Diagnosing tubal pathology: The individual approach

Broeze, K.A.

Publication date
2013

Chapter 5

Are patient characteristics associated with the accuracy of hysterosalpingography in diagnosing tubal pathology? An individual patient data meta-analysis

Kimiko Broeze
Brent Opmeer
Nan Geloven
Sjors Coppus
John Collins
Janneke den Hartog
Paul van der Linden
Piotr Marianowski
Ernest Ng
Jan Willem van der Steeg
Pieternel Steures
Anika Strandell
Fulco van der Veen
Ben Willem Mol

Human Reproduction Update 2011, 17: 293-300

Abstract

Introduction Conventional meta-analysis has estimated sensitivity and specificity of hysterosalpingography (HSG) to be 65% and 83%. The impact of patient characteristics on the accuracy of HSG is unknown. The aim of this study was to assess by individual patient data meta-analysis whether accuracy of HSG is associated with different patient characteristics.

Material & methods We approached authors of primary studies reporting on the accuracy of HSG using findings at laparoscopy as the reference. We assessed whether patient characteristics such as female age, duration of subfertility and a clinical history without risk factors for tubal pathology were associated with the accuracy of HSG, using a random intercept logistic regression model.

Results We acquired data of seven primary studies containing data of 4,521 women. Pooled sensitivity and specificity of HSG were 53% and 87% for any tubal pathology and 46% and 95% for bilateral tubal pathology. In women without risk factors, the sensitivity of HSG was 38% for any tubal pathology, compared to 61% in women with risk factors (P = 0.005). For bilateral tubal pathology, these rates were 13% versus 47% (P = 0.01). For bilateral tubal pathology sensitivity of HSG decreased with age (factor 0.93 per year (P = 0.05). The specificity of HSG was very stable across all subgroups.

Conclusion The accuracy of HSG in detecting tubal pathology was similar in all subgroups, except for women without risk factors in whom sensitivity was lower, possibly due to false positive results at laparoscopy. HSG is a useful tubal patency screening test for all subfertile couples.

Introduction

Worldwide, 10% of couples trying to conceive are suffering from subfertility. One of the major causes of female subfertility is tubal pathology, with a prevalence of around 30% (Evers, 2002). The diagnostic work-up of subfertile women often includes tubal testing by hysterosalpingography (HSG), an invasive procedure in which an oil- or water based contrast medium is injected through the cervical canal into the uterine cavity and the fallopian tubes. Subsequently, the uterine cavity and the patency of the fallopian tubes can be visualized.
The accuracy of HSG as assessed by conventional meta-analysis showed a sensitivity of 65% and a specificity of 83% (Swart et al., 1995). However, not only in this meta-analysis, but also in individual clinical studies, the diagnostic performance of HSG has been assessed in isolation of patient characteristics obtained from clinical history or physical examination, and sensitivity and specificity of HSG were assumed to be stable across subgroups of women (Broeze et al., 2009; Mol et al., 1997; Perquin et al., 2006; Swart et al., 1995).
Since conventional systematic reviews and meta-analyses are based on aggregate data at the study level, and not at the level of subgroups of women, this is unavoidable. The use of data at the patient level in an individual patient data (IPD) meta-analysis could overcome this limitation and integrate the information of patient characteristics into the analysis of test accuracy (Janes and Pepe 2008).
The aim of this study was to assess whether the diagnostic performance of HSG in diagnosing tubal pathology is associated with patient characteristics by performing an individual patient data meta-analysis.

Methods

Literature search

In a previous meta-analysis on the accuracy of HSG, we identified studies published until June 1994, comparing HSG and laparoscopy results on tubal pathology (Swart et al., 1995). A computerized updated search was performed in Medline and Embase from July 1994 to January 1st 2010, using the words 'hysterosalpingography' or 'hysterosalpingogram' or 'HSG' and 'tubal pathology' or 'tubal disease' or 'fallopian tube disease' or 'tubal occlusion' or 'tubal obstruction' or 'tubal infertility'. Cross-references of the selected articles were searched for other eligible articles. Language restrictions were not applied. Two independent reviewers (KAB and SFC) screened the electronic search results for eligible articles by reading the title and abstract. We asked authors of eligible articles to examine the provisional study list to identify any additional studies they may be aware of. In this way, also data from studies that were missed by our search criteria, or that have not been published at all, were eligible for inclusion. We also considered inclusion of studies that collected relevant data,

but were excluded from the previous meta-analysis due to the inability to extract two by two tables.

Data acquisition

For each of the eligible articles, we obtained contact information on the first, second or last author on Medline, Embase or the internet. We approached authors by mail and invited them to share their data in this collaborative project. In case contact information on the first author was not available or the first author did not respond, we contacted the second or last author. We provided authors that were willing to participate with a more detailed study proposal, and asked them to send their original dataset. We requested the complete database in original format, as to minimise their efforts to select the appropriate variables or to convert data to a specific format. If variables and categories were not adequately labelled within the dataset, a separate data dictionary was requested.

Datasets should at least include the following variables: anonymous patient identifiers, patient characteristics obtained from clinical history or physical examination (e.g. female age or type of subfertility), HSG results and the results of diagnostic laparoscopy (tubal pathology absent or present). Tubal pathology was subdivided in any tubal pathology or bilateral tubal pathology. Any tubal pathology was defined as the presence of occlusion of the fallopian tubes, with or without hydrosalpinges or peritubal adhesions, in at least one of the tubes. Occlusion of the fallopian tubes was considered to be present when there was no filling or spillage of dye at laparoscopy. Bilateral tubal pathology was present when such abnormalities were seen in both tubes. Duration of subfertility was defined as the time between child wish and performance of HSG. The approached authors were asked to indicate whether tubal pathology was unilateral or bilateral. If authors had follow-up data available, they were asked to share these data as well. Approval of the ethical commission was acquired by the original authors.

Quality assessment

We scored the quality of the included studies according to the criteria of the QUADAS checklist (Whiting et al., 2003). Additional items were created for the description of selection criteria, execution of tests and the diagnostic strategy that was used. Completeness of the datasets was described, based on the availability of data on patient identifiers, diagnostic test results and target disease. We compared the acquired data and the published results for consistency. We also checked the included studies for their study characteristics, including study design, inclusion criteria and diagnostic strategy. Participating authors were contacted to confirm missing data or to discuss major discordant results between acquired data and reported data. In addition to this, we organized a collaborators meeting, where authors could clarify details of their original study designs and the performed tests. We used RevMan 5 software (Cochrane Collaboration) to summarize the quality indicators of the included studies according to QUADAS.

Statistical analyses

We merged the data into a summary database when variables were compatible. Incompatible data were recoded and also added to the summary database. First, we estimated prevalences of both any as well as bilateral tubal pathology for the individual studies and for the complete set of included studies. We also estimated sensitivity and specificity of HSG, based on two by two tables comparing the results on HSG and laparoscopy, constructed from the individual patient data.

Second, we performed multiple imputations for missing patient characteristics per individual study (Janssen et al., 2010; Koopman et al., 2008). We also performed multiple imputations per study to correct for missing laparoscopy results, thereby reducing verification bias. To perform such analyses we assumed that, within one study, women that did not have a diagnostic laparoscopy had a tubal status comparable to the tubal status of women with the same HSG result, but who did have a diagnostic laparoscopy (Begg and Greenes, 1983; de Groot et al., 2008). All imputation procedures were performed within each study, and for the multicentre study within each center.

Third, we re-estimated sensitivity and specificity of HSG after imputation of laparoscopies.

Fourth, we estimated the accuracy of HSG for subgroups of women, based on the following characteristics; female age, BMI and type and duration of subfertility. We also created subgroups on history of pelvic inflammatory disease (PID), Chlamydia Antibody Test (CAT) results as well as a subgroup with a clinical history without risk factors, consisting of women without previous PID and with a negative CAT result. Logistic regression models were used to quantitatively estimate the association of each of these patient characteristics on the accuracy of HSG. A random intercept in these models accounted for the heterogeneity in accuracy across studies. Female age, duration of subfertility and BMI were included in these analyses as continuous variables. We used splines to assess the assumption of linearity and performed appropriate transformations. We used two different models. In the first model we estimated sensitivity of HSG in women with tubal pathology. In the second model we estimated specificity of HSG in women without tubal pathology. Patient characteristics were added to these models as covariates. The effects of the covariates in the models on the accuracy of HSG indicated the differences in accuracy across patient subgroups. P-values below 0.05 were considered statistically significant.

Finally, estimated sensitivity and specificity of HSG for different relevant patient subgroups were calculated from these models. All analyses were performed both for any tubal pathology as well as for bilateral tubal pathology. Data were analysed using SPSS 17.0 (SPSS Inc., Chicago, Il, USA) and SAS 9.1 (SAS Institute Inc., Cary, NC, USA).

Results

Literature search and data acquisition

In the previous meta-analysis, 19 studies on tubal pathology were included (Swart et al., 1995). In our current search for new studies on the subject, we detected 2,642 potential relevant titles on Medline and 1,673 potential relevant titles on Embase, reporting on tubal pathology. After reading the abstracts, 181 studies were eligible for full reading. Of the 181 studies, 71 studies were studies containing diagnostic data on HSG. No additional studies were identified in cross-references of the selected articles or by the approached authors. Of the 71 selected articles, 2 studies were duplicate studies and 32 authors were untraceable. Therefore, we contacted 37 authors by mail, of which 23 did not respond and 14 responded in a positive way. Seven authors reported that the data were lost, while the other 7 authors provided their data (Collins, et al., 1993; den Hartog et al., 2008; Marianowski et al., 2007; Mol et al., 1997; Mol et al., 2001; Ng et al., 2001; Steures et al., 2007; Strandell et al., 1999;

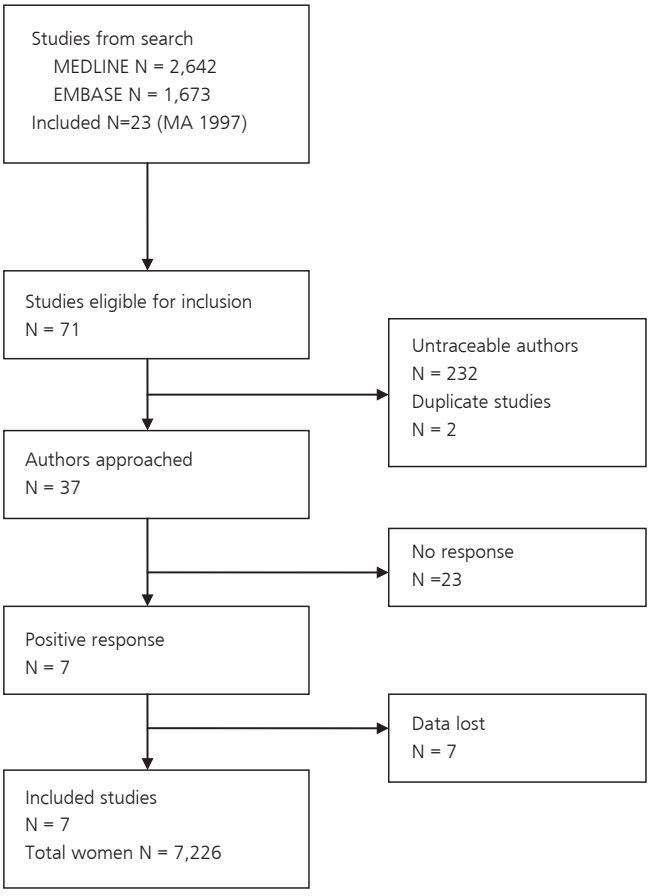

Figure I. Flowchart of included studies

van der Steeg et al., 2008; Veenemans and van der Linden, 2002). The study of van der Steeg et al. was a multicenter trial that contained data of 38 centers. In the analyses all data from this multicenter study were processed as originating from one study. In all included studies in this IPD meta-analysis, the HSG was used for tubal patency testing. A flow chart of the inclusion of studies is shown in figure I.

Finally, data on 7,226 individual women from 7 studies were included in the summary database. For 4,521 women data on tubal status on HSG were available, which were used in the analyses. All included women were referred to a fertility clinic after at least one year of unfulfilled child wish.

Quality assessment

The quality of the received data was considered sufficient for all included studies. An overview of the methodological quality of the included studies according to the criteria of the QUADAS checklist is shown in figure II.

The comparison of consistency between the received data and the published results showed only minimal differences in mean female age and percentage of primary subfertility for 6 studies and were therefore ignored. Study characteristics of the included articles are listed in table I.

Statistical analyses

The number of women available for analysis was 4,521. In 2,632 of these women a laparoscopy had not been performed. These data were imputed.

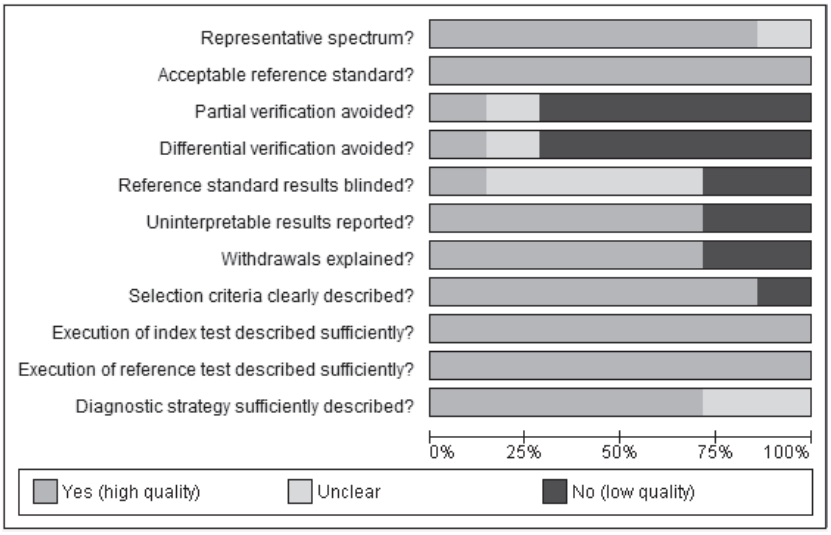

Figure II. Overview of methodological quality of reporting of included studies, according to the QUADAS checklist

Table I. Study characteristics of included studies

Study	Year	Total number women	Study design	Inclusion criteria	Exclusion criteria	Diagnostic strategy
Van der Steeg/ Steures	2007	3,716	Prospective cohort study	Women referred for subfertility work-up	Previous tubal testing Previous tubal surgery	CAT - → No TT CAT + → HSG/DLS
Ng	2001	110	Prospective cohort study	Women referred for subfertility work-up	Previous pelvic surgery Severe male factor	CAT → DLS
Van der Linden	2002	395	Prospective cohort study	Women referred for subfertility work-up	Unknown	CAT - → HSG HSG + → DLS HSG - → EXP → DLS
Den Hartog	2008	642	Prospective cohort study	Women referred for subfertility work-up	Previous pelvic surgery	CAT - → HSG CAT + → DLS HSG + → DLS HSG - → EXP → DLS
Strandell	2004	103	Clinical trial	Women referred for tubal investigation	Unknown	HSG (& HyCoSy) → DLS *1
Marianowski	2007	42	Clinical trial	Women referred for tubal investigation	Unknown	HSG → DLS/ micro DLS
Collins	1999	2,198	Prospective cohort study	Women referred for subfertility work-up		HSG + → DLS HSG - → EXP → *2 DLS

TT: tubal testing (HSG or DLS); EXP: expectative management for at least six months after HSG; HyCoSy: hysterosalpingocontrastsonography; *1 Only a subset of women underwent DLS

The overall prevalence of any tubal pathology, as defined on diagnostic laparoscopy, was 30% (95% CI 29% - 32%), with a range across studies from 12% to 38%. The overall prevalence of bilateral tubal pathology was 15% (95% CI 14% - 17%), with a range across studies from 9% to 21%. These prevalences, as well as the baseline patient characteristics of these studies are shown in table II.

Across the individual studies, sensitivity ranged between 46% and 100% and specificity ranged between 73% and 100% when diagnosing any tubal pathology. The unadjusted pooled accuracy of HSG showed a sensitivity of 70% (95 % CI 0.66 - 0.74) and a specificity of 78% (95 % CI 0.75 - 0.80). After imputation of missing laparoscopy results, these rates

Table II. Patient characteristics of included studies

Study	Female age (years) mean (5th – 95th percentile)	Duration subfertility (years) median (range)	BMI (kg/m^2) mean (5th – 95th percentile)	Primary subfertility (%)	HxPID (%)	CAT positivity (%)	Prevalence tubal pathology (%)	
							Any tubal pathology	Bilateral tubal pathology
Vd Steeg/Steures	32.4 (25 – 39)	1.6 (0 - 12)	24 (19 – 34)	62	3	29	27	12
Ng	31.9 (25 – 37)	3.0 (1 - 12)	21 (18 – 26)	75	3	26	27	16
Vd Linden	31.9 (25 – 39)	2.0 (0 - 13)	na	65	na	22	27	na
Den Hartog	30.8 (24 – 37)	1.4 (0 - 9)	na	71	na	21	18	9
Strandell	31.4 (24 – 40)	2.0 (0 - 7)	na	na	na	na	30	12
Marianowski	32.5 (30 – 38)	na	na	na	21	na	12	10
Collins	29.5 (23 – 37)	3.0 (1 - 15)	na	78	na	na	38	21
Pooled data	31.3 (24 - 39)	2.0 (0 - 15)	24 (19 – 33)	64	3	28	30	15

Table III. Overview of sensitivity & specificity of HSG for the individual studies and for the pooled data before and after imputation of laparoscopies

Study	Number of women	Accuracy of HSG (%)			
		Any tubal pathology		Bilateral tubal pathology	
		Sensitivity	Specificity	Sensitivity	Specificity
Van der Steeg/ Steures	710	76	73	62	91
Ng	48	70	86	na	na
Van der Linden	69	73	67	na	na
Den Hartog	96	67	83	na	95
Strandell	41	46	86	na	na
Marianowski	42	100	100	100	97
Collins	883	67	80	65	87
Pooled accuracy before imputation (95% CI)					
Empirical pooled	1,889	70 (66-74)	78 (75-80)	66 (55-75)	91 (89-93)
Pooled accuracy after imputation (95% CI)					
Empirical pooled	4,521	53 (50-57)	87 (86-88)	46 (41-51)	95 (94-95)
Random intercept logistic regression model	4,521	54 (50-58)	88 (86-89)	39 (25-52)	97 (96-97)

were 53% (95% CI 0.50 - 0.57) and 87% (95% CI 0.86 - 0.88) for sensitivity and specificity, respectively.

The results of the logistic regression models, in which we adjusted for the heterogeneity between studies, showed that most patient characteristics, i.e. duration of subfertility, BMI, type of subfertility, history of PID and CAT were not significantly associated with the accuracy of HSG.

In women with a low risk clinical history, the sensitivity of HSG for detecting unilateral tubal pathology was 38% versus 61% in women with a high risk history (P = 0.005). This 'risk' variable was available for 1656 women, in whom the overall sensitivity was 45% and overall specificity was 89%.

For bilateral tubal pathology, sensitivity ranged between 0% and 100% and specificity ranged between 87% and 97% across the individual studies. The pooled estimates for sensitivity and specificity were 66% (95% CI 0.55 - 0.75) and 91% (95% CI 0.89 - 0.93), respectively. After imputation of laparoscopy results, these rates were 46% (95% CI 0.41 - 0.51) and 95% (95% CI 0.94 - 0.95). An overview of the accuracy per study is shown in table III.

In women with a low risk history, sensitivity was only 13 % compared to 47% in women with a high risk history (P = 0.01). This variable was available for 1607 women, in whom the overall sensitivity was 19% and overall specificity was 98%. Specificity of HSG was very stable across all subgroups. For bilateral tubal pathology, sensitivity of HSG decreased with

Table IV. Association between patient characteristics and accuracy of HSG

Table IVA. Association between patient characteristics and accuracy of HSG for any tubal pathology, assessed by random effects logistic regression model

	Accuracy of HSG (%) Any tubal pathology			
	Sensitivity	**P-value**	**Specificity**	**P-value**
Age (mean accuracy)	51 (N =1101)		89 (N=3420)	
Age 25 yrs	52	0.69	89	0.16
Age 30 yrs	51		89	
Age 35 yrs	50		88	
Duration (mean accuracy)	51 (N=1101)		89 (N=3420)	
Duration 1.4 yrs	50	0.23	89	0.24
Duration 1.7 yrs	50		89	
Duration 2.0 yrs	51		89	
BMI (mean accuracy)	51 (N=1101)		89 (N=3420)	
BMI 20	49	0.55	89	0.53
BMI 25	51		88	
BMI 30	53		88	
Type (mean accuracy)	51 (N=1101)		89 (N=3420)	
Type primary	50	0.57	89	0.20
Type secondary	52		87	
HxPID (mean accuracy)	51 (N=1101)		89 (N=3420)	
No HxPID	51	0.98	89	0.54
HxPID	51		85	
CAT (mean accuracy)	51 (N=1101)		89 (N=3420)	
CAT -	48	0.17	89	0.33
CAT +	56		87	
Clinical risk (mean accuracy)	45 (N=335)		89 (N=1321)	
Low risk	38	0.005	90	0.15
High risk	61		86	

Table IVB. Association between patient characteristics and accuracy of HSG for bilateral tubal pathology, assessed by random effects logistic regression model

	Accuracy of HSG (%) Bilateral tubal pathology			
	Sensitivity	**P-value**	**Specificity**	**P-value**
Age (mean accuracy)	35 (N=520)		97 (N=3926)	
Age 25	47	0.05	98	0.64
Age 30	38		98	
Age 35	30		97	
Duration (mean accuracy)	35 (N=520)		97 (N=3926)	
Duration 1.4	33	0.10	97	0.47
Duration 1.7	35		97	
Duration 2.0	36		97	
BMI (mean accuracy)	35 (N=520)		97 (N=3926)	
BMI 20	34	0.72	97	0.87
BMI 25	35		97	
BMI 30	37		97	
Type (mean accuracy)	35 (N=520)		97 (N=3926)	
Type prim	40	0.06	97	0.38
Type sec	29		97	
HxPID (mean accuracy)	35 (N=520)		97 (N=3926)	
No HxPID	35	0.93	97	0.50
HxPID	37		96	
CAT (mean accuracy)	35 (N=520)		97 (N=3926)	
CAT -	33	0.32	97	0.51
CAT +	40		97	
Clinical risk (mean accuracy)	19 (N=107)		98 (N=1500)	
Low risk	13	0.01	98	0.17
High risk	47		97	

increasing age (factor 0.93 per year (P = 0.05)). An overview of differences in the accuracy of HSG for several subgroups is shown in table IV.

Discussion

The accuracy of HSG has often been estimated in previous studies, but always in isolation of patient characteristics.

In this IPD meta-analysis, we assessed this association for two distinct definitions of tubal pathology; one in which any type of tubal pathology was considered abnormal, and one

in which only bilateral abnormalities were considered abnormal. In our opinion, the latter definition is clinically the one most relevant, as these women have virtually no chance of conceiving either spontaneously or after intrauterine insemination. Since we had to combine data from different studies, we used a relatively broad definition of tubal pathology. All data on tubal pathology in the included studies could be matched using that definition. We did not make a distinction between proximal and distal tubal occlusion, because not all studies reported this level of detail and because clinical management and pregnancy chances are the same for proximal and distal tubal pathology (Farhi et al., 2007).

This IPD meta-analysis showed that sensitivity of HSG was not associated with patient characteristics, except for women with a clinical history without risk factors for tubal pathology, consisting of no previous PID and a negative CAT result, in whom sensitivity was significantly lower than in women with risk factors. Specificity of HSG was relatively high and very stable across all subgroups.

An important strength of this study is the availability of a large number of data, from around the world. For continuous patient characteristics as female age, BMI and duration of subfertility, the complete range of values could be included in the analyses, without loss of information. This enabled us to estimate the association between the accuracy of HSG and patient characteristics with a robust statistical power.

IPD meta-analyses are prone to limitations as well. One of the major issues in IPD meta-analysis is the problem of missing data, both on the study level, as well as on the patient level, including both missing patient characteristics as well as missing laparoscopy results.

As shown in the flowchart, not all eligible studies could be included in this meta-analysis, due to lack of information on the authors, lack of response from the authors or loss of data. The exclusion of these missing studies may have altered the absolute accuracy estimate of HSG, but not the associations between several patient characteristics and the accuracy of HSG, which is the main outcome of this study. Comparison of the patient selections in the studies that were not available for this IPD meta-analysis with the selection of women in the included studies showed no major differences.

Not all original studies contained the same patient characteristics in their databases and even if they had, often data were missing for some women. We decided to impute such missing patient characteristics, since this would prevent the exclusion of observed HSG results from women for which some patient characteristics were not available. It has been reported in literature that imputing such variables is a better option than ignoring them (Janssen et al., 2010). We also compared the patient characteristics between patients included in the analyses and patients excluded from the analyses, which showed no differences.

Another point of attention is the issue of missing laparoscopy results and the presence of partial verification. Partial verification occurs when a set of women that does not undergo the reference standard, is not comparable to the set of women that does undergo the reference standard. When the non-verified women (i.e. women without laparoscopy) are excluded from the analyses, verification bias is introduced (de Groot et al., 2008; van der

Heijden et al., 2006). As shown in table I, in some studies women with a normal HSG received expectative management and no laparoscopy was performed. This was also illustrated by the amount of missing values on laparoscopy for these women, which was 70%, versus 34% missing values on laparoscopy in HSG positive women. To correct for the verification bias that would be introduced when all women with missing laparoscopy results were omitted and analyses were restricted to complete cases, we performed multiple imputations, in which we imputed these missing laparoscopies (van der Heijden et al., 2006). The resulting decrease in sensitivity and increase in specificity compared to the original studies and conventional meta-analysis can be explained by this correction, since omitting of correction of partial verification bias, usually leads to overestimation of sensitivity, with underestimation or varying effects on specificity (Leeflang et al., 2008; Lijmer et al., 1999; Rutjes et al., 2006). Although imputation of missing laparoscopies has changed the accuracy of HSG, there is no reason to assume that imputing the missing laparoscopies will influence the association between accuracy and patient characteristics, since all missing laparoscopy results were imputed independently from the patient characteristics. Furthermore, restricting the analyses to women that have both had HSG and laparoscopy does not reflect daily practice, where most patients first receive a HSG, followed by expective management in case of a normal HSG. In original studies it is therefore hardly feasible to immediately verify the diagnosis in all women, whereas in this IPD meta-analysis we were able to correct for this partial verification, thereby reducing biased accuracies.

The assumption that the diagnostic performance of HSG is stable across patient subgroups has never been tested explicitly before, but can now be supported by this study. This implies that further research to assess the best diagnostic strategy for subfertile women will not be influenced by differences in accuracy across different subgroups of women, except for women with a low risk clinical history, in whom HSG was shown to have a low sensitivity. This finding suggests that these low risk women, without PID and with a negative CAT, have a normal HSG, but show abnormalities on laparoscopy. Clinically, the most likely explanation for this is probably not the failure of HSG to detect tubal pathology, but artefacts at laparoscopy. Although laparoscopy is considered to be the 'gold standard', it might not be a perfect reference standard. The following artefacts may occur at laparoscopy: vaginal leakage of dye, low pressure at chromopertubation, immature ending of the procedure, differences in flow when one tube is patent, or invisible fimbrial ends due to obesity, previous appendectomy, or view-blocking intestines (Mol et al., 1996). These laparoscopic artefacts might result in erroneous interpretation of the HSG. Specificity of HSG was relatively high in all subgroups of women and, also in the low risk group low numbers of false positive HSG results were observed. This means that tubal spasms at HSG are apparently a minor problem. Since HSG was used as a diagnostic test in all included studies in this meta-analysis, pregnancy rates were not reported. Therefore, the possible benefits of flushing of tubal mucus providing potential therapeutic fertility enhancement could not be observed in this study.

In conclusion, our results showed the accuracy of HSG is stable and not associated with any of the patient characteristics assessed in this study, except for women without risk factors. In these women, the sensitivity of HSG was low, which could be possible due to laparoscopic artefacts, leading to false positive laparoscopy results and explaining the decrease in sensitivity. Therefore, HSG can be considered as equally useful to detect tubal pathology for all groups of women. Although some women may still benefit from laparoscopy, HSG can be used as a screening test for all subfertile couples.

References

Begg CB and Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983:39;207-215.

Broeze KA, Opmeer BC, Bachmann LM, Broekmans FJ, Bossuyt PM, Coppus SF, Johnson NP, Khan KS, ter Riet G, van der Veen F, Mol BW. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. BMC Med Res Methodol 2009:9;22.

Collins JA, Burrows EA, and Willan AR. Occupation and the follow-up of infertile couples. Fertil Steril 1993:60;477-485.

Evers JL. Female subfertility. Lancet 2002:360;151-159.

Farhi J, Ben-Haroush A, Lande Y, and Fisch B. Role of treatment with ovarian stimulation and intrauterine insemination in women with unilateral tubal occlusion diagnosed by hysterosalpingography. Fertil Steril 2007:88;396-400.

de Groot JA, Janssen KJ, Zwinderman AH, Moons KG, and Reitsma JB. Multiple imputation to correct for partial verification bias revisited. Stat Med 2008:27;5880-5889.

den Hartog JE, Lardenoije CM, Severens JL, Land JA, Evers JL, and Kessels AG. Screening strategies for tubal factor subfertility. Hum Reprod 2008:23;1840-1848.

van der Heijden GJMG, Donders ART, Stijnen T, and Moons KG. Imputation of missing values is superior to complete case analysis ans the missing-indicator method in multivariable diagnostic research: A clinical example. J Clin Epi 2006:59;1102-1109.

Janes H and Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. Am J Epidemiol 2008:168;89-97.

Janssen KJM, Donders ART, Harrell FE jr, vergouwe Y, Chen Q, Grobbee DE, and Moons KGM. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epi 2010:63;721-727.

Koopman L, van der Heijden GJ, Grobbee DE, and Rovers MM. Comparison of methods of handling missing data in individual patient data meta-analyses: an empirical example on antibiotics in children with acute otitis media. Am J Epidemiol 2008:167;540-545.

Leeflang MM, Deeks JJ, Gatsonis C, and Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008:149;889-897.

Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, and Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999:282;1061-1066.

Marianowski P, Kaminski P, Wielgos M, Szymusik I, and Ludwikowski G. Comparison of tubal patency assessment during microlaparoscopy and laparoscopy, and its compatibility with previous histerosalpingography results. Neuro Endocrinol Lett 2007:28;149-152.

Mol BW, Swart P, Bossuyt PM, van Beurden M, and van der Veen. Reproducibility of teh interpretation of hystersalpingography in the diagnosis of tubal pathology. Human Rep 1996:11;1204-1208.

Mol BW, Swart P, Bossuyt PM, and van der Veen. Is hysterosalpingography an important tool in predicting fertility outcome? Fertil Steril 1997:67;663-669.

Mol BW, Collins JA, van der Veen F, and Bossuyt PM. Cost-effectiveness of hysterosalpingography, laparoscopy, and Chlamydia antibody testing in subfertile couples. Fertil Steril 2001:75;571-580.

Ng EH, Tang OS, and Ho PC. Measurement of serum CA-125 concentrations does not improve the value of Chlamydia trachomatis antibody in predicting tubal pathology at laparoscopy. Hum Reprod 2001:16;775-779.

Perquin DA, Dorr PJ, de Craen AJ, and Helmerhorst FM. Routine use of hysterosalpingography prior to laparoscopy in the fertility workup: a multicentre randomized controlled trial. Hum Reprod 2006:21;1227-1231.

Rutjes AW, Reitsma JB, Di NM, Smidt N, van Rijn JC, and Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006:174;469-476.

Steures P, van der Steeg JW, Hompes PG, Bossuyt PM, Habbema JD, Eijkemans MJ, Koks CA, Boudrez P, van der Veen F, and Mol BW. The additional value of ovarian hyperstimulation in intrauterine insemination for couples with an abnormal postcoital test and a poor prognosis: a randomized clinical trial. Fertil Steril 2007:88;1618-1624.

Strandell A, Bourne T, Bergh C, Granberg S, Asztely M, and Thorburn J. The assessment of endometrial pathology and tubal patency: a comparison between the use of ultrasonography and X-ray hysterosalpingography for the investigation of infertility patients. Ultrasound Obstet Gynecol 199:14;200-204.

Swart P, Mol BW, van der Veen F, van BM, Redekop WK, and Bossuyt PM. The accuracy of hysterosalpingography in the diagnosis of tubal pathology: a meta-analysis. Fertil Steril 1995:64;486-491.

van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Michgelsen HW, van der Heijden PF, Bossuyt PM, van der Veen F, and Mol BW. Predictive value of pregnancy history in subfertile couples: results from a nationwide cohort study in the Netherlands. Fertil Steril 2008:90;521-527.

Veenemans LM and van der Linden PJ. The value of Chlamydia trachomatis antibody testing in predicting tubal factor infertility. Hum Reprod 2002:17;695-698.

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, and Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol, 2003:3;25.