The predictive value of medical history taking and Chlamydia IgG ELISA antibody testing (CAT) in the selection of subfertile women for diagnostic laparoscopy: a clinical prediction model approach

S.F.P.J.Coppus1,2,3,5, B.C.Opmeer2, S.Logan4, F.van der Veen1, S.Bhattacharya4 and B.W.J.Mol1,3

1Centre for Reproductive Medicine, Department of Obstetrics and Gynaecology, 2Department of Clinical Epidemiology and Biostatistics, Academic Medical Centre, Amsterdam, The Netherlands, 3Department of Obstetrics and Gynaecology, Máxima Medical Centre, Veldhoven, The Netherlands and 4Department of Obstetrics and Gynaecology, Aberdeen Maternity Hospital, Aberdeen, UK
5To whom correspondence should be addressed at: Máxima Medical Centre, Department of Obstetrics and Gynaecology, De Run 4600, 5500 MB Veldhoven, The Netherlands. Tel: +31 40 888 8384; Fax: +31 40 888 8387; E-mail: s.f.coppus@amc.uva.nl

BACKGROUND: Medical history taking as well as Chlamydia antibody titre (CAT) testing are currently used in the selection of patients for diagnostic laparoscopy with tubal patency testing. Most research has focused on the predictive value of CAT in isolation from medical history. We assessed therefore whether the combination of medical history and CAT improves the efficiency of selecting patients for laparoscopy as compared to the use of either medical history or CAT. METHODS: Data of 207 consecutive subfertile women were used to create multivariable logistic regression models for the prediction of tubal disease as diagnosed by diagnostic laparoscopy. RESULTS: The model with data of medical history only had an area under the receiver operating characteristic curve (AUC) of 0.65 (95% CI 0.56–0.74). Addition of CAT increased the AUC to 0.70 (95% CI 0.62–0.78) (P = 0.065). CAT was positive in 40 women and showed a sensitivity of 0.37 (95% CI 0.26–0.49) for a specificity of 0.88 (95% CI 0.82–0.93). In CAT positive women, a blank medical history did not decrease the probability of tubal disease. Of the 167 women tested CAT negative, 23 (14%) still had a high probability of disease due to their medical history and 11 of them (48%) showed tubal abnormalities on diagnostic laparoscopy. CONCLUSIONS: CAT testing adds valuable information to a woman’s risk profile based on her medical history. The combination of medical history taking and CAT testing has a better yield for diagnosing tubal disease than either of these alone.

Key words: tubal pathology/medical history/Chlamydia antibody titer/CAT/laparoscopy

Introduction
Tubal pathology affects approximately 15–30% of subfertile women (Evers, 2002). Currently, laparoscopy is considered to be the best available test for diagnosing tubal abnormalities. It allows direct visualization of the pelvis and, in addition to tubal patency testing, offers the opportunity to detect pelvic adhesions. However, laparoscopy does have several drawbacks. First, it is an invasive surgical procedure that requires general anaesthesia, both of which carry associated risks. Secondly, as an expensive investigation that requires operating time and dedicated personnel, its availability is not unlimited.

Therefore, triage is needed to limit the number of unnecessary laparoscopies, while maintaining a high diagnostic yield.

Since it was reported in 1979 that women with positive Chlamydia IgG antibodies are more likely to have tubal pathology than those without (Punnonen et al., 1979), numerous studies have reported on the value of Chlamydia antibody titre (CAT) testing to predict tubal pathology. Based on these reports, the Dutch Society for Obstetrics and Gynaecology (NVOG) recommends the use of CAT as a first-line test in the basic work-up of subfertile couples, with a fixed cut-off level (IgG MIF >1:32 or ELISA >1.1) above which post-infectious pelvic disease should be ruled out with laparoscopy and chromotubation (Swart et al., 1995; NVOG, 2004). The pre-test probability of disease of a woman, based on specific risk factors or other characteristics from the woman’s medical history is not formally taken into account. In contrast to the Dutch guideline, the
guideline from the National Institute for Clinical Excellence (NICE) in the United Kingdom does not recommend CAT testing in the workup of subfertile women, but advises the use of past medical history in the woman to decide whether diagnostic laparoscopy should be performed or not (NICE, 2004).

Most diagnostic test research is interpreted in isolation from its clinical context. However, this does not reflect daily clinical practice, in which the diagnostic investigation is a consecutive process always starting with history taking and physical examination, followed by laboratory tests and more invasive and costly tests such as imaging (Moons et al., 2004). With multivariable analysis one can reveal the ‘add-on’ value of diagnostic tests (Moons and Grobbee, 2002), in this case the additional value of CAT testing.

The present study therefore aimed to explore the diagnostic consequences of medical history taking by itself as recommended by the NICE guideline, CAT testing, as recommended by the NVOG, and the combined interpretation of the results from medical history and CAT testing, which is intuitively used in daily practice by many clinicians.

### Materials and methods

#### Patients

We used data from a previously reported study, in which 207 consecutive women referred for evaluation of subfertility underwent a diagnostic laparoscopy. Characteristics of included women and the materials and methods used in this study have been previously published (Logan et al., 2003). In brief: timing of tubal assessment was made on clinical grounds and was independent of the result of the CAT test. All couples had undergone assessment of ovulation and a semen analysis. Women who were physically unsuitable for laparoscopy or who had undergone a tubal patency test in the past (either hysterosalpingography or laparoscopy), as well as women with previous tubal surgery (including ectopic pregnancy) were excluded from the analysis. Before surgery, all women completed a standard questionnaire with regard to their demographic data and medical history. For CAT testing, a *Chlamydia trachomatis*-specific IgG antibody peptide-based ELISA (pELISA, Medac, Germany) was used. A positive IgG titre (>1.1) was indicative of a past infection. If the result was equivocal, an infection with *C. trachomatis* in the past could not be excluded and therefore positive and equivocal results were pooled. Tubal pathology (uni- or bilateral) was defined as the presence of adhesions involving the Fallopian tube and ovary, clubbing of the tube, hydrosalpinx or obstruction to the flow of dye in the absence of endometriosis, which was categorized as a separate entity according to the Revised American Fertility classification (American Fertility Society, 1985).

#### Data analysis

We developed multivariable logistic regression models to predict tubal pathology using the following data from the medical history: female age, number of previous term deliveries, previous induced abortion, type of subfertility (i.e. primary or secondary), duration of subfertility, history of pelvic inflammatory disease (PID), history of sexually transmitted disease, previous use of an intrauterine contraceptive device and prior non-gynaecological pelvic surgery. We evaluated the prognostic value of CAT both as a dichotomous and continuous variable. First, we checked the assumption of linearity between the continuous variables age, duration of subfertility, number of previous term deliveries and titre of CAT on the one hand and tubal pathology on the other hand visualized with smoothed piecewise polynomials (splines), and formally tested with a generalized additive model. As we found no significant deviations from linearity in the relationship with the outcome, it was not necessary to transform these continuous variables. Subclasses within categorical variables were dichotomized. For dichotomous and continuous variables, univariable beta coefficients ($\beta$), odds ratios (ORs) and 95% confidence intervals (95% CI) as well as $P$-values were calculated. In the univariable analysis, the $P$-value for statistical significance was set at $\leq 0.05$ in all analyses.

Subsequently, multivariable logistic regression analysis with a stepwise backwards selection procedure was used to construct the prediction models. Since the erroneous exclusion of a variable is more deleterious for a model than including too many factors (Mol et al., 2000; Steyerberg et al., 2000), we used a significance level of 30% for entry in the multivariable model and a significance level of 20% to stay in the model.

To adjust for overfitting, we performed internal validation with bootstrapping. Bootstrapping is a technique to evaluate the robustness of the model by simulating sampling variation in repeated samples from the study population. Cases are selected randomly from the original sample to create a new sample of the same size. As cases are drawn with replacement, some cases may be randomly selected multiple times and others not at all (Steyerberg et al., 2001). We generated 500 bootstrap data sets in which the same multivariable logistic regression model was estimated. By analysing the difference between the model based on the original dataset and the bootstrap estimates, a shrinkage factor was calculated and used to correct the original model coefficients and associated probabilities for overfit.

We developed two models. The first model (‘clinical history’ model) was based on items from medical history only. In the second model (‘clinical history and CAT’ model) previous infection with *C. trachomatis* as determined with CAT ELISA was combined with medical history. This second model was developed to explore what the result of CAT adds to the information already obtained from medical history taking. To make comparisons, as a third ‘model’, CAT was dichotomized into positive or negative test results, as recommended by the Dutch guideline.

The discriminative capacity of the two constructed models was evaluated with receiver operating characteristic (ROC) curve analysis, in which the area under the curve (AUC) was calculated. Accuracy of CAT was plotted into the ROC curve as a comparison. Sensitivity was defined as the number of women with tubal pathology that was predicted correctly, and specificity was defined as the number of women without tubal pathology predicted correctly. Calibration of the model, which is the agreement between predicted and observed probabilities, was evaluated with the goodness-of-fit Hosmer and Lemeshow test statistic and by visual inspection of the calibration plots. Finally, we explored the diagnostic consequences of different strategies. Data were analysed using SPSS 12.0.1 (SPSS Inc., Chicago, IL, USA) and S-PLUS 2000 (MathSoft Inc.)

### Results

The prevalence of tubal pathology was 30.4% (63/207). Demographic data of included women have been published elsewhere (Logan et al., 2003). In the univariable analysis, a higher number of previous term deliveries, a history of PID, and a positive CAT were statistically significant predictors of tubal pathology. The other variables were not statistically significantly associated with tubal subfertility (Table I). Four predictive variables, i.e. the number of previous term...
deliveries, duration of subfertility, a history of PID and previous non-gynaecological pelvic surgery were found to be associated with tubal pathology at a \( P \)-level of <0.20 in the ‘clinical history’ model (Table I).

The backward logistic regression selection procedure excluded duration of subfertility due to a \( P \)-value of 0.38 in the ‘clinical history and CAT’ model. However, as one of the aims of the present study was to compare medical history taking with an integrated approach of history taking and CAT testing, exclusion of this variable in the second model would have prohibited this exploration. Analysis of the regression coefficients and 95% CI for the OR of this variable showed that duration of subfertility only had a minor contribution towards the estimated probabilities. To allow comparison, we therefore decided to force duration of subfertility into the second model (Table I). The possible additional value of use of CAT as a continuous variable in the prediction model was explored, but did not improve the discriminatory capacity (data not shown). As a result, CAT was used as dichotomous variable in the ‘clinical history and CAT’ model.

Internal validation with bootstrapping showed a 17.5% overfit for the clinical history model, and a 14% overfit for the ‘clinical history and CAT’ model. Therefore the regression coefficients of the model variables were corrected with these shrinkage factors.

The ROC curves of the two models are shown in Figure 1. The ‘clinical history’ model showed an AUC of 0.65 (95% CI 0.56–0.74). ‘The clinical history and CAT’ model increased the AUC to 0.70 (95% CI 0.62–0.78) \( (P = 0.065) \) (De Long et al., 1988). As a comparison, the performance of CAT alone—sensitivity 37% (95% CI 26–49), specificity 88% (95% CI 82–93)—is plotted in the ROC space.

Calibration of the models was acceptable, as confirmed with the Hosmer and Lemeshow goodness-of-fit test statistic. This test statistic had a \( P \)-value of 0.30 for the ‘clinical history’ model and a \( P \)-value of 0.35 for the ‘clinical history and CAT’ model. Table II shows the agreement between different categories of predicted and observed probabilities in the two prediction models. Both models show an acceptable agreement between the mean predicted probabilities and the observed probability of tubal pathology. However, above a 30% predicted probability, both models underestimate the chance of disease. In Tables II, the categories of predicted probability of the ‘clinical history’ and ‘clinical history and CAT’ model are also cross-tabulated against the results of laparoscopy and CAT. Using the ‘clinical history model’, 28 women who tested CAT positive are classified as having a predicted probability of tubal pathology \( \geq 30\% \). However, 12 of these 28 women (43%) still showed abnormalities at laparoscopy (Table II). Using the ‘clinical history and CAT model’ the same 28 women who would have been classified as low risk patients based on clinical history only, shift to a high risk

---

**Table I. Results of the uni- and multivariable analysis**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariable analysis</th>
<th>Multivariable analysis clinical history model(^a)</th>
<th>Multivariable analysis clinical history and CAT model(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>95% CI</td>
<td>( P )</td>
<td>OR 95% CI  ( P )</td>
</tr>
<tr>
<td>Age (per year)</td>
<td>1.03 0.96–1.09</td>
<td>2.14 1.27–3.63 0.004</td>
<td>2.05 1.17–3.58 0.01</td>
</tr>
<tr>
<td>Number of deliveries</td>
<td>2.05 1.23–3.43</td>
<td>2.14 1.27–3.63 0.004</td>
<td>2.05 1.17–3.58 0.01</td>
</tr>
<tr>
<td>Induced abortion</td>
<td>1.59 0.77–3.29</td>
<td>1.01 1.00–1.03 0.11</td>
<td>1.01 1.00–1.03 0.38</td>
</tr>
<tr>
<td>Secondary infertility</td>
<td>1.59 0.87–2.88</td>
<td>1.01 1.00–1.03 0.11</td>
<td>1.01 1.00–1.03 0.38</td>
</tr>
<tr>
<td>Duration of subfertility</td>
<td>1.02 1.00–1.03</td>
<td>3.28 0.95–11.3 0.06</td>
<td>3.27 0.88–12.2 0.08</td>
</tr>
<tr>
<td>History of Pelvic inflammatory disease</td>
<td>3.48 1.06–11.4</td>
<td>3.28 0.95–11.3 0.06</td>
<td>3.27 0.88–12.2 0.08</td>
</tr>
<tr>
<td>History of Sexually transmitted disease</td>
<td>1.15 0.33–3.98</td>
<td>3.18 0.89–11.4 0.08</td>
<td>4.49 1.10–14.7 0.03</td>
</tr>
<tr>
<td>Intra-uterine contraceptive device use in history</td>
<td>2.40 0.67–8.59</td>
<td>4.07 1.88–8.81 &lt;0.001</td>
<td>4.07 1.88–8.81 &lt;0.001</td>
</tr>
<tr>
<td>CAT positivity</td>
<td>2.93 0.86–9.97</td>
<td>4.49 1.10–14.7 0.03</td>
<td>4.07 1.88–8.81 &lt;0.001</td>
</tr>
</tbody>
</table>

OR, odds ratio; CI, confidence interval; \( P \)-value.

\(^a\) Backward selection procedure.

\(^b\) Backward selection procedure with forced entry of duration of subfertility.
classification (i.e. a predicted probability of tubal disease >30%) (Table II). On the other hand, Table II also shows that among the 167 women who tested CAT negative, 23 (14%) still had a predicted probability of >30% of tubal disease due to their medical history. In this selected group of CAT negative women 48% (11/23) still showed tubal abnormalities on diagnostic laparoscopy.

To show the consequences of medical history taking by itself, CAT testing alone and a combination of medical history taking and CAT testing, diagnostic and clinical parameters are listed in Table III. For comparison, the consequences of performing diagnostic laparoscopy in all women without any selection strategy are shown in the first column. The number of laparoscopies that has to be performed to detect one woman with tubal pathology is comparable when using history, CAT or history and CAT and much lower than without any workup. However, the detection rate of tubal pathology (true positive rate) is highest when using the combined model (54% versus 38 and 37% respectively).

**Discussion**

In this study we evaluated the efficiency of medical history taking, CAT testing and a combination of both in selecting women for laparoscopy to detect tubal pathology. We found that the discriminative capacity of the ‘medical history’ model did not differ from that of CAT testing alone. However, combined interpretation of both history and CAT identified the women at highest risk for tubal disease, and increased sensitivity for a similar specificity. In CAT positive women, the probability of tubal disease is so high that a prompt diagnostic laparoscopy is indicated, irrespective of whether the woman has a suspect or non-suspect medical history. A small subgroup of CAT negative women (14%) still have a high probability of tubal disease due to their history, and laparoscopy will identify tubal abnormalities in around half of these women. The majority of women (70%) in this cohort were classified as having a probability >30% of tubal disease due to a negative CAT status and a non-suspect clinical history. Since 80% of these women will show no tubal abnormalities, laparoscopy in these women can be deferred.

Our study has some limitations. First, both our study cohort and the actual number of patients with positive factors in the medical history were relatively small. This limits the statistical power of this study and leads to relatively large confidence intervals. However, the strength of using this cohort of women is the fact that the decision to perform a diagnostic laparoscopy was irrespective of the result of CAT. This makes it possible to obtain estimates of diagnostic performance without partial verification bias, a limitation encountered in many other diagnostic studies on this subject (Logan et al., 2003; Whiting et al., 2004).

A second limitation of this study is the fact that we dichotomized the result of the CAT ELISA assay. A progressive

<table>
<thead>
<tr>
<th>Diagnostic parameter</th>
<th>All laparoscopy</th>
<th>Laparoscopy if probability &gt;30% as predicted by history model</th>
<th>Laparoscopy if positive CAT</th>
<th>Laparoscopy if probability &gt;30% as predicted by history and CAT model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of laparoscopies</td>
<td>207</td>
<td>42</td>
<td>40</td>
<td>63</td>
</tr>
<tr>
<td>Tubal disease diagnosed</td>
<td>63</td>
<td>24</td>
<td>23</td>
<td>34</td>
</tr>
<tr>
<td>Tubal disease missed</td>
<td>0</td>
<td>39</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>Unnecessary laparoscopies</td>
<td>144</td>
<td>18</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>NND¹</td>
<td>3.3</td>
<td>1.8</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>NA</td>
<td>38%</td>
<td>37%</td>
<td>54%</td>
</tr>
<tr>
<td>Specificity</td>
<td>NA</td>
<td>88%</td>
<td>88%</td>
<td>80%</td>
</tr>
</tbody>
</table>

¹Number needed to diagnose: number of laparoscopies needed to detect one abnormality.
increase in the likelihood of tubal pathology in patients with a higher titre of the Chlamydia antibody test has been shown for MIF tests (Thomas et al., 2000) as well as for ELISA (Tani-kawa et al., 1996). The antibody titres of women in our cohort showed only little diversity, as a result of which we were not able to demonstrate this relation in the present study. It may be possible that use of a MIF test would have resulted in a larger diversity between IgG titres of the patients, leading to extra diagnostic information if incorporated into the prediction model. MIF tests however have been shown to suffer from a significant amount of cross-reactivity with Chlamydia pneumoniae. Furthermore, interpretation of these tests is subject to inter-observer variation (Gijsen et al., 2001; Land et al., 2003). Moreover, a recent study showed the pELISA used in this study to be the one most specific for a previous C. trachomatis infection (Land et al., 2003).

A third limitation concerns our definition of tubal pathology. Broad criteria were used, namely presence of adhesions involving the Fallopian tubes and ovaries, clubbing of the tubes and hydrosalpinges or obstruction to the flow of dye irrespective of severity or whether the pathology was uni- or bilateral. Previous studies have shown that the diagnostic performance of CAT increases when the label ‘tubal disease’ is limited to the more severe cases (Land et al., 1998). Due to a low number of women with bilateral tubal disease in this study, we were unable to restrict our analysis to these cases. Detection of endometriosis by laparoscopy was not part of the intention to screen by medical history, since CAT will be unable to predict adhesions or damage due to endometriosis. Although this approach might be open to debate, this has been done previously in studies on the subject (Akande et al., 2003; Logan et al., 2003). Data on the capacity of medical history to predict endometriosis has been published previously (Calhaz-Jorge et al., 2004).

One previous study has attempted to predict tubal pathology by means of a multivariable logistic regression model (Hubacher et al., 2004). Women were divided into a ‘high probability’ and ‘low probability’ group based on a logistic regression model that included four variables, namely previous symptoms of PID, vaginal discharge, lower genital tract infection and presence of antibodies to C. trachomatis. The results demonstrated a sensitivity of 84% with a specificity of 29%. The authors concluded that ‘the role of history taking in the evaluation of women with tubal factor infertility is limited’. Although clinical history does not have the ultimate discriminative power to distinguish between patients with and without tubal pathology, we have demonstrated in the present study that it can be used to create individualized patient risk profiles and have shown the additional value of CAT testing upon medical history taking.

In our study we included women with both uni- and bilateral tubal disease. Unilateral tubal occlusion or tubo-ovarian adhesions may compromise fertility prospects moderately, but only bilateral tubal occlusion is known to virtually exclude all possibility of a spontaneous pregnancy (Mol et al., 1999). Therefore, it is questionable whether all women with minor tubal pathology should be labelled as having tubal disease. To date, no evidence-based policies for the treatment of unilateral tubal disease have been published (Ahmad et al., 2006), and therefore detection of one-sided tubal disease is unlikely to result in a major shift in treatment. Since most of these women have a low predicted probability of disease, postponing laparoscopy in these women may allow them to conceive spontaneously (Van der Steeg et al., 2006), thereby making laparoscopy redundant. Unfortunately, we did not have data about treatment independent pregnancy rates in this cohort of patients, and cannot answer this question.

Our model allows differentiation between women with a high predicted chance of tubal pathology and women with a low chance. What probability of diagnostic uncertainty is acceptable for women is not known at the moment. It may be possible that women prefer to have 100% certainty about their tubal integrity. In that case, since no workup will ever be 100% accurate, all women will require diagnostic laparoscopy. This will increase costs and enlarge the pool of women with a diagnosis of minimal pathology of questionable prognostic significance (Collins et al., 1995). Use of a multivariable prediction model including history and CAT allows a conjoint decision of whether or not to perform a laparoscopy, based on the woman’s as well as the doctor’s preferences. Depending on these preferences, one can perform a diagnostic laparoscopy in case of a high predicted probability, postpone laparoscopy for several months in cases where the predicted probability is low or use a less invasive alternative, i.e. hysterosalpingography or hysterosalpingo-contrast-sonography. Women should be adequately informed regarding the chance of delaying the diagnosis in case of low predicted probability and negative CAT status.

In this study, we explored the value of medical history taking and Chlamydia antibody testing in predicting tubal subfertility by constructing diagnostic prediction models. Whether the use of such models is feasible and clinically effective in daily routine practice is not known. Further research will have to evaluate women’s preferences regarding the trade-off between likelihood of detecting tubal pathology and the risks and discomfort associated with a potentially unnecessary diagnostic laparoscopy.

In conclusion, combined use of medical history taking and CAT testing has a superior diagnostic accuracy than one of these alone. CAT testing adds additional predictive value to a woman’s medical history risk profile.

Acknowledgements
This work was supported by grant 91.46.364 in the VIDI-program of ZonMW, The Hague, The Netherlands, and by the Máxima Medical Centre, Veldhoven, The Netherlands. The funding sources had no involvement in the design, analysis or reporting of this study.

Appendix
The chance of a patient of having tubal pathology can be calculated with the following formula: probability of tubal pathology = 1/[1 + exp(-β)] in which the β for the different models are: ‘clinical history’ model: β = -1.764 + (0.629 × number of term deliveries) + (0.0116 × duration of subfertility) + (0.979 × positive history of PID) + (0.955 × history of...
non-gynaecologic pelvic surgery). ‘history + cat’ model: $\beta = -1.874 + (0.616 \times \text{number of term deliveries}) + (0.0069 \times \text{duration of subfertility}) + (1.020 \times \text{positive history of PID}) + (1.201 \times \text{history of non-gynaecologic pelvic surgery}) + (1.208 \times \text{positive CAT})$.

REFERENCES


Submitted on July 14, 2006; resubmitted on November 16, 2006; accepted on December 14, 2006