Diagnosing tubal pathology: The individual approach

Broeze, K.A.

Citation for published version (APA):
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 (Hull et al., 1985). Data from historical populations estimate the average prevalence of subfertility to be 5.5%, 9.4% and 19.7%, respectively, at ages 25-29 years, 30-34 years and 35-39 years (Bongaarts, 1982). In the Netherlands, this percentage is between 12% and 17%, depending on female age (Bonsel and van der Maas, 1994).

A diagnosis of unexplained subfertility is usually made after it has been demonstrated that a woman has a regular ovulation, patent Fallopian tubes with no evidence of peritubal adhesions, fibroids or endometriosis and that the male partner has normal semen quality (Simon and Laufer, 1993). The standard fertility work-up starts with a thorough medical history and physical examination followed by the assessment of ovulation and a semen analysis. Assessment of tubal patency is usually reserved as last test in this work-up. Tubal disease includes tubal obstruction and pelvic adhesions due to infection, endometriosis and previous surgery and accounts for subfertility in 10-30% of the couples (Evers, 2002). Several diagnostic tests to assess the tubal condition are available to the clinician.

**Hysterosalpingography (HSG)**

HSG was first performed in 1910 by Rindfleish, who injected a bismuth solution in the uterine cavity (Rindfleish, 1910). In 1914, the first report on HSG for determining tubal patency with an oil-soluble contrast medium was published (Carey, 1914). Carey used Collargol, which could cause significant tissue damage and was painful to the women (Hanlo, 1930). Because of these serious adverse events, its use was abandoned and a tubal insufflation test was introduced by Rubin in 1920 (Rubin, 1920). Rubin insufflated oxygen (later carbon dioxide) under pressure through the cervical canal into the uterine cavity. Tubal patency was determined by presence of air under the diaphragm on X-ray, by auscultation of airflow into the abdomen or a drop in pressure during insufflation. In 1925, Heuser was the first to report on the use of lipiodol in HSGs (Heuser, 1925). Lipiodol, like Collargol an oil-soluble contrast medium but with a lower viscosity, proved less toxic and was widely accepted (Hanlo, 1930). Gradually lipiodol was replaced by a water-soluble contrast medium (WSCM) for several reasons: lower cost, better imaging of the tubal mucosal folds and ampullary rugae (Soules and Spandoni, 1982), lower viscosity and more prompt demonstration of tubal patency reducing the need for a delayed film, lower likelihood of persistence of contrast medium within the pelvic cavity and of complications such as anaphylaxis or long-term lipogranuloma formation. At HSG the contrast medium is slowly injected through the cervical canal into the uterine cavity. By fluoroscopy (continuous X-ray imaging) the flow of contrast can be followed and the uterine cavity and lumen of the fallopian tube can be visualized. HSG is performed as an outpatient procedure and can be painful. Some clinics still use oil-soluble contrast medium (OSCM), because no severe side-effects have been reported since the introduction of fluoroscopy screening during injection of contrast and the abandonment of the procedure when intravasation of contrast occurs (Lindequist et al.,...
Laparoscopy and dye
The first diagnostic laparoscopy (DL) in humans was applied by Jacobeus in Sweden in 1910, by inducing a pneumoperitoneum and introducing a cystoscope into the peritoneal cavity. Initially this technique to diagnose intra-abdominal abnormalities was mainly used by physicians. Kalk in Germany was principally responsible for developing laparoscopy into an effective diagnostic and surgical procedure in the early 1930s (Te Linde’s Operative Gynecology, 1992; Trimbos-Kempers, 1981). It was not until 1946 before Palmer in France introduced laparoscopy and dye test (synonyms: dye pertubation, chromopertubation) into gynaecologic practice for the assessment of tubal patency and published his first 250 cases (Palmer, 1946; Palmer 1962). In this procedure, which requires general anaesthesia and operating facilities, a water-soluble dye (usually methylene blue), is injected into the uterine cavity. Through the laparoscope, the pelvic cavity can be inspected for presence of pelvic-peritoneal adhesions, endometriosis and tubal patency. Although the uterine anatomy can be assessed, intrauterine abnormalities can only be visualized if the procedure is combined with hysteroscopy. This can easily be performed in the same setting. Like HSG, DL is an invasive procedure and can result in complications such as infection, vascular-, intestinal- and urologic injuries, which have been reported to vary between 0.06 and 1.5% (Jansen et al., 1997; Chapron et al., 1998; Härkki-Siren et al., 1999). In the nineties of the previous century, laparoscopy was integrated as an essential diagnostic procedure in infertility protocols (ASRM guidelines, 1992; WHO/ Rowe et al., 1993), based on one publication that reported abnormal findings at laparoscopy in 75% of 24 couples with otherwise unexplained subfertility (Drake et al., 1977). In 1997, it was reported that 96% of the reproductive endocrinologists in the USA routinely performed an HSG and 89% a diagnostic laparoscopy in the diagnostic fertility work-up (Glatstein et al., 1997).

A diagnostic laparoscopy can be combined with interventions like coagulation or excision of minimal to mild endometriosis, adhesiolysis or cystectomy. In a randomized controlled trial, it was found that laparoscopic resection or ablation of minimal and mild endometriosis and adhesiolysis enhances fecundity in subfertile women (Marcoux et al., 1997). They reported
that one in eight women with this condition should benefit from surgery, but the monthly
fertility rate among women who underwent surgery (6.1%) was much lower than the rate
expected in fertile women (20%). Whether DL with dye perturbation has a positive effect
on pregnancy rate is not known. One randomised controlled trial showed no difference
between perturbation with an OSCM versus a WSCM (Al-Fahdli et al., 2006).

_Chlamydia Trachomatis antibody testing_
Chlamydia Trachomatis infection, a sexually transmitted disease (STD), is a major cause
of pelvic inflammatory disease (PID), which can lead to chronic abdominal pain, ectopic
pregnancy and tubal pathology (Weström and Wolner-Hanssen, 1993; Paavonen and
Eggert-Kruse, 1999). The causal relationship between Chlamydia Trachomatis and genital
tract infection was established in 1958 (Collier et al., 1958). C. Trachomatis targets the
epithelial cells of the cervix. In the Netherlands an estimated 35 000 cases of Chlamydia
cervicitis occur annually (Health Council of the Netherlands, 2004). Early infection can be detected by culture of cervical swabs and treated effectively with
antibiotics. However, the majority of these infections is asymptomatic and remains therefore
undiagnosed. Most women will effectively clear a Chlamydia trachomatis infection by a
normal immune response and have a low risk of complications. Still, untreated infection can
through ascendance result in PID and tubal factor subfertility (Peipert, 2003; Den Hartog et
al., 2006). There are no prospective controlled trials available on how frequently women
tested positive for Chlamydia cervicitis will develop tubal infertility (Land et al., 2010). A
correlation between the age of a woman, severity of infections, the number of PID episodes
and the risk of tubal pathology has been demonstrated (Weström, 1980; Weström et al.,
1992). The risk of tubal pathology varied from 10% after one episode of PID to 20% after
two episodes and 40% after three episodes (Weström et al., 1992).

Since an association between C. Trachomatis IgG antibodies in serum and tubal pathology
was noted, Chlamydia Antibody tests (CAT) were introduced as screening tests in the work-
up of sub fertile couples to provide the clinician with a risk estimate of the presence of tubal
pathology (Punnonen et al., 1979; Den Hartog et al., 2006). CAT is a non-invasive, non-
expensive and easy to perform test, but does not provide images of the uterine and tubal
anatomy nor of the pelvis.

_Other diagnostic tests_
Apart from HSG, Laparoscopy and CAT testing, other tools are also used to assess tubal
patency. Falloposcopy is a technique with which through a thin fibreoptic scope the lumen
of the fallopian tubes can directly be visualized (Kiss et al., 1993; Dechaud et al., 1998;
Lundberg et al., 1998). Hystero-contrast-sonography (HyCoSy) is a method where tubal
patency is tested by injection of contrast into the uterine cavity and patency is assessed by
transvaginal ultrasound (Schlief and Deichert, 1991). Transvaginal hydrolaparoscopy (THL)
is performed as a needle puncture technique of the pouch of Douglas using an adjusted
Veress-needle trocart system and warm normal saline is injected through this system into the pelvic cavity. It allows direct visualization of peritubal- and peri-adnexal adhesions and endometriotic implants. Tubal patency can be assessed by injection of methylene blue into the uterine cavity like at laparoscopy (Gordts et al., 1998). These techniques will not be further discussed, because few studies are available that evaluate the diagnostic accuracy and predictive performance on pregnancy rates of these tests.

The performance of HSG, laparoscopy and CAT in terms of accuracy and fertility prognosis

Since the early nineteen seventies many studies on the performance of HSG and CAT for the assessment of tubal pathology have been published. In these studies laparoscopy is usually the reference test. The ideal (or ‘gold standard’) test for tubal disease would correctly identify all women with tubal disease. It would be a sensitive test (i.e. the test result would be positive in all women with the disease) and it would also be specific (i.e. the test result would be negative only in women without the disease).

In a systematic review from 20 studies published between 1968 and 1994 on the diagnostic accuracy of HSG compared with laparoscopy, only 3 studies could be identified in which the judgement of tubal pathology at laparoscopy was independent of the knowledge of HSG results and thus preventing observer bias (Swart et al., 1995). A meta-analysis of these studies showed a sensitivity of 65% and a specificity of 83% for HSG for diagnosing tubal patency. HSG was unreliable for diagnosing peri-adnexal adhesions (Swart et al., 1995). The accuracy of serum Chlamydial antibodies (CAT) was assessed in another meta-analysis (Mol et al., 1997). This study showed that the discriminative capacity of Chlamydia antibody titers depended on the immuno-assays used and that the diagnosis of any tubal pathology was comparable with that of HSG in diagnosing tubal occlusion (Mol et al., 1997). The negative predictive value (NPV) of CAT in subfertile women was 75-90%, making the presence of tubal pathology in women with a negative CAT less likely, whereas the positive predictive value (PPV) of CAT ranged between 30 and 65%, indicating a high false-positive rate if laparoscopy is performed in CAT positive women (Den Hartog et al., 2006).

In a study performed in 1995, diagnostic laparoscopy failed to be an ideal predictor for future fertility (Collins et al., 1995). In a prospective cohort study of 794 women, laparoscopy performed better than HSG as a predictor of future fertility (Mol et al., 1999). In this study HSG showed a fecundity rate ratio (FRR) of 0.8 and 0.49 for one-sided and two-sided tubal occlusion, whereas in laparoscopy these were 0.51 and 0.15, respectively. After a normal HSG or a HSG with one-sided occlusion, laparoscopy showed two-sided occlusion in 5% of these women and their fertility prospects were almost zero. If two-sided tubal occlusion was detected on HSG but not during laparoscopy, fertility prospects were slightly impaired. Fertility prospects after bilateral occlusion on HSG were strongly impaired in case laparoscopy showed one-sided and two-sided occlusion, with FRRs of 0.38 and 0.19, respectively.
Background of the research

Many reports of diagnostic tests, including the before mentioned studies, contain methodological shortcomings that can result in an overestimation of diagnostic test accuracy (Lijmer et al., 1999). A number of issues in the design and conduct of diagnostic accuracy studies can lead to bias or variation. The sources of bias and variation for which there is the most evidence are demographic features, disease prevalence and severity, partial verification bias, clinical review bias and observer and instrument variation (Whiting et al., 2004). In diagnostic accuracy studies the results of a test are compared with those of a reference standard, the best available method to detect presence or absence of disease. These results are then expressed in accuracy statistics, such as sensitivity and specificity. In studies on the prognostic performance of tests, test results are compared with a clinically relevant end point and statistics are calculated to express the predictive capacity of the test. A group of scientist and editors developed the STARD (Standards for Reporting of Diagnostic Accuracy) statement to improve the reporting quality of diagnostic accuracy and to help authors and readers to assess the potential for bias in a study and to evaluate the generalizability of the results (Bossuyt et al., 2003).

Similarly, for improvement on the reporting of meta-analyses, QUORUM (Quality of Reporting of Meta-analyses) and MOOSE (Meta-analyses of observational Studies in Epidemiology) are available (Moher et al., 2001; Stroup et al., 2000). A problem with these conventional meta-analyses in diagnostic studies is that they statistically pool the results of individual diagnostic studies which typically examine the accuracy of a test in isolation from medical history and clinical examination, or do not adjust for overlap of information captured by clinical history, physical examination and additional tests. How useful a test will be in practice remains therefore uncertain (Bachman et al., 2003; Khan et al., 2003; Miettinen and Caro, 1994; Broeze et al., 2009).

Not only in meta-analyses, but also in individual studies, the diagnostic performance of HSG and CAT has been assessed in isolation of patient characteristics, and the sensitivity and specificity of HSG and CAT have been assumed to be stable across subgroups of women (Swart et al., 1995; Mol et al., 1997; Perquin et al., 2006). Since conventional systematic reviews and meta-analyses are based on aggregate data at the study level and not at the level of subgroups, this is unavoidable. The use of data at the patient level in an individual patient data (IPD) meta-analysis could overcome this limitation, because the accuracy of tests can be assessed in relation to other patient characteristics and allows the development or evaluation of diagnostic algorithms for individual patients (Broeze et al., 2009).

Despite clinical research done so far and reflecting the limitations of this research as discussed, there is still inconsistency between fertility guidelines about the best diagnostic strategy for tubal patency testing (NICE, 2004; ASRM, 2006; NVOG, 2004) and there is no consensus on which test should be initially used, or on the most effective or cost-effective sequence of tests (Fatum et al., 2002; Perquin et al., 2006; Bosteels et al., 2007; Den Hartog
et al., 2008). In many fertility clinics, CAT is used ahead of more invasive tests such as HSG or laparoscopy, whereas in other clinics CAT is not used at all and all women are directly referred for HSG or even laparoscopy. Because of these inconsistencies in tubal assessment of subfertile women and the importance of early identification of women with bilateral tubal pathology, who have significantly reduced conception chances, our research has focused on the effective and efficient use of tools and resources available in clinical practice. We therefore critically reappraised the literature dealing with tubal pathology in the subfertile population by using STARD criteria and IPD meta-analysis of diagnostic and prognostic studies if possible and used the data obtained from a large multicentre prospective cohort study among subfertile couples that had completed their basic fertility work-up collected in 38 clinics in the Netherlands between 2002 and 2004. The results of these studies will be presented in 3 theses.
- Diagnostic tests for tubal pathology from a clinical and economic perspective (H.R. Verhoeve)
- Diagnostic and prognostic aspects of tubal patency testing (S.F. Coppus)
- Diagnosing tubal pathology -the individual approach- (K.A. Broeze)

Background and outline of thesis

In daily practice, all diagnostic processes start with medical history taking and physical examination before diagnostic tests are performed. For diagnosing tubal pathology the current guidelines all recommend medical history and physical examination as the primary evaluation in the fertility work-up. Over decades, many studies have been performed to establish the optimal strategy to diagnose tubal pathology in subfertile women. Initially, diagnostic studies only focused on the tubal patency test of interest. Later, also information from medical history and physical examination was taken into account to identify women at risk for tubal pathology. To estimate the optimal work-up for tubal pathology for an individual patient requires integrating all available data from medical history, physical examination, tubal patency tests and subsequent pregnancy rates. Using these data from existing, original, studies and pooling them in a meta-analysis supplies a maximum amount of data and provides the highest level of evidence. Unfortunately, the methodology of diagnostic and prognostic studies hampers optimal use of their data in a meta-analysis, since they typically examine the accuracy of a test in isolation from medical history and clinical examination. The use of data at the patient level in an individual patient data (IPD) meta-analysis could overcome this limitation, because the accuracy of tests can be assessed in relation to other patient characteristics and this approach allows the development or evaluation of diagnostic algorithms for individual patients.
In this thesis we perform four individual patient data (IPD) meta-analyses in which we approach all authors of eligible studies for inclusion of their complete, original data in a summary database.

Since no IPD meta-analyses have ever been performed in diagnostic research before, we first focus on the applicability of this type of research for diagnostic studies in reproductive medicine. We describe the advantages and flaws of this new approach of diagnostic evidence synthesis.

Hereafter, we perform two IPD meta-analyses by which we determine the best performing Chlamydia antibody test (CAT) essay, and investigate the influence of patient characteristics on the diagnostic accuracy of hysterosalpingography (HSG). Next, all available information from clinical history, physical examination is combined with the results of CAT and HSG in a third IPD meta-analysis to estimate the best diagnostic strategy for the diagnosis of unilateral and bilateral tubal pathology.

Finally, in a fourth IPD meta-analysis, we use all available information to predict the chances of natural conception after HSG or diagnostic laparoscopy.

In Chapter 2 we show the advantages and disadvantages of conventional systematic reviews and meta-analyses. We address the differences between this conventional approach and IPD meta-analyses and describe the opportunities for IPD meta-analysis in reproductive medicine, using three clinical examples; ovarian reserve testing, tubal patency testing and single versus double embryo transfer.

In Chapter 3 we introduce IPD meta-analysis for diagnostic and prognostic research in obstetrics, gynaecology and reproductive medicine. The general methodology of this innovative research approach and its applicability for diagnostic studies on, for example, tubal pathology is described.

In Chapter 4 an IPD meta-analysis is performed to re-assess the diagnostic accuracy of three essays of the Chlamydia antibody test. For this meta-analysis we use continuous test results from 6191 women to determine the best performing CAT essay.

In Chapter 5 data from seven primary studies are pooled in an IPD meta-analysis to describe the association of individual patient characteristics from clinical history and physical examination and diagnostic accuracy of hysterosalpingography.

In Chapter 6 we present a multivariable diagnostic approach for subfertile women in which patient characteristics and the results of CAT and HSG are integrated. Based on this IPD meta-analysis we provide a diagnostic score chart to calculate the predicted probability of unilateral and bilateral tubal pathology.

In Chapter 7 a prognostic IPD meta-analysis is performed in which we included data of 6160 women from 8 original studies. Using these data, we assess the predictive capacity of HSG and diagnostic laparoscopy for spontaneous pregnancy chances in women with different profiles and investigate the possible added value of these tubal patency tests over patient characteristics obtained from medical history and physical examination.
In Chapter 8 we provide a general discussion of the results of the studies presented in this thesis and comment on the implications of these findings. In Chapter 9 we give a summary of the data presented in this thesis. Chapter 10 provides an epilogue in which the results of three theses on tubal pathology from our study group are integrated.
References


Chapter 1


