Can history, ultrasound, or ELISA chlamydial antibodies, alone or in combination, predict tubal factor infertility in subfertile women?

Susan Logan1,3, Rafet Gazvani1, Hamish McKenzie2, Allan Templeton1 and Siladitya Bhattacharya1

1Department of Obstetrics and Gynaecology, Aberdeen Maternity Hospital, Foresterhill, Aberdeen AB25 2ZD and 2Department of Medical Microbiology, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK

3To whom correspondence should be addressed. E-mail: ogy167@abdn.ac.uk

BACKGROUND: This study aimed to determine whether medical history, transvaginal ultrasound (TVU) or Chlamydia trachomatis antibody testing (CAT), alone or in combination, could provide a non-invasive, clinically useful screening test for predicting tubal factor infertility (TFI) in subfertile women. METHODS: Prior to tubal evaluation, relevant medical history, TVU findings, and enzyme-linked immunosorbet assay (ELISA) IgG CAT results were collected. Sensitivity, specificity, likelihood ratios (LR) and accuracy for predicting TFI, as determined by laparoscopy and dye hydrotubation, were calculated for each test alone, and in parallel and series combination. RESULTS: Thirty per cent (63/207) were diagnosed with TFI. The highest sensitivity (67%, 95% CI: 54±77) included any positive test, yet missed one in three women with TFI. The highest specificity (100%, 95% CI: 97±100) required all three tests positive, but identified only three women. Only the combination of CAT and TVU rated as a good clinical test, but confidence intervals were wide due to the small numbers affected. The combination of CAT or TVU and CAT alone reported the highest accuracy (73%, 95% CI: 66±78), misdiagnosing one in four women. CONCLUSION: Medical history, TVU appearances, and ELISA IgG CAT alone, or in combination, failed to predict accurately TFI in subfertile women.

Key words: Chlamydia trachomatis antibody test/medical history/subfertility/transvaginal ultrasound/tubal infertility

Introduction

The pivotal role of chlamydial pelvic inflammatory disease (PID) in the aetiology of tubal factor infertility (TFI) is well recognized (Chief Medical Officer, 1998; Paavonen and Eggert-Kruse, 1999). During 1989–1999, tubal factor infertility was diagnosed in 22% of couples attending the Aberdeen Fertility Centre. For those with secondary infertility, tubal disease accounted for 40% of all cases (S.Bhattacharya, personal communication). As 93% of couples fall pregnant within a year (Evers, 2002), it would be useful to have a non-invasive test that could reliably predict tubal disease at 12 months.

Laparoscopy is the gold standard for detecting pelvic adhesions and endometriosis (Corson, 1977; Siegler, 1983). However, it is invasive, costly, carries the risk of both major and minor complications (Bateman et al., 1996; Jansen et al., 1997), and, in most cases, will find no abnormality. Compared with laparoscopy, hysterosalpingography (HSG) is less costly, and less risky in terms of anaesthetic complications and organ and blood vessel damage. However, it is invasive, uncomfortable, carries a risk of ionization, and is poor at diagnosing peritubal adhesions (Swart et al., 1995). False positive results can occur due to tubal spasm, dissimilar tubal filling pressure, excessive viscosity, faulty technique, or misinterpreted films (Dabekausen et al., 1994; Swart et al., 1995). This then further delays definitive tubal assessment.

While pelvic inflammatory disease is the major cause of tubal factor infertility, other factors may also be contributory. Women with a past history of sexually transmitted infection (Grodstein et al., 1993), pelvic inflammatory disease (Grodstein et al., 1993), intrauterine contraceptive device (IUCD) use (Cramer et al., 1985; Grodstein et al., 1993) and pelvic/lower abdominal surgery (Macmillan, 2000) are at increased risk of tubal damage. All of these risk factors can be elicited by medical history.

Transvaginal ultrasonography is used extensively in fertility treatment. The Fallopian tube is not usually seen by transvaginal ultrasound (TVU), but inflammatory processes of the tube such as hydrosalpinx, pyosalpinx or tubo-ovarian abscess have typical sonographic appearances. Previous reports had favourably demonstrated an accuracy of 97% in the screening of endometriomata (Mais et al., 1993).
Serological tests are used to demonstrate evidence of past infection by a variety of organisms. The association between chlamydial IgG antibody titres and TFI was first reported by Punnonen et al. (1979). The clinical benefits of antibody testing lie in its potential to avoid invasive diagnostic procedures in those with normal tubal function, or expedite investigations in those thought to be at increased risk of TFI. In theory, the presence of chlamydial antibodies can be determined at low cost and burden to the woman.

Each of these approaches is used routinely by fertility clinics to screen for TFI, but their actual performance as screening tests in unselected subfertile women has been poorly evaluated. When testing or screening for acute chlamydial infection, confirmatory testing improves test performance. Parallel combination testing requires any test to be positive and improves sensitivity, while series combination testing requires all tests to be positive and improves specificity (Griner et al., 1981). Potentially, combination testing could be used as a screen for upper genital tract damage.

The aim of this study was to determine whether medical history, transvaginal ultrasound or the measurement of chlamydial antibodies, alone or in combination, could predict tubal disease in subfertile women.

Materials and methods

Consecutive women attending a dedicated fertility clinic were recruited when referred for tubal evaluation. The timing of tubal assessment was made on clinical grounds, but all had undergone ovulation and semen analysis. Women who were physically unsuitable for laparoscopy, and those who had previously undergone tubal assessment in the form of laparoscopy, HSG or tubal surgery (including ectopic pregnancy), were excluded. Approval for the study was received from the local ethics committee.

Demographic data and medical history were recorded on a standardized datasheet. Screening for acute lower genital tract chlamydial infection was initially by enzyme immunoassay (EIA) (Microtrak II; Behring Diagnostics, UK) and subsequently by BDProbeTec™ (Becton Dickinson, UK) nucleic acid amplification assay. All positive EIA results were confirmed by direct immuno-fluorescence (DFA) by the Syva Microtrak Direct Specimen Test (Dade Behring Diagnostics Ltd, UK). The positive BDProbeTec™ results were confirmed by retesting the sample.

A transvaginal scan was performed by one of four experienced individuals using a Siemens Sonoline SI-250 scanner. Pelvic findings were recorded on a standardized datasheet and categorized as suspicious if adnexal appearances were abnormal and a hydrosalpinx could not be ruled out.

Five millilitres of venous blood was drawn for the measurement of serum Chlamydia trachomatis-specific IgG antibodies by the peptide-based enzyme-linked immunosorbent assay (pELISA; Medac, Germany). Results were interpreted according to the manufacturer’s instructions, issuing a positive, equivocal, or negative result. As C. trachomatis antibody testing was not routinely performed in our clinic, the samples were pooled and assayed at a later date. Women with equivocal results were therefore not retested. A positive IgG result was indicative of a past or persisting infection. If the result was equivocal, past or persisting infection could not be excluded and therefore equivocal and positive results were pooled together.

Laparoscopy and dye hydrodistubation was performed without regard to the Chlamydia antibody test (CAT) result. The surgeon was also blinded to the CAT result. Tubal patency testing was done with methylene blue dye. The operative findings were coded on a standardized operation note. Tubal disease was confirmed by the presence of adhesions involving the Fallopian tubes and ovaries, clubbing of the Fallopian tubes, hydrosalpinges, or obstruction to the flow of dye. Endometriosis was categorized separately as per the Revised American Fertility Society classification (Anonymous, 1985). Women whose staging equated to Stage III (moderate) and Stage IV (severe) were categorized with endometriosis as a cause for their subfertility.

The prevalence of tubal infertility in women attending the Aberdeen Fertility Clinic was estimated at 20–30%. Assuming estimates of sensitivity and specificity of 80 and 90% respectively, we aimed to recruit a minimum of 200 women, 60 of whom would have tubal disease diagnosed at laparoscopy. This sample size would enable an estimation of sensitivity to within 10%, and specificity to within 5%.

Clinical and laboratory data were analysed using Statistical Package for Social Sciences (SPSS) software. Estimates of sensitivity, specificity, likelihood ratio (positive and negative), and accuracy were calculated for each test. Confidence intervals (95% CI) were reported in order for statistical comparisons to be made.

Results

A total of 235 consecutive women were approached. None declined to participate. Twenty-eight women were excluded for the following reasons: previous laparoscopy (n = 4), previous tubal surgery (n = 3), severe male factor (n = 1), body mass index deemed too high for laparoscopy (n = 1), pregnant prior to laparoscopy (n = 5), did not attend laparoscopy (n = 1), and antibody test result not available (n = 13).

The study population comprised 207 women. Overall, the mean (SD) age was 31 (5) years with a range of 20–42 years. Mean (SD) duration of subfertility was 28 (17) months, range 5–120. Subfertility was reported as primary and secondary by 115 (56%) and 92 (44%) women respectively. Following laparoscopy, 100 (48%), 63 (30%), 36 (17%), 18 (9%) and 11 (5%) were categorized with unexplained, tubal, male, anovulation and endometriosis as causes for their subfertility. Ten per cent had combined causes.

Medical history associated with tubal factor infertility

Surgery

Seventy-nine (38%) women recalled a past history of pelvic surgery. Sixty-nine (33%) women had undergone obstetric-and/or gynaecology-related procedures, 16 (8%) women gave a history of lower abdominal surgery, and six (3%) women reported both. The obstetric- and/or gynaecology-related operations included the following: evacuation of retained products of conception—surgical (n = 30) and medical (n = 4); termination of pregnancy—surgical (n = 23) and medical (n = 4); lower uterine Caesarean section (n = 8); manual removal of placenta (n = 1); dilatation and curettage (n = 3); cone biopsy (n = 1); large loop excision of the transformation zone (n = 1); and hysteroscopy (n = 1). The lower abdominal operations were appendicectomy (n = 17) and peritoneal dialysis (n = 1). Sixteen (8%) women reported more than one operation.
Pelvic inflammatory disease (PID)
Twelve (6%) women reported a past history of PID.

Sexually transmitted disease (STD)
Twelve (6%) women reported a past history of STD that included the following:
- Chlamydia (n = 5);
- Genital warts (n = 5);
- Both (n = 1);
- Genital herpes (n = 3).

Intrauterine contraceptive device (IUCD)
Ten (5%) women reported a past history of IUCD use.

Acute lower genital tract chlamydial infection
A total of 188/207 (91%) women were screened for acute chlamydial infection, 180 (96%) by EIA and eight (4%) by BDProbeTec® assay. Two women (1%), screened by EIA, were positive for C. trachomatis.

Transvaginal ultrasound scan
Abnormalities of the uterus, ovaries and Fallopian tubes were visualized in four (2%), 26 (13%), and nine (4%) women respectively. The uterine abnormalities were thought to represent fibroids in all cases. The ovarian abnormalities were thought to represent ovarian cysts (16), polycystic ovaries (n = 7), endometrioma(ta) (n = 3), and in one woman the ovaries were not visualized. The tubal abnormalities were reported as a unilateral hydrosalpinx, but in three cases, the ultrasonic features were recorded as either an ovarian cyst or a hydrosalpinx. Overall, 25 (12%) women had suspicious adnexal masses on scan, where a hydrosalpinx could not be excluded with certainty.

Chlamydial IgG antibody testing
The chlamydial IgG antibody result was reported as negative in 167 (81%) women, equivocal in seven (3%) women, and positive in 33 (16%) women.

Laparoscopy and dye hydrotubation
The majority of the operations, 172 (83%), were performed by a single sub-speciality trainee in reproductive medicine. The remainder were undertaken by a consultant specializing in subfertility and a senior specialist registrar. Pelvic adhesions were documented in 34 (16%) women. With regards to tubal damage, 21 (10%) women had unilateral or bilateral clubbed tubes. Unilateral or bilateral hydrosalpinges were seen in 11 (5%) women. One woman (0.5%) was diagnosed with a chronic ectopic pregnancy as a result of the findings at laparoscopy. Regarding tubal obstruction, 16 (8%) women and three (1%) women had unilateral and bilateral delayed dye spill respectively. All had evidence of salpingitis (tubal inflammation, fibrosis or hydrosalpinx) and/or adhesions. Unilateral and bilateral obstruction to the spill of dye was seen in 31 (15%) women and 18 (9%) women respectively. All had evidence of salpingitis (tubal inflammation, fibrosis or hydrosalpinx) and/or adhesions. Three women with obstruction, but no other abnormality, went onto hysterosalpingogram. Two had bilateral dye flow and one continued to exhibit bilateral obstruction. Overall, 63 (30%) of the study population were diagnosed with tubal factor infertility, with a proportion exhibiting more than one of four pathological findings highlighted above. Endometriosis was diagnosed in 74 (36%) women. An American Fertility Society grading of Stage 1 (minimal),
Stage II (mild), Stage III (moderate) and Stage IV (severe) was made in 54 (73%), nine (12%), nine (12%) and two (3%) women respectively.

In terms of predicting TFI, the performance of medical history, TVU and C. trachomatis IgG antibodies, alone and in combination, are summarized in Table I, Table II and Table III.

Sensitivity measured how good the test(s) were at identifying subfertile women with TFI. Regarding medical history, individual sensitivities ranged from 6 to 46%, increasing to 54% if one or more clinical features were present (Table I). A suspicious TVU carried a sensitivity of only 13%, compared with 37% by testing positive or equivocal by CAT (Table I).

Table II. Performance of tests, in parallel, for diagnosing tubal factor infertility (TFI)

<table>
<thead>
<tr>
<th>No. TFI (n = 63)</th>
<th>No. no TFI (n = 144)</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
<th>Accuracy (%) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One or more positive clinical features or suspicious TVU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>36</td>
<td>72</td>
<td>57</td>
<td>(45–69)</td>
<td>50</td>
<td>(42–58)</td>
</tr>
<tr>
<td>–</td>
<td>27</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive clinical feature(s) or positive/ equivocal Ct Ab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>41</td>
<td>67</td>
<td>65</td>
<td>(53–76)</td>
<td>54</td>
<td>(45–61)</td>
</tr>
<tr>
<td>–</td>
<td>22</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive or equivocal Ct Ab or suspicious TVU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>23</td>
<td>17</td>
<td>35</td>
<td>(25–48)</td>
<td>88</td>
<td>(82–92)</td>
</tr>
<tr>
<td>–</td>
<td>40</td>
<td>127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive clinical feature(s) or positive/ equivocal Ct Ab or suspicious TVU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>42</td>
<td>78</td>
<td>67</td>
<td>(54–77)</td>
<td>46</td>
<td>(38–54)</td>
</tr>
<tr>
<td>–</td>
<td>21</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LR+ = positive likelihood ratio; LR− = negative likelihood ratio; CI = confidence interval; TVU = transvaginal ultrasound; Ct = C. trachomatis; Ab = antibodies.

Table III. Performance of tests, in series, for diagnosing tubal factor infertility (TFI)

<table>
<thead>
<tr>
<th>No. TFI (n = 63)</th>
<th>No. no TFI (n = 144)</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
<th>Accuracy (%) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One or more positive clinical features and suspicious TVU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>(4–19)</td>
<td>97</td>
<td>(92–99)</td>
</tr>
<tr>
<td></td>
<td>57</td>
<td>139</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive clinical feature(s) and positive/ equivocal Ct Ab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>12</td>
<td>27</td>
<td>(18–39)</td>
<td>92</td>
<td>(86–95)</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>132</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive or equivocal Ct Ab and suspicious TVU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>(3–15)</td>
<td>99</td>
<td>(96–99)</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive clinical feature(s) and positive/ equivocal Ct Ab and suspicious TVU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>(2–13)</td>
<td>100</td>
<td>(97–100)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LR+ = positive likelihood ratio; LR− = negative likelihood ratio; CI = confidence interval; TVU = transvaginal ultrasound; Ct = C. trachomatis; Ab = antibodies; NE = not estimable.

Stage II (mild), Stage III (moderate) and Stage IV (severe) was made in 54 (73%), nine (12%), nine (12%) and two (3%) women respectively.

In terms of predicting TFI, the performance of medical history, TVU and C. trachomatis IgG antibodies, alone and in combination, are summarized in Table I, Table II and Table III. Sensitivity measured how good the test(s) were at identifying subfertile women with TFI. Regarding medical history, individual sensitivities ranged from 6 to 46%, increasing to 54% if one or more clinical features were present (Table I). A suspicious TVU carried a sensitivity of only 13%, compared with 37% by testing positive or equivocal by CAT (Table I). Table II shows that combining two or three of the tests, in parallel, either maintained the same sensitivity or significantly increased it. Table III shows that combining two or three tests, in series, either maintained the same sensitivity or significantly decreased it. The highest sensitivity reported was 67% (95% CI: 54–77) in the test combination that included positive clinical feature(s) or suspicious TVU or positive/equivocal CAT.

Specificity measured how good the test(s) were at identifying subfertile women without TFI. The specificity of individual clinical features ranged from 65 to 97%, decreasing to 58% if one or more clinical features were present (Table I). A suspicious TVU carried a specificity of 88%, identical to testing positive or equivocal by CAT (Table I). Table II shows that combining two or three tests in parallel either maintained the same specificity or significantly decreased it. Table III shows that combining two or three tests in series either maintained the same specificity or significantly increased it.
Table IV. Studies comparing *C. trachomatis* IgG antibody testing to laparoscopy findings

<table>
<thead>
<tr>
<th>Authors and assay</th>
<th>No. of women</th>
<th>Prevalence of TFI (%)</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
<th>Accuracy (%) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Veenemans and van der Linden (2002), IF</td>
<td>145</td>
<td>24</td>
<td>80 (64–90)</td>
<td>55 (45–64)</td>
<td>1.8 (1.4–2.3)</td>
<td>0.4 (0.2–0.7)</td>
<td>61 (53–68)</td>
</tr>
<tr>
<td>Johnston <em>et al.</em> (2000), ELISA/MIF</td>
<td>80</td>
<td>45</td>
<td>81 (65–90)</td>
<td>86 (73–94)</td>
<td>5.9 (2.8–12.6)</td>
<td>0.2 (0.1–0.4)</td>
<td>84 (74–90)</td>
</tr>
<tr>
<td>Thomas <em>et al.</em> (2000), MIF</td>
<td>113</td>
<td>20</td>
<td>87 (68–96)</td>
<td>59 (49–69)</td>
<td>2.1 (1.6–2.8)</td>
<td>0.2 (0.1–0.6)</td>
<td>65 (55–73)</td>
</tr>
<tr>
<td>Land <em>et al.</em> (1998), MIF</td>
<td>227</td>
<td>30</td>
<td>19–55</td>
<td>73–99</td>
<td>2.0–29.7</td>
<td>0.6–0.8</td>
<td>NE</td>
</tr>
<tr>
<td>Eggert-Kruse <em>et al.</em> (1997), IF</td>
<td>1383</td>
<td>34</td>
<td>30 (26–35)</td>
<td>85 (82–87)</td>
<td>2.0 (1.6–2.4)</td>
<td>0.8 (0.8–0.9)</td>
<td>66 (64–69)</td>
</tr>
<tr>
<td>Meikle <em>et al.</em> (1994), IF</td>
<td>218</td>
<td>40</td>
<td>78 (68–86)</td>
<td>64 (56–72)</td>
<td>2.2 (1.7–2.8)</td>
<td>0.3 (0.2–0.5)</td>
<td>70 (63–75)</td>
</tr>
</tbody>
</table>

*Range of test performance results depending on Ab titre threshold level.
TFI = tubal factor infertility; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; CI = confidence interval; IF = immunofluorescence; MIF = microimmunofluorescence.

The highest specificity was 100% (95% CI: 97–100) in the test combination that included positive clinical feature(s) and suspicious TVU and positive/equivocal CAT.

The likelihood ratio of a positive test (LR+) reflected how much more likely a positive result was to be found in a woman with TFI. Only the combination of positive or equivocal CAT and a suspicious TVU rated as a good clinical test. All the rest rated either fairly or poorly.

The likelihood ratio of a negative test (LR−) reflected how much more likely a negative result was to be found in a person without TFI. All tests and combinations rated poorly.

Accuracy measured what proportion of all tests gave the correct results. The highest and lowest accuracy values were 73 and 52% respectively.

**Discussion**

Tubal factor infertility not associated with *C. trachomatis* infection is found in up to 25–50% of cases, which limits the use of CAT on its own (Rice and Schachter, 1991). Few studies have investigated suitable alternatives. Relevant medical history was chosen as it is routinely obtained and there is abundant literature correlating specific clinical features with tubal damage. The format of the case notes used in this study ensured uniformity and completeness of history taking. Recall bias could not be excluded, but would have been prevented to some extent as most general practitioners used the clinic’s proforma referral letter. In terms of test performance, medical history may have benefited from the addition of other factors such as a relevant partner history or symptoms of acute/chronic pelvic infection.

Transvaginal ultrasound was assessed as it is widely available and hydrosalpinges have characteristic ultrasonographic appearances. Labelling the TVU as suspicious, if any adnexal abnormality was seen, improved sensitivity at the expense of specificity. Endometriosis was separately categorized, but the test performance of TVU for diagnosing TFI or endometriosis was not significantly different (data not shown).

An ELISA CAT was chosen for antibody testing in this study as they are widely available and easy to use. Moreover, a meta-analysis reported that the test performance of ELISA was comparable to the microimmunofluorescence (MIF) antibody test (Mol *et al*., 1997). In this study, equivocal antibody results were not repeated as the assay was performed some months later. As an equivocal result cannot exclude past infection, they were included with the positive results. However, no significant difference was found between positive and positive/equivocal results (data not shown).

Test performance was evaluated in five different ways, with the three screening tests considered alone, in parallel combination, and in series combination. Sensitivity, specificity, likelihood ratios, and accuracy were highlighted as, compared with predictive values, they are not affected by prevalence and can be used to compare test outcome in populations with different prevalence rates. Confidence intervals were included to allow statistical comparisons to be made. The prevalence of tubal disease was within the range assumed for the power calculation, but as the sensitivity and specificity of CAT by ELISA were lower than anticipated, confidence intervals were wider. This resulted from less patient selection, blinding of the CAT result at laparoscopy, and tubal disease secondary to causes other than *Chlamydia trachomatis* infection. However, no studies to date have attempted a power calculation.

A number of studies have assessed the test performance of CAT (Table IV) and one paper has reviewed its accuracy by meta-analysis of papers published between 1981 and 1994 (Mol *et al*., 1997). Some studies compared CAT with HSG (Dabekausen *et al*., 1994; Meikle *et al*., 1994; Mol *et al*., 1997; Veenemans and van der Linden, 2002), but only those women who underwent laparoscopy are included in the test performance analysis reported that the test performance of ELISA was comparable to the microimmunofluorescence (MIF) antibody test (Mol *et al*., 1997). In this study, equivocal antibody results were not repeated as the assay was performed some months later. As an equivocal result cannot exclude past infection, they were included with the positive results. However, no significant difference was found between positive and positive/equivocal results (data not shown).

Test performance was evaluated in five different ways, with the three screening tests considered alone, in parallel combination, and in series combination. Sensitivity, specificity, likelihood ratios, and accuracy were highlighted as, compared with predictive values, they are not affected by prevalence and can be used to compare test outcome in populations with different prevalence rates. Confidence intervals were included to allow statistical comparisons to be made. The prevalence of tubal disease was within the range assumed for the power calculation, but as the sensitivity and specificity of CAT by ELISA were lower than anticipated, confidence intervals were wider. This resulted from less patient selection, blinding of the CAT result at laparoscopy, and tubal disease secondary to causes other than *Chlamydia trachomatis* infection. However, no studies to date have attempted a power calculation.

A number of studies have assessed the test performance of CAT (Table IV) and one paper has reviewed its accuracy by meta-analysis of papers published between 1981 and 1994 (Mol *et al*., 1997). Some studies compared CAT with HSG (Dabekausen *et al*., 1994; Meikle *et al*., 1994; Mol *et al*., 1997; Veenemans and van der Linden, 2002), but only those women who underwent laparoscopy are included in the test performance analysis reported that the test performance of ELISA was comparable to the microimmunofluorescence (MIF) antibody test (Mol *et al*., 1997). In this study, equivocal antibody results were not repeated as the assay was performed some months later. As an equivocal result cannot exclude past infection, they were included with the positive results. However, no significant difference was found between positive and positive/equivocal results (data not shown).

Test performance was evaluated in five different ways, with the three screening tests considered alone, in parallel combination, and in series combination. Sensitivity, specificity, likelihood ratios, and accuracy were highlighted as, compared with predictive values, they are not affected by prevalence and can be used to compare test outcome in populations with different prevalence rates. Confidence intervals were included to allow statistical comparisons to be made. The prevalence of tubal disease was within the range assumed for the power calculation, but as the sensitivity and specificity of CAT by ELISA were lower than anticipated, confidence intervals were wider. This resulted from less patient selection, blinding of the CAT result at laparoscopy, and tubal disease secondary to causes other than *Chlamydia trachomatis* infection. However, no studies to date have attempted a power calculation.

A number of studies have assessed the test performance of CAT (Table IV) and one paper has reviewed its accuracy by meta-analysis of papers published between 1981 and 1994 (Mol *et al*., 1997). Some studies compared CAT with HSG (Dabekausen *et al*., 1994; Meikle *et al*., 1994; Mol *et al*., 1997; Veenemans and van der Linden, 2002), but only those women who underwent laparoscopy are included in the test performance analysis reported that the test performance of ELISA was comparable to the microimmunofluorescence (MIF) antibody test (Mol *et al*., 1997). In this study, equivocal antibody results were not repeated as the assay was performed some months later. As an equivocal result cannot exclude past infection, they were included with the positive results. However, no significant difference was found between positive and positive/equivocal results (data not shown).

Test performance was evaluated in five different ways, with the three screening tests considered alone, in parallel combination, and in series combination. Sensitivity, specificity, likelihood ratios, and accuracy were highlighted as, compared with predictive values, they are not affected by prevalence and can be used to compare test outcome in populations with different prevalence rates. Confidence intervals were included to allow statistical comparisons to be made. The prevalence of tubal disease was within the range assumed for the power calculation, but as the sensitivity and specificity of CAT by ELISA were lower than anticipated, confidence intervals were wider. This resulted from less patient selection, blinding of the CAT result at laparoscopy, and tubal disease secondary to causes other than *Chlamydia trachomatis* infection. However, no studies to date have attempted a power calculation.
studies that specified a threshold titre for the CAT (Meikle et al., 1994; Land et al., 1998; Johnston et al., 2000; Thomas et al., 2000; Veenemans and van der Linden, 2002). Increasing the antibody titre reduced the number of women identified with TFI and increased confidence intervals, which were rarely reported. Endometriosis was not separately categorized in two studies (Meikle et al., 1994; Johnston et al., 2000) and the majority failed to blind the surgeon to the CAT result (Eggert-Kruse et al., 1997; Land et al., 1998; Johnston et al., 2000; Thomas et al., 2000; Veenemans and van der Linden, 2002). Overall, published accuracy of CAT was not significantly different to that reported here when the same definition of tubal disease was used (Dabekausen et al., 1994; Meikle et al., 1994; Eggert-Kruse et al., 1997; Land et al., 1998; Johnston et al., 2000; Thomas et al., 2000; Veenemans and van der Linden, 2002). Even Mol et al. (1997) concluded that while CAT has a discriminative capacity comparable to HSG, it provided no information on severity of TFI and therefore on prognosis.

Regarding medical history, only the study by Johnston et al. (2000) assessed clinical features alone and in combination with CAT (Table V). Their features included pelvic pain, cervical intraepithelial neoplasia, and abnormal clinical findings, but only reported on 79/80 women. Accuracy of their positive clinical features was similar to that reported here. The combination of any clinical features or a positive CAT had significantly better test performance in all categories, except specificity, than our parallel combination. This resulted from patient selection, confirmation by MIF, and possibly more predictive clinical feature choices.

Regarding the use of TVU, Ubaldi et al. (1998) compared transvaginal sonographic diagnosis with laparoscopic findings in 133 women undergoing laparoscopy for infertility, chronic pelvic pain or adnexal masses. Reported sensitivity and likelihood ratios were significantly better because of patient selection, fewer scanning personnel, and their definition of adhesions (Table V).

The use of CAT has been widely embraced by clinicians working in the field of subfertility and viewed as a simple and inexpensive screening test for TFI (Cahill and Wardle, 2002). However, little attention has been paid to its limitations. While the ELISA CAT was a significantly more discriminatory test than positive clinical feature(s) or TVU, on its own, it only rated as a fair screening test, labelling approximately one in four women inaccurately. A false positive test may unfairly stigmatize a women as having had a STD, with significant levels of emotional and social burden documented in the case of lower genital tract chlamydial infection (Duncan et al., 2001). The psychological effect of positive antibody results on subfertile women has not been assessed. In contrast, a false negative test result may result in unjustified expectant management or the institution of ineffective medical treatment.

Attempting to use medical history to identify pertinent clinical features was, in terms of accuracy, almost no better than flipping a coin. Transvaginal ultrasound, itself, was a poor screening test, misdiagnosing approximately one in three women. Combining medical history, TVU and CAT also gave disappointing results. Overall, the highest reported sensitivity (67%) required any of the three tests to be positive, yet still missed one in three women with tubal disease. The highest recorded specificity (100%) only identified three women. Similarly, while a combination of CAT and ultrasound rated as a good clinical test, confidence intervals ranged from a poor to excellent test due to the small number of women (four) positive for both. More than 200 antibody tests and ultrasound scans were performed in order to identify 6% (4/63) of the women with TFI. Such low returns are impractical in screening terms.

As attractive as it might be to reduce the number of normal laparoscopies, it is unlikely that this will happen in the near future. Recognized fertility treatments in the form of ovulation induction, intrauterine insemination and donor insemination all depend on normal Fallopian tube function and endometriosis can only be diagnosed and staged at laparoscopy. The multifactorial basis of TFI means that, realistically, only a proportion of women with TFI will be diagnosed by CAT. New recombinant protein ELISA are beginning to appear on the market and might overcome the broadness of antigens, but prior to this, the psychological and social impact of CAT and its overall acceptability in a subfertile population should be fully assessed. Rather than channelling resources into antibody testing, it is likely to be more clinically useful to research less invasive tubal evaluation techniques. In this context, ultrasound should not be dismissed. One small study reported that with the addition of three-dimensional power Doppler, free spill could be seen in 85% of cases compared to 43% using conventional hysterosalpingo-contrast sonography (Sladkevicius et al., 2000).

### Table V. Studies comparing clinical features, a combination of clinical features and C. trachomatis IgG antibody testing, and transvaginal ultrasound to laparoscopy

<table>
<thead>
<tr>
<th>Authors, assay, and screening test</th>
<th>No. of women</th>
<th>Prevalence of TFI (%)</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR– (95% CI)</th>
<th>Accuracy (%) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnston et al. (2000), ELISA/MIF</td>
<td>Clinical features</td>
<td>79</td>
<td>45</td>
<td>36 (23–52)</td>
<td>81 (67–90)</td>
<td>1.9 (0.9–4.2)</td>
<td>0.8 (0.6–1.0)</td>
</tr>
<tr>
<td></td>
<td>Clinical features and CAT</td>
<td>80</td>
<td>45</td>
<td>91 (78–97)</td>
<td>71 (56–82)</td>
<td>3.1 (1.9–4.9)</td>
<td>0.1 (0.04–0.4)</td>
</tr>
<tr>
<td>Ubaldi et al. (1998), TVU</td>
<td>133</td>
<td>65</td>
<td>86 (77–92)</td>
<td>98 (89–99)</td>
<td>39.7 (5.7–276.1)</td>
<td>0.1 (0.1–0.2)</td>
<td>90 (84–94)</td>
</tr>
</tbody>
</table>

TFI = tubal factor infertility; LR+ = positive likelihood ratio; LR– = negative likelihood ratio; CI = confidence interval; ELISA = enzyme immunoassay; MIF = microimmunofluorescence; CAT = C. trachomatis antibody test; TVU = transvaginal ultrasound.
Screening using medical history relating to TFI, TVU scan and C. trachomatis IgG antibody testing by ELISA, alone or in combination, failed to predict accurately TFI in subfertile women.

Acknowledgements
I wish to thank Mrs J. Skinner for excellent technical assistance and Dr Karen Dreesbach (Medac, Germany) for providing the ELISA kits. I am grateful for the statistical assistance provided by Dr Jill Mollison.

References

Submitted on March 18, 2003; accepted on May 30, 2003