Diagnosing tubal pathology: The individual approach
Broeze, K.A.

Citation for published version (APA):
Broeze, K. A. (2013). Diagnosing tubal pathology: The individual approach

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.
Chapter 8

General discussion
Tubal pathology counts for 30% of female subfertility (Evers, 2002). In the fertility work-up, clinical history, physical examination and several tubal patency tests are used to diagnose tubal pathology and to provide optimal management strategies for couples suffering from subfertility.

Since the late 1980’s many diagnostic studies have been performed to assess the diagnostic accuracy of tubal patency tests, such as Chlamydia antibody testing, hysterosalpingography and diagnostic laparoscopy, of which the latter is regarded as the reference test, the gold standard to diagnose tubal pathology (Anestad et al. 1987, Ismajovich et al. 1986, La Sala et al. 1987, Fayez et al. 1988). Despite all available evidence published in the 30 years after these first diagnostic accuracy studies and more recent guidelines (NICE 2004, NVOG 2004, ASRM 2006), there is still no consensus on the best diagnostic strategy to diagnose tubal pathology. It is unknown which test should be initially used and whether the tests do have additional value over clinical history and physical examination. A major reason for the lack of good research on tubal patency testing is the difficulty of integrating patient characteristics with the results from CAT, HSG and laparoscopy - in several possible sequences- since the diagnostic performance of the tubal patency tests are assessed in isolation of these patient characteristics.

This creates important clinical uncertainty because the accuracy of a test can be different for patients with a different profile – e.g. its discriminative ability may be better or worse in women with a particular characteristic and because the target condition may be related to particular risk factors, i.e. the prevalence varies across women with different characteristics.

In this thesis we used a relatively new approach, called individual patient data meta-analysis, to integrate information from clinical history, physical examination and tubal patency tests to diagnose tubal pathology and, finally, to predict natural conception in subfertile women.

**Individual patient data (IPD) meta-analysis in reproductive medicine**

In medical research, systematic reviews and meta-analyses are known to provide the highest level of scientific evidence (Sackett et al. 2007). In reproductive medicine, many systematic reviews and meta-analyses are performed, mainly to estimate pooled treatment effects of assisted reproductive therapy. In diagnostic research, systematic reviews and meta-analyses are less often performed, due to typical difficulties in this type of research. For example, diagnostic studies are rarely randomised, they show a poor quality of reporting and the most common measurement of diagnostic accuracy is a combination of sensitivity and specificity, which is influenced by many external factors such as disease prevalence and patient characteristics; all factors that lead to major difficulties in pooling aggregated data in a meta-analysis (Rutjes et al. 2006, Leeflang et al. 2008, Mallett et al. 2006). Clinically, and even more important, a diagnostic strategy containing both patient characteristics and several diagnostic tests in different inter-dependent sequences, hampers the performance of a meta-analysis at the study level. Therefore, it was suggested to use data of the individual
patients, available from studies included in the systematic reviews (Khan et al. 2003, Stewart et al. 1993).

Before the start of this thesis, such individual patient data meta-analyses were only performed at the therapy domain in oncology and cardiovascular research. In reproductive medicine, IPD meta-analyses were rarely used, in contrast to ‘conventional’ meta-analyses. Therefore, we first discussed the advantages and disadvantages of both approaches and described the opportunities of IPD meta-analyses for research purposes in reproductive medicine (chapter 2). The anticipated benefits of IPD meta-analyses seemed to be manifold. Using data of individual patients allows analysis of continuous test results, of follow-up data (i.e. live birth or ongoing pregnancy in stead of chemical pregnancy), and of differently defined outcome measurements (i.e. unilateral versus bilateral tubal pathology or pregnancy per treatment cycle versus pregnancy per woman). Also, information of the patient characteristics and several combinations and sequences of tests can be integrated in the analyses. Direct contact with the original authors might improve detection of unpublished or ingoing studies and poor quality of reporting can be clarified and missing information supplied.

On the other hand, the most important challenges of IPD meta-analyses are the required time investment and the problem of missing data at both the study level as well as at the patient level, due to untraceable authors, lost data and missing variables between databases.

For the four performed IPD meta-analyses in this thesis the amount of untraceable authors varied from 10 to 45%. On average, 34% of the approached authors did not respond to repeated emails and phone calls. Of the authors that did respond to our approach 49% had lost their original data. The amount of available data for the variables in the acquired databases showed a major variability between variables and between studies. Only female age was known for all included patients and for only one study, data on the included variables were available for all patients. All other databases had missing data at the patient level, leaving a patchwork dataset for the IPD meta-analyses. To overcome this missing data problem, multiple imputation was used to impute missing data. By imputing laparoscopy results the risk of verification bias was decreased and analysis of a greater amount of data was allowed, leading to a more reliable result (van der Heijden et al. 2006, de Groot et al. 2008).

Direct contact with the participating authors had several advantages. The authors were easy to approach for questions on their study protocols or clarifications on their data. Some authors also provided non-published follow-up data, increasing the total amount of available data, and annual meetings often resulted in improved analyses and manuscripts.

A detailed study protocol was described in which we outlined the main methodological objectives of using IPD meta-analysis, such as the development of a framework to perform IPD meta-analyses, the exploration of missing data and the comparison with ‘conventional’ meta-analyses (chapter 3). The general methods of IPD meta-analyses that were performed in this thesis were also outlined in this chapter; from the selection of studies and acquisition of original data, the assessment of methodological quality, to the description of the statistical analyses. Four clinical topics were described, for which diagnostic algorithms and prediction
rules could be provided by using IPD meta-analyses; the diagnosis of endometrial cancer in women with postmenopausal bleeding, prediction of preterm birth, ovarian response tests in women for IVF and, as presented in this thesis, the diagnosis of tubal pathology.

**Diagnosing tubal pathology**

The Chlamydia antibody test, hysterosalpingography and diagnostic laparoscopy are the most often used diagnostic tests to indicate tubal pathology. Although less invasive alternatives for HSG and laparoscopy, are available, like hysterosalpingo-contrast-sonography (Schlief and Deichert, 1991) or transvaginal hydrolaparoscopy (Gordts et al., 1998), these newer tests have not gained widespread use in daily clinical practice, as a result of which most clinicians still use hysterosalpingography and/or diagnostic laparoscopy besides CAT to finalize the fertility work-up.

In most clinics the Chlamydia antibody test is used as the first test or screening test in the diagnostic pathway for tubal pathology (Punnonen et al. 1979, Den Hartog et al. 2008). In a previous meta-analysis it was shown that the diagnostic accuracy of CAT covers a wide range of sensitivities (21% - 90%) and specificities (29% - 100%). Since this 'conventional' meta-analysis was limited by the published 2x2 tables of the original studies, we performed an IPD meta-analysis in which we acquired continuous CAT results from 14 primary studies, containing data from 6,191 women (chapter 4). In this way, the accuracy of three types of CAT assays could be evaluated for any as well as for bilateral tubal pathology.

We defined any tubal pathology as the occlusion of one or both fallopian tubes, with or without hydrosalpinges or peritubal adhesions, since most original studies described the tubal status in this way. Because it is known that only bilateral tubal pathology significantly decreases the chance of natural conception as well as the chance of conception after intra-uterine insemination (IUI), we also performed the analyses for bilateral tubal pathology (Mol et al. 1999, Steures et al. 2004, Verhoeve et al. 2011).

For all three assays, low sensitivities combined with high specificities were found; implicating that CAT is able to avoid unnecessary, invasive testing, in women without tubal pathology, but that women with tubal pathology might remain undiagnosed. Furthermore, it was shown that the diagnostic accuracy of the micro-immunofluorescence (MIF) test was significantly superior to that of the immunofluorescence (IF) test and the Chlamydia trachomatis specific enzyme-linked immunosorbent assay (ELISA), with an AUC of 0.75 for any tubal pathology and an AUC of 0.77 for bilateral tubal pathology. This result is in line with the study of Land et al. in which a direct comparison of MIF and ELISA showed an odds ratio twice as high for MIF (Land et al. 2003). Therefore we concluded that, when CAT is used in the initial screening for tubal pathology, MIF should be the test of first choice (chapter 4).

Whether CAT has additional value for the diagnosis of tubal pathology, after the performance of HSG or laparoscopy, has been debated. A recent study showed that women without visible tubal pathology on HSG or laparoscopy, but with a positive CAT had a 33% lower probability of natural conception, as compared to CAT negative women (Coppus et al. 2011).
This might imply a useful prognostic factor for pregnancy chances in women suffering from subfertility. In most clinics, hysterosalpingography is often used to visualize tubal patency after the performance of CAT. In a previous meta-analysis, the diagnostic accuracy of HSG showed a sensitivity of 65% with a specificity of 83%. However, these single figures imply that the accuracy of HSG is invariant across women. Differences in accuracy across subgroups are generally not taken into account in conventional meta-analysis, where a pooled estimate is obtained for overall sensitivity and specificity, but also in the original studies, the capacity of HSG to diagnose tubal pathology was assessed in isolation of patient characteristics, such as female age or previous PID. This does not resemble clinical practice, where clinical history taking, physical examination, and the performance of CAT, precede hysterosalpingography and this information is integrated by the physician in a clinical diagnosis. In a second IPD meta-analysis in which we combined data from patient characteristics with HSG results, we found that hysterosalpingography had a lower sensitivity in women with a previous PID or a positive CAT (chapter 5). The diagnostic accuracy of HSG did not vary across women for any of the other patient characteristics that could be included in the meta-analysis.

To develop an improved diagnostic strategy by incorporating risk factors for tubal pathology, in our third IPD meta-analysis we combined the results from clinical history, physical examination, CAT, HSG and laparoscopy (chapter 6). From a previous meta-analysis, it was known that previous PID, ectopic pregnancy, endometriosis, complicated appendicitis and pelvic surgery are strong risk factors for tubal pathology (Luttjeboer et al. 2009). Another study based on the data from a multicentre cohort study, found that referral by a gynaecologist, female age above 32 years, dysmenorrhoea and smoking of the woman, a pregnancy of both partners in a previous relationship, male smoking habits, history of PID, history of Chlamydia infection, ectopic pregnancy and previous pelvic surgery increased the likelihood of bilateral tubal pathology. In contrast, a previous pregnancy in the current relationship and a higher male age lowered the chances of bilateral tubal pathology (Coppus et al. 2007). For our IPD meta-analysis data of this multicentre study were used, combined with data of den Hartog et al, Land et al, van der Linden et al and Ng et al. Unfortunately, due to the unavailability of some of the above variables in the additional four studies we were only able to use data on female age, duration of subfertility, previous pregnancies, BMI, previous PID, pelvic surgery, and previous Chlamydia infection. We found duration of subfertility, previous pregnancies, previous PID, pelvic surgery and previous Chlamydia infection to be significant predictors in the patient characteristics model, which is comparable to the results of the above-mentioned study (Coppus et al. 2007). Also, the predictive performance of the model for bilateral tubal pathology was comparable to the previous study (AUC 0.63). Furthermore, our IPD meta-analysis showed that addition of both CAT and HSG significantly increased the predictive performance of the model (AUC 0.76) when diagnosing tubal pathology.
Predicting natural conception

In the end, the ultimate goal of the subfertility work-up is to assess the chances of natural conception. These chances influence future treatment strategy, since couples with a probability of natural conception of more than 30% can be offered expectant management (Steures et al. 2006). Therefore, it would be of great advantage when natural conception can be predicted by the tubal patency tests performed in the diagnostic pathway for subfertile couples. From previous literature, it was known that both hysterosalpingography, as well as diagnostic laparoscopy, both show a slight reduction in pregnancy chances when unilateral tubal pathology was shown. In women with bilateral tubal pathology a significant decrease in pregnancy chances was seen after both tests.

We performed a fourth IPD meta-analysis to evaluate whether natural conception can be predicted from both patient characteristics, as well as from CAT, HSG and laparoscopy (chapter 7). Using data from 6,326 women, we found that increasing female age and a longer duration of subfertility lead to a significant lower probability of natural conception, whereas secondary subfertility showed a significantly higher probability of natural conception, results comparable to the model of Hunault (Hunault et al. 2004). BMI, history of PID, previous Chlamydia infection, previous pelvic surgery and a positive CAT result were no significant predictors of natural conception. According to previous studies, we also found that for both tests, unilateral tubal pathology showed a slight reduction in natural conception (FRR 0.83-0.84), whereas bilateral tubal pathology showed a major decrease in natural conception rates (FRR 0.41-0.47). These rates were even lower when patient characteristics were taken into account, especially when bilateral tubal pathology was seen on HSG (FRR 0.19).

Previous studies that did not take patient characteristics into account found similar predictive capacities for HSG and laparoscopy (Mol et al. 1999, Verhoeve et al. 2011). We found that when information from clinical history and physical examination was integrated in the prediction of natural conception, laparoscopy does not have added value for women with bilateral tubal pathology on HSG. This corresponds to the results of another study, where in 95% of patients, laparoscopic findings did not change the HSG-based treatment plan (Lavy et al. 2004). The possible therapeutic effect of HSG in enhancing natural conception is currently studied in a RCT, comparing the use of oil-based contrast medium versus water-based contrast medium (H2Oil study NTR3270). Obviously, laparoscopy could still have additional value in women with endometriosis or in women with inconclusive results at HSG. Previously, a randomised controlled trial was performed to assess the value of the routine use of HSG before laparoscopy was done (Perquin et al. 2006). In this study women were randomised between routine HSG followed by laparoscopy or immediate laparoscopy. Women without tubal pathology on HSG were managed expectantly for 6 months. When no pregnancy occurred after this time period, they underwent a laparoscopy. Women with tubal pathology on HSG underwent laparoscopy within 1 to 2 months. At 18 months of follow-up, no differences in pregnancy rates between both strategies were seen. It was concluded that HSG has no role in the fertility work-up. However, Coppus et al showed from
the same data that a strategy starting with HSG results in a 30% reduction in the number of laparoscopies (Coppus et al. 2006). Also, in a recent cost-effectiveness study it was shown that performing hysterosalpingography as the first tubal patency test, only followed by laparoscopy when tubal pathology was seen, was more cost effective than immediately performing laparoscopy (Verhoeve et al. 2013). Interestingly, the data of the RCT of Perquin et al were taken into account in the present IPD meta-analysis, showing a high predictive value of HSG after patient characteristics were noticed.

**Conclusion and clinical implications**

From the present thesis we can conclude that IPD meta-analysis offers substantial advantages for diagnostic and prognostic research in reproductive medicine. Better data sharing policies, advanced data storage facilities and standardized data collection with pre-defined variables in original studies are needed since they will increase the amount of available data, thereby allowing the analysis of multivariable models, without missing data problems. Clinically, the integration of patient characteristics from clinical history and physical examination with the results of tubal patency tests could improve both the diagnosis of tubal pathology as well as the prediction of natural conception in women suffering from subfertility. Further validation of the model and prospective evaluation of its benefits when used in the clinical setting are required before its implementation in practice. The Chlamydia antibody test might have a role in the diagnosis of bilateral tubal pathology before the performance of the HSG, but the use of MIF is then strongly recommended. There was no evidence that CAT is also useful in the prediction of natural conception. Hysterosalpingography should be performed before laparoscopy since it is proven to be a useful tubal patency screening test, with a better predictive capacity compared to laparoscopy. Laparoscopy can be withheld for women with inconclusive results on HSG or for women suspected for endometriosis.
Implications for future research

Future IPD meta-analyses will enhance their level of scientific evidence when it would be possible to include all available data from original, prospective, studies. Therefore, the original studies should collect their data in a standardized way following pre-defined protocols and definitions. For all included patients, all potentially relevant information from clinical history and physical examination should be documented and reported alongside the diagnostic accuracy findings. Also, when evaluating multiple tests, all tests should be performed in all patients, at least in an arbitrary subset of patients, to avoid or allow adjustment for partial verification bias. In ideal circumstances, after study completion data of the original studies should be centrally archived and made accessible for inclusion in future ‘conventional’ or IPD meta-analyses. Widespread implementation of data sharing policies in combination with central data archival facilities could enhance the acquisition of complete data, as IPD-researchers will become less dependent on whether individual authors are willing to share their data, or whether authors can still be traced years after the original publication. As the guidelines prescribe, the diagnostic strategy should start with clinical history taking and physical examination. Tubal patency tests should have added value in the selection of women with a high probability of bilateral tubal pathology, since these women will benefit from an early start of assisted reproductive therapy. After more than 30 years of research, it is still debated which tests should be performed to indicate that specific women. The ultimate approach to solve this question would be a large multicentre randomised trial, in which different screening strategies including patient characteristics are compared. These strategies should be evaluated for their effectiveness as well as their cost-effectiveness, considering the worldwide need to keep health care expenditures within sustainable limits and the threatening of financial cuts for assisted reproductive therapy.
References


ASRM Practice Committee Reports. Optimal evaluation of the infertile female. Fertil Steril 2006; 86, s264-s267.


Stewart LA and Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? Lancet 1993; 341: 418-22.

