Hysterosalpingosonography for diagnosing tubal occlusion in subfertile women: a systematic review with meta-analysis

S. Maheux-Lacroix1,2,*, A. Boutin2, L. Moore2, M.-E. Bergeron1,2,3, E. Bujold1,2, P. Laberge1,2, M. Lemyre1,2, and S. Dodin1,2

1Department of Obstetrics, Gynecology and Reproduction, Université Laval, 2325, rue de l’Université, Québec, QC, Canada G1V 0A6 2CHU de Québec Research Center, 2705, boul. Laurier, Québec, QC, Canada G1V 4G2 3Oxford Fertility Unit, Institute of Reproductive Sciences, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford OX4 2HW, UK

* Correspondence address. E-mail: sarah.maheux.lacroix@gmail.com

Submitted on September 23, 2004; resubmitted on January 15, 2014; accepted on January 23, 2014

STUDY QUESTION: Is hysterosalpingosonography (sono-HSG) an accurate test for diagnosing tubal occlusion in subfertile women and how does it perform compared with hysterosalpingography (HSG)?

SUMMARY ANSWER: sono-HSG is an accurate test for diagnosing tubal occlusion and performs similarly to HSG.

WHAT IS KNOWN ALREADY: sono-HSG and HSG are both short, well-tolerated outpatient procedures. However, sono-HSG has the advantage over HSG of obviating ionizing radiation and the risk of iodine allergy, being associated with a greater sensitivity and specificity in detecting anomalies of the uterine cavity and permitting concomitant visualization of the ovaries and myometrium.

STUDY DESIGN, SIZE, DURATION: A systematic review and meta-analysis of studies published in any language before 14 November 2012 were performed. All studies assessing the accuracy of sono-HSG for diagnosing tubal occlusion in a subfertile female population were considered.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We searched Medline, Embase, Cochrane Library, Web of Science and Biosis as well as related articles, citations and reference lists. Diagnostic studies were eligible if they compared sono-HSG (+ HSG) to laparoscopy with chromotubation in women suffering from subfertility. Two authors independently screened for eligibility, extracted data and assessed the quality of included studies. Risk of bias and applicability concerns were investigated according to the Quality Assessment of Diagnostic Accuracy Study (QUADAS-2). Bivariate random-effects models were used to estimate pooled sensitivity and specificity with their 95% confidence intervals (95% CIs), to generate summary receiver operating characteristic curves and to evaluate sources of heterogeneity.

MAIN RESULTS AND THE ROLE OF CHANCE: Of the 4221 citations identified, 30 studies were eligible. Of the latter, 28 reported results per individual tube and were included in the meta-analysis, representing a total of 1551 women and 2740 tubes. In nine studies, all participants underwent HSG in addition to sono-HSG and laparoscopy, allowing direct comparison of the accuracy of sono-HSG and HSG. Pooled estimates of sensitivity and specificity of sono-HSG were 0.92 (95% CI: 0.82–0.96) and 0.95 (95% CI: 0.90–0.97), respectively. In nine studies (582 women, 1055 tubes), sono-HSG and HSG were both compared with laparoscopy, giving pooled estimates of sensitivity and specificity of 0.95 (95% CI: 0.78–0.99) and 0.93 (95% CI: 0.89–0.96) for sono-HSG, and 0.94 (95% CI: 0.74–0.99) and 0.92 (95% CI: 0.87–0.95) for HSG, respectively. Doppler sonography was associated with significantly greater sensitivity and specificity of sono-HSG compared with its non-use (0.93 and 0.95 versus 0.86 and 0.89, respectively, \( P = 0.0497 \)). Sensitivity analysis regarding methodological quality of studies was consistent with these findings. We also found no benefit of the commercially available contrast media over saline solution in regard to the diagnostic accuracy of sono-HSG.

LIMITATIONS, REASONS FOR CAUTION: Methodological quality varied greatly between studies. However, sensitivity analysis, taking methodological quality of studies into account, did not modify the results. This systematic review did not allow the distinction between distal and proximal occlusion. This could be interesting to take into account in further studies, as the performance of the test may differ for each localization.

WIDER IMPLICATIONS OF THE FINDINGS: Given our findings and the known benefits of sono-HSG over HSG in the context of subfertility, sono-HSG should replace HSG in the initial workup of subfertile couples.

© The Author 2014. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
Introduction

Hysterosalpingosonography (sono-HSG), an ultrasound-based technique, has been proposed as an alternative to hysterosalpingography (HSG) to assess tubal patency in the initial workup of subfertile couples (NICE, 2013). Sono-HSG and HSG are both short, well-tolerated outpatient procedures (Dessole et al., 2003; Savelli et al., 2009; Socolov et al., 2010). However, sono-HSG has the advantage of obviating ionizing radiation and the risk of iodine allergy associated with HSG (Saunders et al., 2011). Compared with HSG, sono-HSG also has greater sensitivity and specificity in detecting anomalies of the uterine cavity (Saules et al., 2000; Acholonu et al., 2011) and permits concomitant visualization of the ovaries and myometrium (Saunders et al., 2011).

Both sono-HSG and HSG are substitutes for laparoscopy, which is largely accepted as the gold standard for diagnosing tubal occlusion (Mol et al., 1999; Saunders et al., 2011; NICE, 2013). However, being a more costly and invasive test (Saunders et al., 2011), laparoscopy is usually indicated for women who could also benefit from a laparoscopy for the assessment or treatment of another pelvic pathology (NICE, 2013).

A systematic review with meta-analysis comparing sono-HSG and HSG to laparoscopy for diagnosing tubal occlusion was published in Holz et al. (1997). In this review, sono-HSG was associated with a 10% rate of false occlusion and 7% of false patency compared with 13 and 11%, respectively, with HSG. Since then, several reports have been published (Saunders et al., 2011) and the techniques implemented for sono-HSG have greatly improved with the arrival of new contrast media, three-dimensional (3D) ultrasonography, colour-coded 3D power Doppler imaging and heightened ultrasound resolution (Kiyokawa et al., 2000; Sladkevicius et al., 2000; Sankpal et al., 2001).

Our primary objective was to determine the diagnostic accuracy of sono-HSG for detecting tubal occlusion in women suffering from subfertility. Our secondary objective was to compare the diagnostic accuracy of sono-HSG and HSG. We planned to investigate sources of heterogeneity, namely differences in sono-HSG techniques and methodological quality of the included studies.

Materials and Methods

Design

The design of this systematic review was elaborated by a multidisciplinary group of experts (in reproductive endocrinology and infertility, ultrasound in obstetrics and gynaecology, minimally invasive gynaecology and research methodologies) according to approaches outlined in the ‘Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy’ (Deeks et al., 2008). This article was written in accordance with ‘Preferred Reporting Items for Systematic Reviews and Meta-Analyses’ (PRISMA) (Liberati et al., 2009). Our protocol was registered with PROSPERO (#CRD42013003829) and published a priori (Maheux-Lacroix et al., 2013).

Search strategy

We searched Medline, Embase, Cochrane Library and the Web of Science from their inception to 14 November 2012. Biosis was used to identify relevant abstracts and conference proceedings. The search strategy for Medline is presented in Supplementary data, Fig. S1. As recommended in the literature (Leeflag et al., 2006; de Vet et al., 2008; Reitsma et al., 2012), we developed a search strategy with terms related to the index test (sono-HSG) and the target condition (tubal occlusion) and did not use any filter for diagnostic studies to maximize the sensitivity of the search. The strategy was revised by a healthcare librarian and all authors. Finally, we looked at reference lists and citations of relevant articles (previous reviews and included studies) to identify additional eligible reports.

Eligibility criteria and study selection

We considered all studies assessing the accuracy of sono-HSG for diagnosing tubal occlusion in a subfertile female population. There were no restrictions in terms of publication date or language. Articles written in languages other than English or French were translated before completing the selection process and data abstraction. Studies including women suffering from recurrent spontaneous miscarriages were eligible. We excluded studies that undertook sono-HSG to confirm occlusion after tubal sterilization, these samples being significantly different from the population of interest.

We only considered studies using exclusively laparoscopy as reference standard. Studies considering other modalities as reference standard, such as HSG, hysteroscopic selective tubal cannulation under fluoroscopic guidance or vaginal laparoscopy, were excluded. For studies using HSG as a comparator test (that is, additionally assessing the accuracy of HSG compared with laparoscopy), data on the diagnostic accuracy of HSG was retained in order to make a direct comparison of accuracy between sono-HSG and HSG.

We considered consecutive and random series of patients as well as case–control designs.

Studies utilizing random partial verification and non-random partial verification were eligible, provided the determinants of partial verification were known and verification in each strata was random and in known proportions (Irwig et al., 1994; de Groot et al., 2011; de Groot et al., 2012).

Study eligibility was assessed independently by two reviewers screening titles, abstracts and full-text publications, when required. If disagreements were not resolved by consensus, a third reviewer was consulted. Attempts were made to contact the author for further information on studies that fulfilled the eligibility criteria but did not have sufficient data to build 2-by-2 tables. We collected reasons for full-text exclusion. To avoid duplication, author names, sample sizes and study results were compared.

Data abstraction

Two authors independently extracted data from included studies, and disagreements were resolved by discussion. If consensus was not reached, a third reviewer was consulted. We developed a standardized data abstraction form, which was pilot-tested on three studies (Spalding et al., 1997; Inki et al., 1998; Reis et al., 1998) and refined accordingly. The following information was extracted from each of them:

1. Study characteristics and methods: study design, inclusion and exclusion criteria, flow diagram, setting, country, language of publication.
(2) Description of the sono-HSG technique: resolution, two-dimensional (2D) or 3D device, vaginal or abdominal probe, contrast type, catheter type and Doppler sonography.

(3) Measures of diagnostic accuracy of sono-HSG (and HSG when available) in reference to laparoscopy.

In some studies, a positive test was defined as the presence of an occluded tube and, in others, by the presence of a patent tube. In order to pool the results, we reported the data of all studies by considering a positive test as an occluded tube. If published data did not allow us to obtain or derive the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), we attempted to contact the corresponding author of the study.

Assessment of methodological quality

Two persons independently assessed the risk of bias and applicability concerns with a quality checklist derived from the Quality Assessment of Diagnostic Accuracy Study (QUADAS-2; Whiting et al., 2011) and available in our published protocol (Maheux-Lacroix et al., 2013). In case of discrepancy, a third reviewer was consulted. An interval of no more than 1 month between tests was considered appropriate. Sono-HSG (+ HSG) results had to be interpreted without knowledge of the laparoscopy results. Reviewers’ judgments about risks of bias and applicability concerns were used in sensitivity analysis to examine the effect of methodological quality of studies.

Statistical analysis and data synthesis

Meta-analysis was performed by computing the number of TP, TN, FP and FN of each study in bivariate hierarchical random-effects models using SAS 9.3 (SAS Institute, Inc., Cary, NC, USA, 2011). The results were presented with Cochrane Review Manager version 5.2 (The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark, 2012). Pooled and individual estimates of sensitivity and specificity and 95% confidence intervals (95% CIs) were presented in paired forest plots. We generated summary receiver operating characteristic (SROC) curves with point estimates for each study as well as symmetrical summary curves, summary point estimates, 95% confidence region and 95% prediction region. The magnitude of heterogeneity was assessed with the 95% prediction region on SROC curves.

Direct comparison of sono-HSG and HSG accuracy as well as subgroup and sensitivity analyses were achieved using bivariate models. We a priori planned to examine the effect of differences in sono-HSG technique [2D versus 3D, low (<5 MHz) versus high resolution (≥5 MHz), Doppler versus standard sonography, vaginal versus abdominal probe, saline versus other contrast and flexible versus rigid catheter] and in the methodological quality of studies (low versus high or unclear global risks of bias and applicability concerns). We calculated P-values (a P < 0.05 was considered statistically significant) by computing the change in the \(\chi^2\) statistic.

Results

Search results

We identified 4221 citations, with 160 studies that were considered potentially eligible for our systematic review after screening titles and abstracts (Fig. 1). One hundred and twenty-six studies were excluded because they did not meet the inclusion criteria. Four additional studies were excluded because we did not have sufficient data to build 2-by-2 tables and attempts to contact the authors failed. At the end of this process, a total of 30 studies were included in the systematic

---

Figure 1 | Search results for a systematic review of the accuracy of sono-HSG for diagnosing tubal occlusion in subfertile women.
Accuracy of sono-HSG for diagnosing tubal occlusion

Twenty-eight studies, representing 1551 women and 2740 tubes, served to estimate the global accuracy of sono-HSG for diagnosing tubal occlusion. Pooled estimates of sensitivity and specificity were 0.92 (95% CI: 0.82–0.96) and 0.95 (95% CI: 0.90–0.97), respectively. Paired forest plots and the corresponding SROC curve are presented in Fig. 2. The large prediction region on the SROC curve reflects heterogeneity between studies.

Two studies (Randolph et al., 1986; de Almeida et al., 2000) were excluded from the meta-analysis because they did not report results per individual tube. The study by de Almeida et al. (2000) (n = 30 women) achieved sensitivity and specificity of 1.00 (95% CI: 0.40–1.00) and 0.91 (95% CI: 0.80–0.97), respectively, for the detection of bilateral obstruction, and Randolph et al. (1986) study (n = 61 women) reached a specificity of 0.92 (95% CI: 0.74–0.99) for the detection of at least one occluded tube (TP number was unknown, which prevented the calculation of sensitivity).

Direct comparison of sono-HSG and HSG for diagnosing tubal occlusion

Sono-HSG and HSG were directly compared with laparoscopy by chromotubation in nine studies, representing 582 women and 1053 tubes. Figure 3 presents individual and pooled estimates of diagnostic accuracy parameters of both tests and SROC curves. For sono-HSG, pooled estimates of sensitivity and specificity were 0.95 (95% CI: 0.78–0.99) and 0.93 (95% CI: 0.89–0.96), respectively. For HSG, pooled estimates of sensitivity and specificity were 0.94 (95% CI: 0.74–0.99) and 0.92 (95% CI: 0.87–0.95), respectively. Differences between the two techniques were not statistically significant with a P-value of 0.4.

Methodological quality of studies

Figure 4 summarizes the risks of bias and applicability concerns of studies based on QUADAS-2 (the assessment of each individual study is presented in Supplementary data, Fig. S2). Only six studies specified that they presented consecutive series of patients (low risk of bias in terms of patient selection). Most studies (n = 19) reported a clear definition of a positive test and how they ensured adequate blinding (low risk of bias in terms of ‘index test’). In terms of ‘flow and timing’, 10 studies were considered a high risk of bias either because an interval of more than 1 month separated the tests or more than 10% of tubes were excluded from the final analysis. Reasons reported for exclusions were: poor visualization, cervical stenosis, pain, suspicion of hydrosalpinx at sonography (sono-HSG cancelled), pregnancy and the women failed to return for the other test. All included studies had a complete verification design (gold standard performed on all participants; low risk of bias in terms of reference standard). Applicability concerns were raised in 11 studies that either exclusively recruited subfertile women with planned laparoscopy, included women suffering from recurrent pregnancy loss or underwent sono-HSG under general anaesthesia just before laparoscopy (applicability concerns in terms of ‘patient selection’). In summary, 14 studies were considered a high risk of bias or raised applicability concerns for at least one item of the QUADAS-2 tool.

Subgroup and sensitivity analyses

Subgroup and sensitivity analyses are presented in Table II. Doppler sonography for sono-HSG was associated with significantly higher sensitivity and specificity compared with its non-use (0.93 and 0.95 versus 0.86 and 0.89, respectively, P = 0.0497). Estimates were not statistically different when comparing 3D with 2D and saline with other contrasts. Data available for the estimation of pooled sensitivity and specificity with abdominal probes, low-resolution devices (<5 MHz), and rigid catheters came from only four studies (Allahbadia et al., 1992; Omibgodun et al., 1992; Allahbadia, 1993; Kozarzewski et al., 1995) of unclear methodological quality (3/7 to 5/7 items on QUADAS-2 were unclear); therefore, these subgroup analyses were not performed.

Analysis based on global score of risk of bias and applicability concerns revealed that the methodological quality of the studies did not modify pooled estimates of sono-HSG sensitivity or specificity (P = 0.5).

Discussion

In this systematic review, we observed high diagnostic accuracy of sono-HSG for tubal occlusion with overall sensitivity of 0.92 (95% CI: 0.82–0.96) and specificity of 0.95 (95% CI: 0.90–0.97). We also found that the diagnostic accuracy of sono-HSG and HSG was comparable with no significant difference in performance of the two tests.

Sono-HSG presents some advantages over HSG. First, sono-HSG avoids the risk of allergy and ionizing radiation associated with iodine contrast and fluoroscopic guidance of HSG. Secondly, sono-HSG was associated with greater sensitivity and specificity for the detection of intrauterine pathologies, such as fibromas, polyps and synechiae in subfertile populations (Safoares et al., 2000; Acholonu et al., 2011). Finally, unlike HSG, sono-HSG permits concomitant visualization of the ovaries and myometrium, which can result in relevant findings in a context of subfertility, such as polycystic ovaries, endometriomas, other ovarian cysts and Mullerian anomalies (Saunders et al., 2011).

Sono-HSG and HSG are both short and well-tolerated outpatient procedures (Ayida et al., 1996; Dessole et al., 2003; Savelli et al., 2009; Lim et al., 2011). In a series of 1153 sono-HSG (Dessole et al., 2003), adverse effects occurred in only 8.8% of cases (moderate or severe pelvic pain 3.8%, vasovagal symptoms 3.5%, nausea 1.0%, vomiting 0.5% and...
### Table I  Characteristics of included studies in a systematic review of the accuracy of sono-HSG for diagnosing tubal occlusion in subfertile women.

<table>
<thead>
<tr>
<th>Studies</th>
<th>Participants</th>
<th>Tubes (excluded)</th>
<th>Country</th>
<th>Language</th>
<th>Population</th>
<th>Prevalence</th>
<th>Contrast</th>
<th>Catheter</th>
<th>Resolution (MHz)</th>
<th>probe</th>
<th>2D or 3D</th>
<th>Doppler</th>
<th>Comparison with HSG</th>
<th>In MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allahbadia (1992)</td>
<td>129</td>
<td>129 (0)</td>
<td>India</td>
<td>English</td>
<td>Subfertility</td>
<td>0.11</td>
<td>Saline + air</td>
<td>Flexible with balloon (8 Fr) or rigid</td>
<td>3.5–5</td>
<td>Vaginal or abdominal</td>
<td>2D</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Allahbadia (1993)</td>
<td>27</td>
<td>54 (0)</td>
<td>India</td>
<td>English</td>
<td>Subfertility</td>
<td>0.22</td>
<td>Saline</td>
<td>Rigid</td>
<td>2.5</td>
<td>Abdominal</td>
<td>2D</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Allahbadia (1994)</td>
<td>53</td>
<td>106 (0)</td>
<td>India</td>
<td>English</td>
<td>Subfertility</td>
<td>0.11</td>
<td>Saline</td>
<td>Rigid or flexible with balloon (8 Fr)</td>
<td>2.5–5</td>
<td>Vaginal or abdominal</td>
<td>2D</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Allahbadia et al. (1992)</td>
<td>50</td>
<td>100 (0)</td>
<td>India</td>
<td>English</td>
<td>Subfertility</td>
<td>0.15</td>
<td>Saline + air</td>
<td>Flexible with balloon (8 Fr)</td>
<td>5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ayida et al. (1997)</td>
<td>20</td>
<td>32 (7)</td>
<td>England</td>
<td>English</td>
<td>Subfertility</td>
<td>0.12</td>
<td>Galactose</td>
<td>Flexible with balloon (5–7 Fr)</td>
<td>5–6</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Battaglia et al. (1996)</td>
<td>59</td>
<td>114 (4)</td>
<td>Italy</td>
<td>English</td>
<td>Subfertility</td>
<td>0.17</td>
<td>Saline</td>
<td>Flexible with balloon</td>
<td>5–6.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Chan et al. (2005)</td>
<td>21</td>
<td>34 (8)</td>
<td>China</td>
<td>English</td>
<td>Subfertility</td>
<td>0.26</td>
<td>Galactose</td>
<td>Flexible with balloon (8 Fr)</td>
<td>7</td>
<td>Vaginal</td>
<td>3D</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>de Almeida et al. (2000)</td>
<td>30</td>
<td>60 (8)</td>
<td>Brazil</td>
<td>Portuguese</td>
<td>Subfertility</td>
<td>0.04</td>
<td>Galactose</td>
<td>Flexible with balloon (10 Fr)</td>
<td>5</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>De Jonge et al. (2001)</td>
<td>100</td>
<td>178 (22)</td>
<td>South Africa</td>
<td>English</td>
<td>Primary subfertility</td>
<td>0.89</td>
<td>Galactose</td>
<td>Flexible with balloon (5 Fr)</td>
<td>6</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Dietrich et al. (1996)</td>
<td>20</td>
<td>39 (1)</td>
<td>Germany</td>
<td>English</td>
<td>Subfertility</td>
<td>0.13</td>
<td>Saline + Galactose</td>
<td>Flexible with balloon (6 Fr)</td>
<td>5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Dijkman et al. (2000)</td>
<td>100</td>
<td>200 (106)</td>
<td>Netherlands</td>
<td>English</td>
<td>Subfertility</td>
<td>0.28</td>
<td>Saline + Galactose</td>
<td>Flexible with balloon</td>
<td>7.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Friberg and Joergensen (1994)</td>
<td>14</td>
<td>28 (10)</td>
<td>Sweden</td>
<td>English</td>
<td>Subfertility with planned laparoscopy</td>
<td>0.50</td>
<td>Saline</td>
<td>Flexible with balloon</td>
<td>6.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Guerriero et al. (1996)</td>
<td>30</td>
<td>60 (0)</td>
<td>Italy</td>
<td>English</td>
<td>Subfertility</td>
<td>0.30</td>
<td>Galactose</td>
<td>Flexible with balloon (5 Fr)</td>
<td>5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hauge et al. (2000)</td>
<td>33</td>
<td>66 (0)</td>
<td>Norway</td>
<td>English</td>
<td>Subfertility with planned laparoscopy</td>
<td>0.21</td>
<td>Saline + Galactose</td>
<td>Flexible with balloon (4 Fr)</td>
<td>7.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Heikkinen et al. (1995)</td>
<td>31</td>
<td>61 (0)</td>
<td>Finland and Italy</td>
<td>English</td>
<td>Subfertility</td>
<td>0.18</td>
<td>Saline + air</td>
<td>Flexible with balloon (6-10 Fr)</td>
<td>5–6.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Inki et al. (1998)</td>
<td>32</td>
<td>53 (11)</td>
<td>Finland</td>
<td>English</td>
<td>Subfertility</td>
<td>0.23</td>
<td>Saline + air</td>
<td>Flexible with balloon (6 Fr)</td>
<td>6</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Kozarzewski et al. (1995)</td>
<td>25</td>
<td>50 (0)</td>
<td>Poland</td>
<td>Polish</td>
<td>Subfertility</td>
<td>0.20</td>
<td>Galactose</td>
<td>Rigid</td>
<td>3.75</td>
<td>Abdominal</td>
<td>2D</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Kupesic and Plavsic (2007)</td>
<td>268</td>
<td>536 (0)</td>
<td>Croatia</td>
<td>English</td>
<td>Subfertility (91%) and RPL (9%)</td>
<td>0.46</td>
<td>Galactose</td>
<td>Flexible with balloon (5 Fr)</td>
<td>5–7</td>
<td>Vaginal</td>
<td>2D</td>
<td>3D</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Studies</th>
<th>Participants</th>
<th>Tubes(^a) (excluded)</th>
<th>Country</th>
<th>Language</th>
<th>Population</th>
<th>Prevalence(^b)</th>
<th>Contrast</th>
<th>Catheter</th>
<th>Resolution (MHz)</th>
<th>Probe</th>
<th>2D or 3D</th>
<th>Doppler Comparison with HSG(^2)</th>
<th>In MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luciano et al. (2011)</td>
<td>62</td>
<td>121 (0)</td>
<td>USA</td>
<td>English</td>
<td>Subfertility with planned laparoscopy</td>
<td>0.41</td>
<td>Saline + air</td>
<td>Flexible with balloon (5 Fr)</td>
<td>5–9</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Omigbodun et al. (1992)</td>
<td>31</td>
<td>62 (6)</td>
<td>Nigeria</td>
<td>English</td>
<td>Subfertility</td>
<td>0.36</td>
<td>Saline</td>
<td>Rigid</td>
<td>3.5</td>
<td>Abdominal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Radic et al. (2005)</td>
<td>68</td>
<td>135 (0)</td>
<td>Croatia</td>
<td>English</td>
<td>Subfertility (66%) and RPL (34%)</td>
<td>0.35</td>
<td>Saline + Galactose</td>
<td>Flexible with balloon (8 Fr)</td>
<td>6.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Randolph et al. (1986)</td>
<td>61</td>
<td>122 (2)</td>
<td>United States</td>
<td>English</td>
<td>Subfertility and RPL</td>
<td>0.07</td>
<td>Saline</td>
<td>Rigid</td>
<td>3.5</td>
<td>Abdominal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Reis et al. (1998)</td>
<td>44</td>
<td>88 (0)</td>
<td>Brazil</td>
<td>English</td>
<td>Subfertility</td>
<td>0.31</td>
<td>Saline + Galactose</td>
<td>Flexible with balloon</td>
<td>—</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Schwarzler et al. (1997)</td>
<td>57</td>
<td>108 (0)</td>
<td>Austria</td>
<td>German</td>
<td>Subfertility</td>
<td>0.22</td>
<td>—</td>
<td>(6 Fr)</td>
<td>7.5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Socolov et al. (2009)</td>
<td>95</td>
<td>181 (0)</td>
<td>Romania</td>
<td>Romanian</td>
<td>Subfertility</td>
<td>0.36</td>
<td>Saline + air</td>
<td>Flexible with balloon</td>
<td>7</td>
<td>Vaginal</td>
<td>3D</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Spalding et al. (1997)</td>
<td>32</td>
<td>59 (4)</td>
<td>Finland</td>
<td>English</td>
<td>Subfertility with planned laparoscopy</td>
<td>0.22</td>
<td>Saline + air or Galactose</td>
<td>Flexible with balloon</td>
<td>6</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Tanawattanacharoen et al. (1998)</td>
<td>15</td>
<td>25 (5)</td>
<td>Thailand</td>
<td>English</td>
<td>Subfertility</td>
<td>0.36</td>
<td>Galactose</td>
<td>?</td>
<td>5</td>
<td>Vaginal</td>
<td>2D</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Tufekci et al. (1992)</td>
<td>44</td>
<td>88 (16)</td>
<td>Turkey</td>
<td>English</td>
<td>Subfertility</td>
<td>0.28</td>
<td>Saline</td>
<td>Flexible with balloon (8 Fr)</td>
<td>5</td>
<td>Vaginal</td>
<td>2D</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Wang et al. (2012)</td>
<td>70</td>
<td>140 (0)</td>
<td>China</td>
<td>Mandarin</td>
<td>Subfertility</td>
<td>0.46</td>
<td>SonoVue</td>
<td>?</td>
<td>9</td>
<td>Vaginal</td>
<td>2D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watermann et al. (2004)</td>
<td>21</td>
<td>42 (0)</td>
<td>Germany</td>
<td>German</td>
<td>Subfertility with planned laparoscopy</td>
<td>0.12</td>
<td>Galactose</td>
<td>Flexible with balloon (5 Fr)</td>
<td>5–8</td>
<td>Vaginal</td>
<td>3D</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

2D, two-dimensional; 3D, three-dimensional; Fr, French; MA, meta-analysis; RPL, recurrent pregnancy loss.

\(^a\)Number of tubes may not be consistent with the number of patients because of past salpingectomy and unicornuate uterus. In parenthesis is the number of tubes excluded from the analyses for each study. The reasons reported for exclusions were: poor visualisation, cervical stenosis, pain, suspicion of hydrosalpinx at sonography (sono-HSG cancelled), pregnancy and the women failed to return for the other test.

\(^b\)Prevalence of tubal occlusion per tube.
fever 0.8%). Patient tolerance of sono-HSG was compared with HSG in a RCT (Ayida et al., 1996) and no difference in pain, need for analgesia or side-effects was observed at 2 h, 24 h and 28 days after the procedure. There was also no significant difference in terms of procedure duration. The costs of sono-HSG and HSG can vary but are nonetheless considered similar (Lim et al., 2011). However, sono-HSG could be considered as more cost-effective since it allows a complete assessment of the pelvis.

Some authors argue that a complete pelvic ultrasound scan should be part of the evaluation of subfertile women as it provides useful information for treatment decisions and prognosis (Kelly et al., 2001). In fact, complete assessment of the uterus, uterine cavity, endometrium, ovaries, follicles, tubes and their patency could detect relevant anomalies that would otherwise result in prolonged, invasive or unnecessary interventions (Kelly et al., 2001).

We observed heterogeneity of our results, which is to be expected in diagnostic test accuracy reviews (Macaskill et al., 2010). Some between-study heterogeneity could be due to differences in study populations (e.g. BMI, pain tolerance) and in provider expertise (Exacoustos et al., 2009).

Subgroup analyses showed that heterogeneity is partly explained by variations in the sono-HSG technique. Doppler sonography leads to coloration of fluid flow through the tubes and was associated in our review with significantly greater sensitivity and specificity of sono-HSG. Spilling of contrast from the fimbrial end of the tube is difficult to distinguish from the bowel, both of them having similar echogenicity (Sladkevicius et al., 2004).

Figure 2 (A) Forest plot of sono-HSG and (B) summary ROCs (SROC) curve for diagnosing tubal occlusion (per tube).
The benefits of Doppler sonography could be explained by its capacity to enhance the visualization of fluid flow from tubes to the pelvis (Sladkevicius et al., 2000). Hyperechogenic contrast media, e.g. Echovist-200 (Schering AG, Berlin, Germany), ExEm-gel foam (GynaecologIQ, Delft, The Netherlands) and SonoVue (Bracco, Milan, Italy), have been commercialized to facilitate the liquid visualization in the tubes (Exacoustos et al., 2009). In our review, we found no benefit of these contrast media over saline solution in regard to the diagnostic accuracy of sono-HSG. Saline solution mixed with air also has a hyperechoic appearance.
Hysterosalpingosonography for tubal occlusion

et al and is far less expensive than commercial contrast media (Exacoustos et al, 2009). This economical aspect further enhances the advantages of using sono-HSG over HSG.

We observed no significant increase in the diagnostic accuracy of sono-HSG with a 3D device. However, 3D has other benefits compared with 2D that must be considered, namely, it requires less time (Sladkevičius et al, 2000), avoids difficult probe movements and is less dependent on operator skill (Exacoustos et al, 2009). Image acquisition also permits storage and later analyses of captures (Exacoustos et al, 2009).

Our systematic review has some strengths and limitations. First, laparoscopy with chromotubation is largely accepted as the gold standard for diagnosing tubal occlusion (Mol et al, 1999; Saunders et al, 2011; NICE, 2013) as its findings are highly correlated with spontaneous pregnancy rates (Mol et al, 1999), but still diagnostic errors can occur with technical problems (e.g. improper catheter placement, lack of sealing) or severe adhesions (Saunders et al, 2011). In some studies, tubes could not be assessed by either sono-HSG or laparoscopy given a lack of visualization. Such exclusions from the analysis may have affected the validity of the results of the concerned studies. Also, applicability concerns were raised in 2 of the 30 included studies that did not only consider subfertile women, but also women suffering from recurrent pregnancy losses. These aspects were taken into account in the assessment of risk bias and applicability concerns using the QUADAS-2 tool. In a sensitivity analysis, the pooled estimates of sensitivity and specificity of sono-HSG did not significantly differ for the 15 studies that were attributed at least one high risk of bias or applicability concern compared with the other studies included in our systematic review. In other words, the variation in the methodological quality of the included studies did not significantly modify the results of our pooled estimates.

In 28 of the 30 included studies, results were reported per tube, not per woman, which allowed us to calculate pooled estimate per tube. Reporting the results per woman brings the problem of multiple possible definitions of a positive test (both tubes occluded or at least on tube occluded) and, thereby, different values of sensitivity and specificity (Broeze et al, 2012). However, the data for each tube of a woman are not independent. In our analyses, we have not been able to account for this cluster effect, as the results reported for each individual study did not allow us to link tubes together in regards to their belonging to a single woman. To ignore the cluster effect has no effect on the point estimates of sensitivity and specificity but can affect the width of the 95% CI; however, given the high number of women (cluster), the estimates were probably hardly affected. Finally, this systematic review did not allow the distinction between distal and proximal occlusion. This could be interesting to take into account in further studies as the performance of the test may differ for each localization.

Finally, important strengths of our review lay in the a priori nature of our protocol (Maheux-Lacroix et al, 2013) and methods adhering to recent guidelines for diagnostic test accuracy reviews (de Vet et al, 2008; Liberati et al, 2009; Macaskill et al, 2010; Whiting et al, 2011; Reitsma et al, 2012).

In conclusion, sono-HSG is an accurate test for diagnosing tubal occlusion and performs similarly to HSG. Given their comparable patient tolerability and the advantages of sono-HSG over HSG (visualization of ovaries and myometrium, better sensitivity and specificity for the diagnosis of uterine cavity abnormalities, absence of radiation and risk of iodine allergy), sono-HSG should replace HSG in the initial workup of subfertile couples. An economic study comparing cost-utility of sono-HSG and HSG would add an interesting perspective to these considerations. Finally, Doppler sonography potentially improves the diagnostic accuracy of sono-HSG and deserves further evaluation.

### Table II Subgroup and sensitivity analyses.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number of studies*</th>
<th>Number of women</th>
<th>Number of tubes</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doppler</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13</td>
<td>785</td>
<td>1545</td>
<td>0.93 (0.84–0.97)</td>
<td>0.95 (0.92–0.98)</td>
<td>0.0497</td>
</tr>
<tr>
<td>No</td>
<td>15</td>
<td>766</td>
<td>1335</td>
<td>0.86 (0.74–0.93)</td>
<td>0.89 (0.83–0.93)</td>
<td></td>
</tr>
<tr>
<td>2D or 3D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>5</td>
<td>293</td>
<td>641</td>
<td>0.89 (0.77–0.95)</td>
<td>0.94 (0.86–0.98)</td>
<td>0.6703</td>
</tr>
<tr>
<td>2D</td>
<td>26</td>
<td>1258</td>
<td>2281</td>
<td>0.88 (0.80–0.94)</td>
<td>0.92 (0.87–0.95)</td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saline</td>
<td>14</td>
<td>643</td>
<td>1223</td>
<td>0.91 (0.82–0.96)</td>
<td>0.93 (0.88–0.96)</td>
<td></td>
</tr>
<tr>
<td>Other contrasts</td>
<td>16</td>
<td>851</td>
<td>1657</td>
<td>0.87 (0.76–0.94)</td>
<td>0.92 (0.86–0.96)</td>
<td>0.7046</td>
</tr>
<tr>
<td>Risk of bias*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>15</td>
<td>952</td>
<td>1701</td>
<td>0.88 (0.77–0.94)</td>
<td>0.91 (0.85–0.95)</td>
<td></td>
</tr>
<tr>
<td>Low/unclear</td>
<td>13</td>
<td>599</td>
<td>1179</td>
<td>0.91 (0.80–0.96)</td>
<td>0.94 (0.89–0.97)</td>
<td>0.5488</td>
</tr>
</tbody>
</table>

*Studies that reported several techniques were retained in subgroup analyses if they provided data separately for each technique. For this reason, some studies may be counted more than once in each analysis.

Global appreciation for risk of bias and applicability concerns based on the QUADAS-2. CI, confidence interval; 2D, two-dimensional; 3D, three-dimensional.
Authors’ roles

S.M.-L. was involved in data collection, statistical analysis, manuscript preparation and revision, construction of figures and tables and submission of manuscript. A.B. was involved in data collection. L.M. served as advisor on analytical methods. A.B., L.M., M.-E.B., E.B., P.L., M.L., S.D. were involved in interpreting the data, as well as writing and revising manuscript.

Funding

This study was funded by personal funds. S.M.-L. is recipient of a PhD award from the Fonds de Recherche Québec-Santé. L.M. is recipient of a New Investigator Awards from the Canadian Institute of Health Research (CIHR). E.B. hold a Clinician Scientist Award from the CIHR and the Jeanne et Jean-Louis Lévesque Research Chair at Université Laval. A.B. is recipient of a PhD Award from CIHR.

Conflict of interest

None of the authors have conflicts of interest to declare.

References


Lim CP, Hasafa Z, Bhattacharya S, Maheshwari A. Should a hysterosalpingogram be a ﬁrst-line investigation to diagnose female tubal subfertility in the modern subfertility workup? Hum Reprod 2011; 26:967–971.


