Diagnostic tests for tubal pathology from a clinical and economic perspective

Verhoeve, H.R.

Citation for published version (APA):
8

General discussion
Tubal pathology is a common cause for in- or subfertility. The prevalence depends on the population studied and varies between primary, secondary and tertiary study populations (Evers, 2002).

Diagnostic tests for tubal patency are usually planned as last tests in the work-up for subfertility. The reasons for postponement of these tests are several fold. They are invasive procedures, uncomfortable to women, need to be planned and generate healthcare costs. Because of their invasive nature, complications such as infection or injury to the genital or internal organs can occur (Stumpf and March, 1980; Pittaway et al., 1983; Forsey et al., 1990; FW Jansen et al., 1997). Also, a planned invasive test may have become unnecessary, because natural conception has occurred after the first consultation of a couple but before the planned diagnostic test has been performed. However, early patency tests are warranted in women at high risk for tubal pathology because in women with bilateral tubal pathology, natural conception is severely reduced and undue delay in referral for surgery or IVF can be prevented (Mol et al., 1999; Verhoeve et al. 2011). For this reason, early identification of women at high risk and those at low risk for tubal pathology is of importance.

Our first aim in this thesis was therefore to identify and quantify which items from the medical history are associated with an increased risk of tubal pathology in subfertile women. For this we performed a systematic review and meta-analysis of the literature (Chapter 3). This meta-analysis of 32 original studies showed that a history of PID, ectopic pregnancy and endometriosis, complicated appendicitis and pelvic surgery, were strongly associated with the presence of tubal pathology. A strong association between a history of sexually transmitted disease and tubal pathology was found in two case-control studies but not in cohort studies. When we pooled the case-control studies on a history of induced abortion we found a weak, but significant, association with the risk of tubal pathology. A comparably significant association was found in a large cohort study of our own hospital population (Chapter 4). In our meta-analysis we were only able to perform univariable analyses and therefore could not correct for the mutual dependence of individual risk indicators. The associations between items in the medical history and risk of tubal pathology may for this reason be overestimated. We were, however, able to perform a multivariable analyses in a large prospective cohort study of risk indicators in the medical history of tubal pathology. In this cohort of subfertile couples, collected from 38 clinics in the Netherlands, the prevalence of tubal pathology was 18%. This study confirmed the relation of many of the before mentioned items in the medical history as risk indicators for tubal pathology and in general showed comparably strong associations between the different items and tubal pathology (Chapter 5). In the analysis we also found that some male characteristics were associated with the probability of tubal subfertility, such as age, smoking habits as well as a pregnancy in a previous relationship. It is obvious that these characteristics cannot causally be related to tubal damage. We therefore think that these
factors act as indicators for some unrecorded variables, such as number of sexual partners, socio-economic status or are a reflection of lifestyle or other unknown factors which we could not extract from the available information.

Systematic reviews and conventional meta-analyses are established methods for generating high-level evidence to support the development of clinical practice guidelines. They are used to summarize the results of multiple primary studies, allowing a solid base for rational balancing of clinical decisions. Although our meta-analysis was performed according to established methodology (Bossuyt et al., 2003), it has limitations. We defined tubal pathology as any abnormality disturbing the integrity of one or both tubes, even though it is known that only women with bilateral tubal pathology have significantly reduced chances to conceive naturally or after intra uterine insemination (Mol et al., 1999, Verhoeve et al., 2011). The original studies we were able to use for the meta-analysis, varied in their definition of tubal pathology. Other factors that influenced the outcome were: One, partial verification bias; not all patients received an HSG as well as a laparoscopy. Two, publication bias; studies with negative or non-significant outcome are less likely to be published. Three, differences in prevalence of disease; study populations varied from secondary to tertiary hospital populations. Four, study type; case-control as well as cohort studies were included. Case-control studies tend to over-estimate the relation between risk factors in the medical history and presence of tubal pathology. Controls are usually women who have conceived naturally and have delivered live born children and although this excludes significant bilateral tubal disease, this does not exclude presence of mild tubal pathology or unilateral tubal pathology. Five, the reliability of the accuracy of the diagnostic test that is used; it has become standard to use diagnostic laparoscopy as the reference test in detecting and diagnosing tubal pathology. DL can, however, not be seen as the gold standard diagnostic test for tubal pathology, as we will discuss later.

Our second aim was to develop a clinical decision rule based on medical history that can help to identify women at high and low risk for tubal disease. For this purpose we developed two decision models (Chapter 5). In the one model we used as definition presence of any possible type of tubal pathology and in the second model we used the definition of presence of bilateral tubal pathology. Within the observed proportions of affected women, the calculated probabilities of the model for bilateral tubal pathology correlated better than that of the model for any type of tubal pathology. These calculations were based on the collection of data in 38 centres in the Netherlands, representing a mixed secondary and tertiary care population. The model is easy to use in a clinical setting and because it was developed from a mixed population, seems generalizable. However, confirmation of this will require external validation (Chapter 5).

Our third aim was to compare the predictive capacity of the most commonly used tubal tests, i.e. HSG and DL with chromopertubation, on natural conception (Chapter 6). For this purpose
we used the data collected from the same large cohort of subfertile women as in chapter 5. This study showed a moderate (non-significant) reduction in the fecundity rate ratio (FRR) if unilateral tubal pathology was seen and a severe (significant) reduction in the FRR if bilateral tubal pathology was seen at HSG and/or DL. The reductions in FRRs for unilateral as well as bilateral tubal pathology were similar between HSG and DL and suggest that HSG and DL have a comparable predictive capacity on natural conception. The outcome on the FRRs confirmed the findings of cohort studies dating from the ninety nineties, but in contrast to these, we could not conclude that the predictive capacity of DL was better than that of HSG (Mol et al., 1997; 1999a; 1999b). The earlier cohort studies consisted of tertiary hospital populations. Two studies were retrospective in design (Mol et al., 1997; 1999a) and one a prospective cohort study (Mol et al., 1999b). In the prospective study, in which HSG and laparoscopy were compared, the prevalence of bilateral tubal pathology was higher than in our cohort of a mixed population and the time between HSG and laparoscopy was markedly shorter in our study (Mol et al., 1999b). The longer delay between HSG and DL may have resulted in a selection bias of a cohort of women with a poorer fertility prognosis than that of ours, because, according to the prognostic model of Hunault (Hunault et al., 2004), duration of subfertility is one of the most important prognostic factors. We therefore think that our findings are more reliable. We also noticed a lack of agreement between HSG and laparoscopy in those women that underwent both tests, which confirmed the findings of a previous study and support the notion that laparoscopy cannot be considered the gold standard for diagnosing tubal pathology (Mol et al., 1999b).

Our fourth aim was to analyse the cost-effectiveness of different diagnostic strategies for tubal pathology, taking patient characteristics as obtained from the medical history, into account. For this purpose we developed a Markov analytic model and defined six different scenarios for women of 30 to 33 years of age (Chapter 7). Our Markov model showed that the most cost-effective scenario is, to perform no diagnostic tubal tests and delay treatment for women aged until 38. Until the age of 30 years and from the age of 40 the prevalence of bilateral tubal pathology was of no influence on the outcome. The threshold value for the prevalence of bilateral tubal disease at which direct treatment (i.e. IVF) became more cost-effective than delayed treatment, declined gradually from 95% at the age of 31 to 10% at the age of 39, meaning that even in the younger woman at very high risk of tubal pathology, it is not cost-effective to plan a diagnostic tubal patency test early in the fertility work-up. The outcome was not influenced by the presence and treatment of minimal or mild endometriosis, which implies that there is no reason to prefer diagnostic laparoscopy (the gold standard test to diagnose presence of endometriosis) above HSG solely for the reason not to miss out on presence of endometriosis. If tubal tests would be performed, HSG and DL for those women with bilateral tubal pathology at HSG, followed by tailored treatment (i.e. expectant management in case of no- or unilateral tubal pathology and IVF treatment in case of bilateral tubal pathology), was more cost-effective than diagnostic laparoscopy followed by tailored treatment. The scenarios of only HSG or DL
were always less cost-effective than the scenarios in which an HSG was followed by DL (in case of abnormal findings at HSG).

We did not consider the Chlamydia Antibody Test (CAT) as a separate scenario. Although CAT can be of help to differentiate between women who are at low or high risk of tubal pathology, the test result is not conclusive whether tubal pathology is present and does not provide additional anatomical information like HSG does. Another reason not to consider CAT as a separate scenario is that a recent Individual Patient Data (IPD)-analysis showed that the combination of CAT and HSG is a better predictor of bilateral tubal pathology than CAT (Broeze et al., 2012).

Summarizing our conclusions we found that items from the medical history are related to presence of tubal pathology and that these can be quantified and used in a prediction model to identify women at high risk for tubal pathology. Unilateral pathology at HSG and/or DL reduces natural conception chances moderately, whereas bilateral tubal pathology reduces these chances severely. HSG and laparoscopy showed a comparable predictive capacity for natural conception, but neither of these tests can be considered the perfect reference test for the assessment of tubal patency. Routine diagnostic tubal patency tests early in the fertility work-up are not cost-effective. Our analysis also showed that when there is a need for information on the tubal status, a strategy based on HSG followed by laparoscopy, if HSG shows bilateral tubal pathology, is superior over laparoscopy alone.

It is humbling to notice that since the introduction of diagnostic tubal tests, a century ago, we still have to conclude that most studies show methodological shortcomings in their design and are for this reason a cause for debate and practice variation. Studies in non-selected subfertile women comparing HSG with diagnostic laparoscopy, taking place on the same day or with a minimal delay between these tests, nor sufficiently powered randomized clinical trials comparing two strategies for tubal testing have ever been performed. This would have taken away many of the methodological problems. It is, however, not likely that studies that perform multiple tests in a short time frame will be performed and we therefore restricted ourselves to the available knowledge and evidence that has been gathered over the past decades. To our knowledge, only two randomised trials that addressed the routine use of HSG and diagnostic laparoscopy in the fertility work-up have been published (Tanahatoe et al., 2005, Perquin et al., 2006). In the first trial, routine diagnostic laparoscopy after a normal HSG and before the start of intra uterine insemination (IUI) was compared with diagnostic laparoscopy after a normal HSG and following six unsuccessful IUI treatment cycles. This trial did not show an improved pregnancy outcome if DL was routinely performed after a normal HSG (Tanahatoe et al., 2005). In the second trial, women were randomised between routine HSG followed by DL or immediate DL. A laparoscopy followed within 1-2 months in case of abnormal HSG findings and after 6 months in case of normal HSG but no pregnancy within 6 months. Routine use of HSG at an
early stage in the fertility work-up, prior to laparoscopy and dye, did not result in a different cumulative pregnancy rate compared with the routine use of laparoscopy and dye without HSG. The conclusion of the authors was that HSG can be omitted from routine use (Perquin et al., 2006). This statement was criticized in a letter by our group, because randomised comparisons of diagnostic strategies can only assess clinical benefits from the use of a test if the test result is followed by explicit treatment protocols. The design and sample size of the study felt short to make this comparison (Bossuyt et al., 2000; Coppus et al., 2006). The authors rebutted this by mentioning that their trial was an effectiveness trial and not an efficacy trial (Helmerhorst et al., 2006). Our conclusion from this study is, that in those women who received routine HSG, there was a reduction of 30% in diagnostic laparoscopies (56% in the routine HSG group compared with 86% in the routine laparoscopy group) without a reduction in the pregnancy rate. Unnecessary laparoscopies can thus be avoided in a substantial number of subfertile women by choosing a strategy of HSG first, followed by diagnostic laparoscopy in case HSG shows tubal abnormalities. This is in line with the findings in this thesis, that if tubal tests are warranted, use of HSG followed by diagnostic laparoscopy in case of tubal abnormalities is more cost-effective than immediate laparoscopy (chapter 7).

Implications for future research

In this thesis we have come to the conclusion that performance of routine tubal tests are not cost-effective and that HSG is preferred above diagnostic laparoscopy and dye when information on tubal status is required. The need for information may be important for women who want to be reassured that the tubes are patent, especially if they have a preference for expectant management or IUI in case of poor prognosis or show reluctance towards IVF treatment. A preference study about tubal patency tests in subfertile women is required to address this. We developed a model based on the medical history, to decipher women at high risk of tubal pathology from those at low risk of tubal pathology. Although this model showed good calibration in the population that we studied, external validation is still required.

Another potential study would be an adequately powered randomised comparison of HSG and DL, or a less invasive alternative for DL, for example transvaginal hydrolaparoscopy (THL). THL has the advantage that it can be performed in an outpatient setting without general anaesthesia. Our group is preparing a randomised comparison of HSG and THL in 1300 patients.

What has not been addressed in our discussion thus far, is the possibility of a therapeutic effect of oil based contrast media (OBCM) which can be used at HSG. In the literature there is a suggestion that if an OBCM is used instead of a water based contrast medium (WBCM), this may result in more natural conceptions in women with unexplained subfertility and at low risk for tubal pathology (Johnson et al., 2010). If this is the case, routine HSG in all subfertile women at low
risk for tubal pathology may be advised. At present a randomised multicentre trial, comparing the use of OBCM and WBCM at HSG in women at low risk of tubal pathology, is being conducted (H2Oil study, NTR3270).

Another issue that should be addressed in future research is what the implications of unilateral tubal pathology are. We found a 20% reduction in natural conception chances if this was present at HSG or diagnostic laparoscopy. Although this reduction was not statistically significant, the clinical consequences of these findings are best addressed by randomised comparison between IUI and IVF in these women, if they have a prognosis for natural conception below 30% according to the Hunault prediction model (Hunault et al., 2004).

We have commented on the limitations of the available evidence. For some of these limitations, Individual Patient Data (IPD) analyses may provide the necessary corrections (Broeze et al., 2009). This will be dealt with in a separate thesis (Broeze KM). In the epilogue we will take the results of the IPD-analyses and other studies into account to come to a framework for an evidence based guideline.
References


H2Oil study: www.studies-obsgyn.nl/h2olie


