Diagnostic and prognostic aspects of tubal patency testing
Coppus, S.F.P.J.

Citation for published version (APA):
Coppus, S. F. P. J. (2012). Diagnostic and prognostic aspects of tubal patency testing

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Subfertility is customarily defined as failure to conceive after regular unprotected sexual intercourse for one year. The prevalence of subfertility is around 14%, affecting one in seven couples. It has been estimated that in around 10-30% of these couples, subfertility is due to tubal pathology. These tubal abnormalities include tubal obstruction and pelvic adhesions which can be due to previous infection, endometriosis, or previous surgery. To rule in or rule out this pathology, the fertility work-up in subfertile couples is usually concluded with tubal patency testing. Currently, laparoscopy is considered to be the best available test for diagnosing tubal abnormalities, but it has 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. Nationally and internationally, a large practice variation exists in the timing of testing, the test used, and the basis on which women are selected for tubal testing. While some guidelines advocate medical history taking to triage women for tubal patency testing, others advocate Chlamydia antibody testing, or testing in all women. At the start of this project, test research on which these guideline are based on, had been conducted in isolation from its clinical context.

The work presented in the studies in this thesis started with a focus on methodological issues in diagnostic and prognostic research in general. Hereafter, we researched multivariable diagnostic models, to estimate pretest probabilities for tubal pathology that can help to distinguish women with a high and low probability of (severe) tubal disease at first consultation, and explored the added value of Chlamydia antibody testing to medical history taking. Finally, we investigated the prognostic significance of various tubal pathology tests like HSG, laparoscopy and CAT testing.

Chapter 1 gives an outline and describes the objectives of this thesis.

Chapter 2 describes the results of a study in which we evaluated the impact of publication of the Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement on the quality of reporting of test accuracy studies in two leading journals in the field of reproductive medicine. 24 articles published in 1999 (pre-STARD) and 27 published in 2004 (post-STARD) that reported on the diagnostic
performance of a test in comparison with a reference standard were included. Each article was scored on a 25 items checklist, dealing with the study question, study participants, study design, test methods, reference standard, statistical methods, reporting of results and conclusions. The mean number of reported STARD items for articles published in 1999 was 12.1 ± 3.3 (range 6.5-20) and 12.4 ± 3.2 (range 7-17.5) in 2004, after publication of the STARD statement. Overall, less than half of the studies reported adequately on 50% or more of the STARD items. The reporting of individual items showed a wide variation. There was no significant improvement in mean number of reported items for the articles published after the introduction of the STARD statement. From these data we concluded that authors of test accuracy studies in the two leading fertility journals poorly reported the design, conduct, methodology and statistical analysis of their study, which hampers the ability to interpret these studies on their scientific soundness. Strict adherence to the STARD guidelines should therefore be encouraged.

Chapter 3 presents an opinion article on methodological considerations when evaluating prediction models in reproductive medicine. Such models are used to calculate the probability of pregnancy without treatment, as well as the probability of pregnancy after ovulation induction, intrauterine insemination or in vitro fertilization. The performance of such prediction models is often evaluated with a receiver operating characteristic (ROC) curve. The area under the ROC curve, also known as c-statistic, is then used as a measure of model performance. However, the value of this c-statistic is low for most prediction models in reproductive medicine. We demonstrate that low values of the c-statistic are to be expected in these prediction models, but we also show that this does not imply that these models are of limited use in clinical practice. Arguments are given why the calibration of the model which is the correspondence between model-based probabilities and observed pregnancy rates, the availability of a clinically useful distribution of probabilities, the ability of a model to correctly classify couples for treatment decisions, and reclassification indices are more meaningful concepts for model evaluation.

Chapter 4 presents an exploratory investigation on whether information obtained from medical history taking can be quantified in a probabilistic model. We also explored whether CAT testing has additional value to medical history taking, given differences in prior probabilities. To answer these questions, 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. 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. From these data we concluded that 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.

Chapter 5 investigates the ability of medical history taking alone to predict tubal pathology at first consultation of the subfertile couple, as the ultimate aim of tubal testing is to identify women with bilateral tubal pathology in a timely manner, so they can be treated with IVF or tubal surgery. For this purpose, data of 3716 women who underwent tubal patency testing as part of their routine fertility workup, and participated in a nationwide cohort study were used. Elements from their medical history were related to the presence of tubal pathology. With multivariable logistic regression, we constructed two diagnostic models. In the first model tubal disease was defined as occlusion and/or severe adhesions of at least one tube, whereas in a second model tubal disease was defined as the presence of bilateral abnormalities. Both models discriminated moderately well between women with and women without tubal disease with an area under the receiver-operating characteristic curve (AUC) of 0.65 (95% CI: 0.63-0.68) for any tubal pathology and 0.68 (95% CI: 0.65-0.71) for bilateral tubal pathology, respectively. However, the models could make an almost perfect distinction between women with a high and a low probability of tubal pathology. A decision rule in the form of a simple diagnostic score chart was developed for application of the models in clinical practice. This decision rule could be used to select women for diagnostic laparoscopy more efficiently.

Chapter 6 explores the prognostic capacity of hysterosalpingography and laparoscopy in a general subfertile population. From data that were prospectively
collected in 38 fertility centres, we selected those women who underwent HSG and/or laparoscopy as part of their subfertility work-up. Follow-up started immediately after tubal testing and follow-up was ended 12 months thereafter. We correlated findings at HSG or laparoscopy with time to pregnancy. Follow-up was censored at the date of last contact, when the woman was not pregnant or at the start of fertility treatment. Kaplan-Meier curves for the occurrence of spontaneous intrauterine pregnancy were constructed for women without tubal pathology, for those with unilateral tubal pathology and for women with bilateral tubal pathology at HSG or laparoscopy. With multivariable Cox regression analysis the association between tubal pathology and the occurrence of an intrauterine pregnancy was expressed in fecundity rate ratios (FRRs). In total, 3301 women were included, of whom 2043 underwent HSG only, 747 underwent diagnostic laparoscopy only and 511 underwent both. At HSG, 322 (13%) women showed unilateral tubal pathology and 135 (5%) showed bilateral tubal pathology. At laparoscopy, 167 (13%) showed unilateral tubal pathology and 215 (17%) showed bilateral tubal pathology. Multivariable analysis resulted in FRRs of 0.81 [95% confidence interval (CI): 0.59-1.1] for unilateral, and 0.28 (95% CI: 0.13-0.59) for bilateral, tubal pathology at HSG. The FRRs at laparoscopy were 0.85 (95% CI: 0.47-1.52) for unilateral, and 0.24 (95% CI: 0.11-0.54) for bilateral, tubal pathology. From this study it was concluded that women with unilateral tubal pathology at HSG or laparoscopy had moderate, non-significant reduction in spontaneous pregnancy chances, whereas those with bilateral tubal pathology at HSG or laparoscopy had a severe reduction in the ability to achieve spontaneous conception. This reduction in fertility prospects was similar for HSG and laparoscopy, suggesting that HSG and laparoscopy have a comparable predictive capacity for natural conception.

Chapter 7 investigates the prognostic significance of positive *Chlamydia trachomatis* antibodies for spontaneous pregnancy once tubal pathology has been ruled out. We studied ovulatory women in whom HSG or laparoscopy showed patent tubes. Women were tested for *C. trachomatis* immunoglobulin G (IgG) antibodies with either micro-immunofluorescence (MIF) or an ELISA. CAT serology was positive if the MIF titre was ≥ 1:32 or if the ELISA index was > 1.1. The proportion of couples pregnant without treatment was estimated at 12 months of follow-up. Time to pregnancy was considered censored at the date of the last contact when the woman was not pregnant or at the start of treatment. The association between CAT positivity and an ongoing pregnancy was evaluated with Cox regression
analyses. Of the 1882 included women without visible tubal pathology, 338 (18%) had a treatment-independent pregnancy within 1 year [estimated cumulative pregnancy rate 31%; 95% confidence interval (CI): 27%-35%]. Because of violation of Cox regression analysis assumptions after 9 months of follow-up, which could not be explained on sound clinical nor statistical grounds, regression analyses were limited to the first 9 months after tubal testing. Nevertheless, positive C. trachomatis IgG serology was associated with a statistically significant 33% lower probability of an ongoing pregnancy [adjusted fecundity rate ratio 0.66 (95% CI 0.49 to 0.89)]. We concluded that CAT testing does not only have diagnostic value, but also prognostic value, even after HSG or laparoscopy has shown no visible tubal pathology, as subfertile women with a positive CAT have lower pregnancy chances than CAT negative women. After external validation, this finding could be incorporated into existing prognostic models.

Chapter 8 presents a discussion of the findings of this thesis, clinical implications and future research recommendations.

Chapter 10 provides an epilogue in which the results of three different theses on tubal pathology from our study group are integrated.