, a precise assessment of the possibility of a disease or disorder will be far more important than the likelihood ratios derived from expensive tests. Third, when an interval LR is used: clinicians naturally assume that with a sperm count of 1 million/ml in a man from a couple who have been trying to conceive for 12 months without success is more likely to present male factor infertility than if the sperm count is 19 million/ml (even though both values are low according to the WHO manual). An interval LR based on internal ranges helps clinicians quantify the differences in diagnostic effects that are instinctively recognized. A separate LR is calculated for every level (interval) of the test result, and this interval LR provides more clinically useful information than when the data are presented in a dichotomized format (Brown and Reeves, 2003; Grimes and Schulz, 2005).

This study has several limitations. First, it has assumed that a low false positive rate is more important than a low false negative rate in the test diagnosis of male factor infertility from a physiological and clinical point of view. But it is not known for sure that this standpoint coincides with that of the patient or with the opinion of all clinicians. In a study of ovarian reserve testing, the patients preferred to avoid false negatives rather than false positives (Mol et al., 2006). This aspect is currently being investigated in the study centres' own patients. Furthermore, large discrepancies between clinics have been described concerning the interpretation of semen parameters (Aguilar et al., 2008; Van der Steeg et al., 2006; Wiegerinck et al., 1999). Second, the severe exclusion conditions for items, which are standard in this type of study, undoubtedly have an extremely selective effect regarding such items. Nevertheless, the above-mentioned heterogeneity among items reinforces the view that the possible selection bias cannot be very large, and thus it may be assumed that the results obtained are valid. Third, the heterogeneity found among the different studies may be indicative of different methodologies being applied. Coppus et al. (2006) showed that the majority of published studies of diagnostic accuracy still have methodological flaws in design or analysis, or provide results with limited practical applicability, despite the publication of STARD, an adequate methodological framework for diagnostic test evaluations (Bossuyt et al., 2003). The heterogeneity that has been identified may be caused by the use of different cut-off values by different studies or be derived from the definition of fertile and infertile populations (Dinnes et al., 2005). The scant number of studies included in the different scenarios prevented us from carrying out a meta-regression analysis. Nevertheless, the technique that was applied for calculating the pooled attributes of the test is the one that is least affected by heterogeneity among studies (Dersimonian and Laird, 1987).

In summary, there is a substantial and recent body of work on the use of DNA staining tests interpreted as DNA damage and DNA fragmentation tests in the diagnostic work-up of male factor infertility. However, the research performed is characterized by substantial heterogeneity and the absence of direct comparisons with other (cheaper and more accessible) seminal tests, such that its utility in

clinical practice is limited. It is, therefore, not possible to determine whether the addition of new test strategies, including the use of SCSA, might lead to an improvement in referral practices. Any further work with existing primary research would require the use of individual patient data. Variations in test accuracy with changes in the clinical characteristics of patients or changes in the application of the diagnostic tests could then be investigated. In addition, there is a need for investigation of the test performance of current normal practice for diagnosing male factor infertility in different clinical scenarios (e.g. the primary care setting). Assessing diagnostic tests without following STARD (Bossuyt et al., 2003) and in isolation, outside the clinical context, is of limited use apart from assessing a test's potential usefulness in phase I and II diagnostic study designs (Knotterus and Van Weel, 2002).

## References

- Aguilar, J., Alvarez, C., Morancho-Zaragoza, J., et al., 2008. Quality specifications for seminal parameters based on clinicians' opinions. *Scand. J. Clin. Lab. Invest.* 68, 68–76.
- Alvarez, C., Castilla, J.A., Martinez, L., et al., 2003. Biological variation in seminal parameters in healthy subjects. *Hum. Reprod.* 18, 2082–2088.
- Bartoov, B., Eltes, F., Pansky, M., et al., 1993. Estimating fertility potential via semen analysis data. *Hum. Reprod.* 8, 65–70.
- Biggerstaff, B.J., 2000. Comparing diagnostic tests: a simple graphic using likelihood ratios. *Stat. Med.* 19, 649–663.
- Bonde, J.P., Ernst, E., Jensen, T.K., et al., 1998. Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *Lancet* 352, 1172–1177.
- Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., et al., 2003. Standards for reporting of diagnostic accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin. Chem.* 49, 7–17.
- Bossuyt, P.M., Lijmer, J.G., Mol, B.W., 2000. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 356, 1844–1847.
- Branigan, E.F., Estes, M.A., Muller, C.H., 1999. Advanced semen analysis: a simple screening test to predict intrauterine insemination success. *Fertil. Steril.* 71, 547–551.
- Brown, M.D., Reeves, M.J., 2003. Evidence-based emergency medicine/skills for evidence-based emergency care. Interval likelihood ratios: another advantage for the evidence-based diagnostician. *Ann. Emerg. Med.* 42, 292–297.
- Bungum, M., Humaidan, P., Axmon, A., et al., 2007. Sperm DNA integrity assessment in prediction of assisted reproduction technology outcome. *Hum. Reprod.* 22, 174–179.
- Bungum, M., Humaidan, P., Spanò, M., et al., 2004. The predictive value of sperm chromatin structure assay (SCSA) parameters for the outcome of intrauterine insemination, IVF and ICSI. *Hum. Reprod.* 19, 1401–1408.
- Check, J.H., 2005. Sperm may be associated with subfertility independent of oocyte fertilization. *Clin. Exp. Obstet. Gynecol.* 32, 5–8.
- Chohan, K.R., Griffin, J.T., Lafromboise, M., et al., 2006. Comparison of chromatin assays for DNA fragmentation evaluation in human sperm. *J. Androl.* 27, 53–59.
- Cocuzza, M., Sikka, S.C., Athayde, K.S., et al., 2007. Clinical relevance of oxidative stress and sperm chromatin damage in male infertility: an evidence based analysis. *International braz j urol: Official Journal of the Brazilian Society of Urology* 33, 603–621.

- Collins, J.A., Barnhart, K.T., Schlegel, P.N., 2008. Do sperm DNA integrity tests predict pregnancy with in-vitro fertilization? *Fertil. Steril.* 89, 823–831.
- Coppus, S.F., Van der Veen, F., Bossuyt, P.M., et al., 2006. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertil. Steril.* 86, 1321–1329.
- Dersimonian, R., Laird, N., 1987. Meta-analysis in clinical trials. *Control. Clin. Trials* 7, 177–188.
- Dinnes, J., Deeks, J., Kirby, J., et al., 2005. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol. Assess.* 9, 1–113.
- Dujardin, B., Van den Ende, J., Van Gompel, A., et al., 1994. Likelihood ratios: a real improvement for clinical decision making? *Eur. J. Epidemiol.* 10, 29–36.
- Erenpreiss, J., Elzanaty, S., Giwercman, A., 2008. Sperm DNA damage in men from infertile couples. *Asian J. Androl.* 10, 786–790.
- Erenpreiss, J., Spanò, M., Erenpreiss, J., 2006. Sperm chromatin structure and male fertility: biological and clinical aspects. *Asian J. Androl.* 8, 11–29.
- Evenson, D.P., Wixon, R., 2008. Data analysis of two in-vivo fertility studies using sperm chromatin structure assay-derived DNA fragmentation index vs. pregnancy outcome. *Fertil. Steril.* 90, 1229–1931.
- Evenson, D.P., Wixon, R., 2006. Meta-analysis of sperm DNA fragmentation using the sperm chromatin structure assay. *Reprod. BioMed. Online* 12, 466–472.
- Evenson, D.P., Jost, L.K., Marshall, D., et al., 1999. Utility of the sperm chromatin structure assay as a diagnostic and prognostic tool in the human fertility clinic. *Hum. Reprod.* 14, 1039–1049.
- Fagan, T.J., 1975. Nomogram for Bayes theorem. *N. Engl. J. Med.* 293, 257.
- Gart, J.J., Zweifel, J.R., 1967. On the bias of various estimators of the logit and its variance with applications to quantal bioassay. *Biometrika* 54, 181–187.
- Glazener, C.M., Coulson, C., Lambert, P.A., 1987. The value of artificial insemination with husband's semen in infertility due to failure of postcoital sperm–mucus penetration – controlled trial of treatment. *Br. J. Obstet. Gynaecol.* 94, 774–778.
- Greenhalgh, T., 1997. How to read a paper: papers that report diagnostic or screening tests. *Br. Med. J.* 315, 540–543.
- Greenland, S., 1987. Quantitative methods in the review of epidemiologic literature. *Epidemiol. Rev.* 9, 1–30.
- Grimes, D.A., Schulz, K.F., 2005. Refining clinical diagnosis with likelihood ratios. *The Lancet* 365, 1500–1505.
- Guzick, D.S., Overstreet, J.W., Factor-Litvak, P., et al., 2001. Sperm morphology, motility, and concentration in fertile and infertile men. *N. Engl. J. Med.* 345, 1388–1393.
- Haugen, T.B., Egeland, T., Magnus, O., 2006. Semen parameters in Norwegian fertile men. *J. Androl.* 27, 66–71.
- Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558.
- Horvath, A.R., Pewsner, D., 2004. Systematic reviews in laboratory medicine: principles, processes and practical considerations. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 342, 23–29.
- Jaeschke, R., Guyatt, G., Lijmer, J., 2002. Diagnostic tests. In: Guyatt, G., Rennie, D. (Eds.), *Users' Guides to Medical Literature: A Manual for Evidence-Based Clinical Practice*. AMA Press, Chicago, pp. 121–140.
- Jeyendran, R.S., Zaneveld, L.J., 1993. Controversies in the development and validation of new sperm assay. *Fertil. Steril.* 59, 726–728.
- Jialiang, L., Fine, J.P., Safdar, N., 2007. Prevalence-dependent diagnostic accuracy measures. *Stat. Med.* 26, 3258–3273.
- Karabinus, D.S., Gelety, T.J., 1997. The impact of sperm morphology evaluated by strict criteria on intrauterine insemination success. *Fertil. Steril.* 67, 536–541.
- Knottnerus, J.A., Van Weel, C., 2002. General introduction: evaluation of diagnostic procedures. In: Knottnerus, J.A. (Ed.), *The Evidence Base of Clinical Diagnosis*. BMA Books, London, pp. 1–16.
- Kvist, U., Björndahl, L., 2002. ESHRE Monographs: Manual on Basic Semen Analysis. Oxford University Press, Oxford.
- Lefèvre, L., Bedu-Addo, K., Conner, S.J., et al., 2007. Counting sperm does not add up any more: time for a new equation? *Reproduction* 133, 675–684.
- Lewis, S.E.M., 2007. Is sperm evaluation useful in predicting human fertility? *Reproduction* 134, 31–40.
- Lindheim, S.R., Barad, D.H., Zinger, M., et al., 1996. Abnormal sperm morphology is highly predictive of an intrauterine insemination pregnancy outcome during controlled ovarian hyperstimulation. *J. Assist. Reprod. Genet.* 13, 569–572.
- Matorras, R., Corcóstegui, B., Perez, C., et al., 1995. Sperm morphology analysis (strict criteria) in male infertility is not a prognostic factor in intrauterine insemination with husband's sperm. *Fertil. Steril.* 63, 608–611.
- Menkveld, R., Wong, W.Y., Lombard, C.J., et al., 2001. Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort towards standardization of in-vivo thresholds. *Hum. Reprod.* 16, 1165–1171.
- Merviel, P., Heraud, M.H., Grenier, N., et al., 2008. Predictive factors for pregnancy after intrauterine inseminations (IUI): an analysis of 1038 cycles and a review of the literature. *Fertil. Steril.* [e-publication ahead of print].
- Misro, M.M., Chaki, S.P., 2008. Development of a rapid, sensitive, and reproducible laboratory test kit for the assessment of plasma membrane integrity of human sperm. *Fertil. Steril.* 89, 223–227.
- Mol, B.W., Verhagen, T.E., Hendriks, D.J., et al., 2006. Value of ovarian reserve testing before IVF: a clinical decision analysis. *Hum. Reprod.* 21, 1816–1823.
- Montanaro Gauci, M., Kruger, T.F., Coetzee, K., et al., 2001. Stepwise regression analysis to study male and female factors impacting on pregnancy rate in an intrauterine insemination programme. *Andrologia* 33, 135–141.
- Moons, K.G., Grobbee, D.E., 2002. Diagnostic studies as multivariable, prediction research. *J. Epidemiol. Community Health* 56, 337–338.
- Nallella, K.P., Sharma, R.K., Aziz, N., et al., 2006. Significance of sperm characteristics in the evaluation of male infertility. *Fertil. Steril.* 85, 629–634.
- Nicopoulos, J.D., Gilling-Smith, C., Almeida, P.A., et al., 2008. Sperm DNA fragmentation in subfertile men: the effect on the outcome of intracytoplasmic sperm injection and correlation with sperm variables. *BJU Int.* 101, 1553–1560.
- Niederberger, C., 2003. Responses to semen analysis CART report. *J. Androl.* 24, 329–331.
- Nyboe, A., Goossens, V., Ferraretti, A.P., et al., 2008. European IVF-monitoring (EIM) Consortium; European Society of Human Reproduction, Embryology (ESHRE). Assisted reproductive technology in Europe, 2004: results generated from European registers by ESHRE. *Hum. Reprod.* 23, 756–771.
- Ombelet, W., Deblaere, K., Bosmans, E., et al., 2003. Semen quality and intrauterine insemination. *Reprod. BioMed. Online* 7, 485–492.
- Ombelet, W., Wouters, E., Boels, L., et al., 1997a. Sperm morphology assessment: diagnostic potential and comparative analysis of strict or WHO criteria in a fertile and a subfertile population. *J. Androl.* 20, 367–372.

- Ombelet, W., Vandeput, H., Van de Putte, G., et al., 1997b. Intrauterine insemination after ovarian stimulation with clomiphene citrate: predictive potential of inseminating motile count and sperm morphology. *Hum. Reprod.* 7, 1458–1463.
- Practice Committee of American Society for Reproductive Medicine, 2006. The clinical utility of sperm DNA integrity testing. *Fertil. Steril.* 86, 35–37.
- Roldan-Nofuentes, J.A., Luna del Castillo, J.D., 2007. Comparison of the likelihood ratios of two binary diagnostic tests in paired designs. *Stat. Med.* 26, 4179–4201.
- Royère, D., 2004. Intrauterine insemination in human: state-of-the-art. *Gynécolog., Obstétrique Fertil.* 32, 873–879.
- Sackett, D.L., Deeks, J.J., Altman, D.G., 1996. Down with odds ratios! *Evidence-based Medicine* 1, 164–166.
- Sackett, D.L., Haynes, R.B., Guyatt, G.H., et al., 1991. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Williams and Wilkins, Boston, pp. 51–69.
- Saleh, R.A., Agarwal, A., Nada, E.A., et al., 2003. Negative effects of increased sperm DNA damage in relation to seminal oxidative stress in men with idiopathic and male factor infertility. *Fertil. Steril.* 79, 597–605.
- Saleh, R.A., Agarwal, A., Nelson, D.R., et al., 2002. Increased sperm nuclear DNA damage in normozoospermic infertile men: a prospective study. *Fertil. Steril.* 78, 313–318.
- Sergerie, M., Laforest, G., Bujan, L., et al., 2005. Sperm DNA fragmentation: threshold value in male fertility. *Hum. Reprod.* 20, 3446–3451.
- Snick, H.K., Snick, T.S., Evers, J.L., et al., 1997. The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum. Reprod.* 12, 1582–1588.
- Spanò, M., Bonde, J.P., Hjøllund, H.I., et al., 2000. Sperm chromatin damage impairs human fertility. The Danish First Pregnancy Planner Study Team. *Fertil. Steril.* 73, 43–50.
- Swan, S.H., 2006. Does our environment affect our fertility? Some examples to help reframe the question. *Semin. Reprod. Med.* 24, 142–146.
- Toner, J.P., Mossad, H., Grow, D.R., et al., 1995. Value of sperm morphology assessed by strict criteria for prediction of the outcome of artificial (intrauterine) insemination. *Andrologia* 27, 143–148.
- Van der Steeg, J.W., Steures, P., Eijkemans, M.J., et al., 2006. Do clinical prediction models improve concordance of treatment decisions in reproductive medicine? *BJOG: An Int. J. Obstet. Gynaecol.* 113, 825–831.
- Van der Steeg, J.W., Steures, P., Hompes, P.G., et al., 2005. Investigation of the infertile couple: a basic fertility work-up performed within 12 months of trying to conceive generates costs and complications for no particular benefit. *Hum. Reprod.* 20, 2672–2674.
- Van Voorhis, B.J., Barnett, M., Sparks, A.E., et al., 2001. Effect of the total motile sperm count on the efficacy and cost-effectiveness of intrauterine insemination and in-vitro fertilization. *Fertil. Steril.* 75, 661–668.
- Van Waart, J., Kruger, T.F., Lombard, C.J., et al., 2001. Predictive value of normal sperm morphology in intrauterine insemination (IUI): a structured literature review. *Hum. Reprod.* 7, 495–500.
- Wiegerinck, M.A., Bongers, M.Y., Mol, B.W., et al., 1999. How concordant are the estimated rates of natural conception and in-vitro fertilization/embryo transfer success? *Hum. Reprod.* 14, 689–693.
- World Health Organization, 1999. *Laboratory Manual for the Examination of Human Semen and Semen–Cervical Mucus Interaction*, fourth ed. Cambridge University Press, Cambridge.
- Wouts, M.H., Duisterhout, J.S., Kuik, D.J., et al., 1987. The chance of spontaneous conception for the infertile couple referred to an academic clinic for reproductive endocrinology and fertility in The Netherlands. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 26, 243–250.
- Zini, A., Boman, J.M., Belzile, E., et al., 2008. Sperm DNA damage is associated with an increased risk of pregnancy loss after IVF and ICSI: systematic review and meta-analysis. *Hum. Reprod.* 23, 2663–2668.
- Zhou, X.H., Obuchowski, N.A., McClish, D.K., 2002. *Statistical Methods in Diagnostic Medicine*. Wiley, New York.

*Declaration: The authors report no financial or commercial conflicts of interest.*

Received 24 February 2009; refereed 9 April 2009; accepted 23 September 2009.