Ahmed and Abo-taleb *Egyptian Journal of Radiology and Nuclear Medicine* (2019) 50:63
https://doi.org/10.1186/s43055-019-0064-1

Egyptian Journal of Radiology and Nuclear Medicine

RESEARCH Open Access

The validity of HSG in infertility work up

Shimaa Abdalla Ahmed[1*] and Hisham Abo-taleb[2]

Abstract

Background: Our purpose was to evaluate the diagnostic accuracy of hysterosalpingography (HSG) in the diagnosis of the uterine cavity and tuboperitoneal abnormalities in infertile women. Reproducibility and consistency were also assessed. Two hundred infertile females underwent HSG, hysteroscopy, and/or laparoscopy as part of infertility workup. HSG examinations were retrospectively reviewed by three radiologists; we compared interobserver variability, and differences between the two results of reading the same examination after 3 months were compared to calculate intraobserver variability.
HSG sensitivity, specificity, PPV, NPV, and accuracy were calculated.

Results: The overall accuracy of HSG in diagnosing tubal, uterine cavity, and peritoneal abnormalities was 95.5%, 95%, and 89%, respectively (*P* value < 0.04).
HSG is reproducible in diagnosing normal versus abnormal examinations. Reproducibility in diagnosing uterine cavity, tubal, and peritoneal abnormalities was (ICC = 0.90), (ICC = 0.70), and (ICC = 0.31), respectively. Best agreement was seen in diagnosing luminal filling defect (sub mucous fibroid/polyp) (ICC = 0.90) (95% CI 0.86–0.98), whereas poorest agreement was found in diagnosing uterine adhesions (ICC = 0.13) (95% CI 0.10–0.13) and pelvic adhesions (ICC = 0.12) (95% CI 0.10–0.13) (*P* value < 0.03).
HSG consistency ranged from moderate to good (*K* = 0.49–0.79). It was highest in diagnosing normal versus abnormal examination (*P* value < 0.01); poorest in diagnosing pelvic adhesions.

Conclusion: HSG has high validity in negative results; it can minimize the use of invasive procedures. Laparoscopy is recommended in patients who had a pelvic disease or showing tubal obstruction on HSG.

Keywords: Hysterosalpingogram, Hysteroscopy, Laparoscopy, Infertility, Interobserver variability

Background

Hysterosalpingography (HSG) is one of the most commonly used imaging modalities in infertility workup [1]. Uterine cavity abnormalities represent 50% of recurrent implantation failure [2]. Sensitivity and specificity of HSG in the detection of intrauterine abnormalities including (polypoid lesions, uterine malformations, and intrauterine adhesions) ranged from 44.4 to 75%, and from 82.5 to 96.4%, respectively [3].

The sensitivity and specificity of HSG in diagnosing tubal abnormalities was 65% and 83% respectively [4].

Tubal damage or uterine cavity abnormality detected by HSG as the cause of infertility will help the gynecologist to decide which operational techniques the patients will undergo (laparoscopy, hysteroscopy, or surgery) [5]. Although laparoscopy is superior to HSG in the evaluation of pelvic pathology and peritoneal factors of infertility, HSG is more economical and less invasive; both diagnostic methods are complementary [6].

HSG' comment or film reading is crucial for the infertility workup, and its interpretation will affect the next additional surgical attempts that will be needed; unfortunately, these interpretations may be affected by inter- and intraobserver variability in reading [5].

Poor to fair reproducibility of HSG among clinicians have been previously reported [5]; only two studies in the literature were designed for assessment of clinicians and radiologist's observer variability in HSG [1, 7]. Okaya et al. [1] reported more compatibility of radiologists than clinicians did. Renbaum et al. [7] described low reproducibility between non-radiology clinicians with generally good consistency.

* Correspondence: shimaaabdalla@aun.edu
[1]Radiodiagnosis, South Egypt Cancer Institute, Assiut University, el Azhar street, Assiut, Egypt
Full list of author information is available at the end of the article

While keeping in mind that high reproducibility is essential for a clinical test to achieve high diagnostic accuracy. To the best of our knowledge, HSG reproducibility exclusively among radiologists has not been studied and the question can be posed, whether observer bias is responsible for variation in observed prevalence of uterine and tubal evaluation in an infertile female. We conducted this study in our own setting to clarify accuracy, reproducibility, and consistency of HSG in diagnosing uterine cavity and tuboperitoneal abnormalities.

Methods

This retrospective study was conducted at a University Hospital Center (divisions of Reproductive Endocrinology and Radiology), and the ethics Committee of the Faculty of Medicine of our University approved this study and written informed consent was not required. All HSG examinations were performed by using Fujifilm computed radiography, USA.

We searched medical records on picture archive and communications system (PACS) over a 2-year period (between September 2015 and December 2017); we included 200 infertile females who attended the infertility clinic for treatment and underwent hysterosalpingograms (HSG), hysteroscopy, and/or laparoscopy as part of infertility workup. Incomplete hysteroscopy and/or laparoscopy reports were excluded.

The three radiologists who were specifically involved in HSG reading on a weekly basis evaluated all HSG examinations.

They had various levels of experience; consultant radiologist with specific women's imaging training for 7 years [1] represent level one, and women imaging fellows with 4 years' experience [1] represent level two, and 3rd-year radiology resident [1] represent level three.

Each reader interpreted ten HSG examinations in each session; the duration of each session was 30 min.

All radiologists were blinded to the patient's identity and diagnosis given by other readers. Each radiologist recorded his/her interpretations independently in a report sheet designed for the study.

We reported normal HSG when there was normal uterine cavity (average size and shape with no detected filling defect or irregularities), normal both tubes (average caliber, well outlined by free contrast spillage, and without peritoneal cavity contrast loculation).

Abnormal HSG was reported when there was a uterine cavity and/or tubal abnormality in the presence or absence of contrast media loculation in the peritoneal cavity or peritubal area.

We diagnosed abnormal tubes when there was a tubal obstruction (unilateral or bilateral) and/or hydrosalpinx in the presence or absence of pelvic/peritubal adhesions.

When tubal obstruction was detected at HSG, it was defined as proximal or distal. Proximal tubal obstruction was diagnosed if there was filling of the intramural or intramural/isthmic part of the tube without passage of contrast to the distal part.

Distal obstruction was diagnosed by absence of the contrast spillage to the peritoneal cavity after its passage to the distal part of the tube with or without ampullary dilatation. Pelvic adhesions were diagnosed if there was any contrast media loculation in the pelvic cavity.

We diagnosed peritubal adhesions (in whom patency of at least one tube was demonstrated) if there were at least two criteria from the following (convoluted tube, peritubal halo effect, contrast media loculation in the peritoneal cavity).

When abnormal uterine cavity (abnormal size and/or shape in the presence or absence of luminal filling defect/irregularities) was detected at HSG, it was further classified into congenital anomalies of uterine shape (septate uterus, bicornuate, and unicornuate, and others), luminal-filling defects (fibroid/polyp or adhesions), and abnormal uterine contour.

The results of first reading and results of second reading after 3 months were recorded to calculate inter- and intraobserver variability.

Finally, HSG diagnosis was compared with the gold standard (hysteroscopy and/or laparoscopy with dye test). All females in this study underwent HSG and hysteroscopy; additional diagnostic laparoscopy was performed in 128 patients only. The mean duration between HSG and laparoscopy was 2 ± 0.7 months, and between HSG and hysteroscopy was 1 ± 0.5 months.

Reference standard

Hysteroscopy and/or laparoscopy with a dye test.

Statistical analysis

IBM SPSS Statistics version 21 (IBM Corp., Armonk, NY) was used for data analysis.

True positive results were considered if HSG diagnosis is confirmed by hysteroscopy and/or laparoscopy; otherwise, false positive results were considered. True negative results were considered if no abnormality were detected by HSG which is confirmed by hysteroscopy and/or laparoscopy with dye test; otherwise, false negative results were considered. Sensitivity, specificity, positive, negative predictive values, and accuracy of HSG were calculated.

Interobserver and intraobserver agreement were tested for the presence of a uterine cavity or tuboperitoneal abnormalities and type of abnormality (as stated in the methodology section). Cohen's kappa coefficient was used for calculation of interobserver agreement. Kappa value of 0.81–1.00 indicates excellent agreement, a k

value of 0.61–0.80 indicates good agreement, a *k* value of 0.41–0.60 indicates moderate agreement, a *k* value of 0.21–0.40 indicates fair agreement, and a *k* value of < 0.20 indicates poor agreement [8]. Reproducibility and consistency in each HSG diagnosis were estimated using interclass correlation (ICC) [9]. *P* value ≤ 0.05 was considered as statistically significant.

Results

Out of 200 infertile women (their mean age was 30.36 ± 3.79 years), 72 women had normal HSG. Among these women, three cases had uterine cavity abnormalities, two cases had tubal obstruction, and five cases had peritoneal abnormalities on laparoscopy. The negative likehood ratio of HSG is 0.09 (95% CI 0.05–0.12).

One hundred twenty-eight women had abnormal HSG. Among these women, 17 cases showed normal study by gold standard procedures. The positive likehood ratio of HSG is 9.0 (95% CI 6.0–9.0) the number of cases detected by HSG and gold standard test was illustrated in (Fig. 1).

Thirty-three women had congenital abnormalities of uterine shape on HSG; among these women, seven cases were originally diagnosed as a bicornuate uterus, four had a complete uterine septum, two had a partial septum, and one had an arcuate uterus on hysteroscopy.

Forty-eight women had a uterine filling defect (submucous fibroid/polyp) on HSG; among these women, hysteroscopy confirmed normal cavity in seven cases. Thirty-eight women had uterine adhesions on HSG; among these women, hysteroscopy confirmed normal cavity in six cases. Thirty-one women had abnormal uterine contour on HSG; among these women, three cases showed normal cavity on hysteroscopy.

Seventy-three women had bilateral tubal patency on HSG; among these women, two cases had (one unilateral and one bilateral) tubal obstruction on laparoscopy.

Thirty-two women had unilateral tubal obstruction on HSG. Among these women, six had bilateral tubal patency, and four were diagnosed with bilateral proximal tubal obstruction on laparoscopy.

Thirty-three women had bilateral tubal obstruction on HSG. Among these women, two cases had unilateral proximal obstruction; three cases had bilateral patent tubes on laparoscopy.

Thirty-five women had pelvic adhesions on HSG; among these women, nine cases showed normal laparoscopic examination. Twenty-three women had peritubal adhesions on HSG; among these women, 11 cases showed normal laparoscopic examination accuracy of HSG was demonstrated in (Table 1).

Interobserver agreement between radiologists in the first and second round readings was demonstrated in (Table 2). We demonstrated general moderate reproducibility, which did not differ between the first and second readings. Agreement between observer 1 and the other observers ranged from fair to moderate (0.36 to 0.59). Agreement between the other two radiologists was moderate. We demonstrated the highest agreement percentage (90.5%) between Obs2 and Obs3 (*P* value < 0.05).

Interobserver variability in each HSG diagnosis was demonstrated in (Table 3). This table indicates good reproducibility in the assessment of normal versus abnormal HSG examination. Reproducibility in diagnosing uterine cavity, tubal, and peritoneal abnormalities was (ICC = 0.90), (ICC = 0.70), and (ICC = 0.31), respectively. Perfect agreement was seen in diagnosing sub mucous fibroid/polyp (ICC = 0.90) (95% CI 0.86–0.98) (Fig. 2) and in diagnosing abnormal uterine contour (ICC = 0.87) (95% CI (0.58–0.91) (Figs. 3 and 4). Poorest agreement was found in diagnosing uterine adhesions (ICC = 0.13) (95% CI 0.10–0.13) (Fig. 5) and pelvic adhesions (ICC = 0.12) (95% CI 0.10–0.13) (*P* value < 0.03).

Consistency between the first and second readings for each radiologist was demonstrated in (Table 4). HSG consistency ranged from moderate to good (*K* = 0.49–0.79). Agreement between first and second readings for observer 1, 2, and 3 was 0.79 (95% CI 0.64–0.79), 0.58 (95% CI 0.50–0.60), and 0.49 (95% CI 0.47–0.49), respectively. Agreement percentages were 81%, 91%, and 79%, respectively (*P* value < 0.01). We demonstrated that after the specified period, observer 1 achieved substantial agreement, while observers 2 and 3 maintained the moderate agreement. Highest consistency was seen in diagnosing normal versus abnormal examination and poorest in diagnosing pelvic adhesions (*P* value < 0.01).

We used a mixed model approach to evaluate the association between accuracy, reproducibility, and consistency in each HSG classified abnormalities (Fig. 6). It indicates the highest reliability in the evaluation of uterine cavity abnormalities and lowest reliability in the evaluation of peritoneal abnormalities.

Discussion

In the current study, the overall sensitivity and specificity of HSG in diagnosing uterine cavity abnormalities were 90.8% and 96.5%, respectively. Wide ranges of sensitivity (21% to 81%) and specificity (70% to 98%) have been reported in the literature [10–15] Taskin et al. [10] attributed their low sensitivity to the predominant malefactor of infertility in the couples visiting their clinic. Nigam et al. [11] reported a 70% PPV with 12.96% false negative rate. Shakya B [12] reported a 90% accuracy with 83.3% false negative rate, and false positive rate close to zero. These conflicting results may be related to the variable sample size and prevalence of each pathology.

Fig. 1 Demonstrates the number of cases detected by HSG and gold standard test in each diagnosis

We demonstrated the highest false negative rate of 2.5% (5/200) in diagnosing uterine luminal filling defect (submucous fibroid/polyps) confirmed by hysteroscopy; these data are in agreement with the data of Soares SR et al. [3]. In our opinion, early views of the uterus are essential to avoid missing small filling defects that may become obscured with advanced opacification of the uterine cavity. A combination of HSG and MR imaging can provide a road map for fibroid locations and to guide fertility-restoring therapy [16].

Table 1 Overall accuracy of HSG diagnosis compared to the gold standard test (hysteroscopy ± diagnostic laparoscopy with dye test)

Number of patients[a]	True positive	True negative	False positive	False negative	sensitivity	specificity	PPV	NPV	accuracy	*P* value
Uterine cavity (*n* = 200)										< 0.02
Congenital abnormalities of uterine shape[b]	29	163	7	1	96.9 (78–97)	95.9 (88–97)	80.6 (58–77)	99.4 (81–99.5)	96 (88–96)	
Luminal filling defect										
Submucosa fibroid/polyp	41	147	7	5	89.1 (88–91)	95.5 (89–96)	85.4 (82–89)	96.7 (88–99)	94 (81–94)	
Adhesion	32	158	6	4	88.9 (85–89)	96.3 (88–94)	84.2 (81–85)	97.5 (89–98)	95 (88–95)	
Abnormal uterine contour [c]	28	166	3	3	90.3 (81–91)	98.2 (86–99)	90.3 (85–92)	96.4 (89–97)	97 (91–97)	
Tubal status (*n* = 128)										< 0.04
Bilateral tubal patency	72	53	1	2	97.3 (89–99)	98.1 (93–99)	98.6 (88–99)	96.6 (81–97)	97.7 (90–98)	
Bilateral obstruction	30	93	3	2	93.8 (87–94)	96.9 (90–97)	90.9 (86–92)	97.9 (90–99)	96.1 (89–98)	
Unilateral obstruction	26	92	6	4	86.7 (81–90)	93.9 (87–95)	81.3.8 (80–90)	95.8 (88–98)	92.2 (89–98)	
Peritoneal abnormalities (*n* = 128)										< 0.01
Peritubal adhesions	12	101	11	4	75 (68–79)	90.2 (83–91)	52.2 (48–57)	96.2 (88–97)	88.3 (80–93	
Pelvic adhesions	26	90	9	3	89.7 (81–90)	90.9 (78–91)	74.3 (68–80)	96.8 (91–98)	90.6 (88–97)	

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy data are percentages. All numbers in parentheses are 95% CIs
[a]200 women underwent hysteroscopy, 128 of them underwent additional laparoscopy with dye test
[b]29 women was diagnosed to have congenital abnormalities of uterine shape on hysteroscopy (septate uterus in 15, bicornuate in 11, and unicornuate in 3)
[c]28 women was diagnosed to have abnormal uterine contour on hysteroscopy (adenomyosis in 11 and fibroid uterus in 17 patients)

Table 2 Agreement between the radiologists (interobserver variability) for the first and second readings (*P* value < 0.05)

Observers[a]	First round K value (95% confidence interval)	% agreement (n/n[a])	Second round K value (95% confidence interval)	% agreement (n/n[a])
Between observer 1,2	0.36 (0.31–0.59)	62.5 (125/200)	0.59 (0.57–0.60)	65 (130/200)
Between observer 1,3	0.38 (0.28–0.49)	71 (142/200)	0.50 (0.47–0.50)	75 (150/200)
Between observer 2,3	0.56 (0.47–0.59)	82.5 (165/200)	0.57 (0.55–0.60)	90.5 (181/200)

n/n[a]: number of agreed cases/total number of cases
[a]Observer 1 is the most experienced, and observer 3 is the least experienced

Interestingly, HSG has high PPV (98.2%) in diagnosing abnormal uterine contour in our study, similarly [12].

Further, 39.3% (11/28) patients with abnormal contour had adenomyosis and pelvic endometriosis on laparoscopy and underwent surgery. Seven cases showed deep pelvic endometriosis with adhesions/tethering between the torus uterinus and rectal serosa, which would be amenable to shaving. Five cases showed a large complex mass involving the posterior myometrial wall and infiltrating into the bowel wall, which would require segmental resection of the involved bowel. In our opinion, MRI is recommended in cases with abnormal uterine contour on HSG and specifically in radiographic appearance suggestive of adenomyosis (irregularity of the uterine contour with small outpouchings of contrast media). MRI can provide more information about disease burden before surgery and allows disease monitoring after the intervention [16].

Good correlation between HSG and hysteroscopy in diagnosing intrauterine adhesions has been previously reported [3, 13]. Fortunately, in our study, HSG has high specificity (96.3%) with a false positive rate of 3% (6/200). We attribute our false positive results to abnormality misclassification of adhesions induced deformed cavity. Two cases with unilateral scarring of one half of the uterine cavity diagnosed as having a unicornuate uterus on HSG and four cases with synechiae indenting the cavity diagnosed as septate uterus on HSG. Ahmadi et al. [17] reported that HSG appearance of intrauterine adhesions varies with the scar sites and severity [17]. HSG can provide a road map for the location and extent before hysteroscopic adhesiolysis, and interval HSG can be used to track the progression of the adhesions [16].

Highest sensitivity of HSG was demonstrated in diagnosing congenital uterine abnormalities. This finding corroborated the results of the previous study [18]. This finding is in disagreement with the data reported by Soares SR et al. [3] due to the high incidence of arcuate uteri in their study (which is a discrete malformation, more subtle and difficult to detect). Diagnostic accuracy of

Table 3 Reproducibility in each HSG diagnosis. Level of interobserver agreement expressed in ICC (as equivalent of the overall k). (*P* value < 0.03)

HSG diagnosis	ICC	95% confidence interval	Interobserver agreement (%)
Uterine cavity			
(normal vs. abnormal)	0.93	(0.77–0.94)	96%
Congenital anomalies of uterine shape	0.80	(0.80–0.95)	91%
Intraluminal filling defect			
Sub mucous fibroid/polyp	0.90	(0.88–0.91)	97%
Uterine adhesion	0.13	(0.10–0.13)	86%
Abnormal uterine contour	0.87	(0.58–0.91)	88%
Tubes			
(normal vs. abnormal)	0.81	(0.81–0.86)	96%
Bilateral patency	0.90	(0.83–0.90)	91%
unilateral obstruction	0.40	(0.39–0.42)	80%
bilateral obstruction	0.69	(0.65–0.80)	90%
Peritoneal abnormalities			
No peritoneal abnormalities	0.54	(0.33–0.56)	74%
Presence of Pelvic adhesions	0.12	(0.10–0.13)	76%
Presence of peritubal adhesions	0.29	(0.21–0.35)	81%

CI confidence interval, *ICC* interclass correlation coefficient

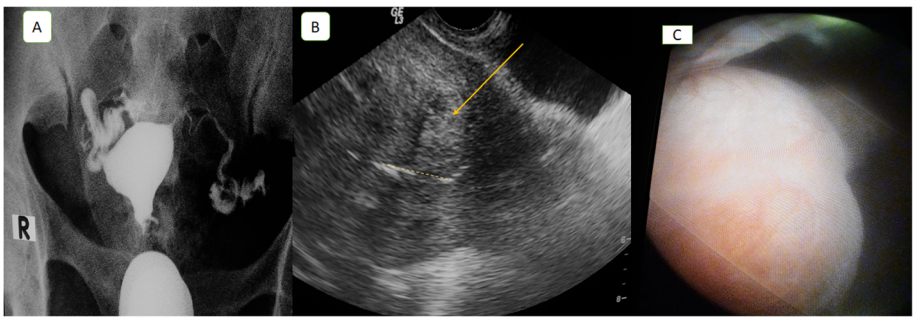

Fig. 2 Female aged 29 years old, G1 P1 A0, with multiple fibroids. Hysterosalpingography first film (**a**) showed an enlarged size of the uterine cavity with distorted outlines with an immediate spill of contrast from both tubes, all radiologists agreed to the possible diagnosis of multiple fibroids (**b**). **b** TVUS image showed enlarged uterus with multiple hypoechoic mural and submucous fibroid, largest one measuring 4 × 3 cm (arrow in **b**), it displaces the endometrial line (dashed arrow in **b**). **c** The hysteroscopic image confirmed fibroids

HSG seems to depend on the type of uterine abnormality; HSG is more sensitive in diagnosing uterine malformations with a more aggressive morphologic expression [3].

As expected, differentiation between septate and bicornuate uterus is a common diagnostic dilemma in HSG, similarly [18]. The present study estimated that seven cases diagnosed as having bicornuate uterus at HSG had a complete uterine septum (four cases), partial septum (two cases), and arcuate uterus (one case) at hysteroscopy.

It is known that the appearance of the external fundal contour represents the primary difference (septate uterus will show a normal convex uterine fundus). In most cases, further evaluation with MRI is required; the septate uterus can be discriminated from the bicornuate uterus by the presence of fundal cleft less than 1 cm [16]. In addition, it allows characterization of the septal nature (fibrous or muscular). Hysteroscopic metroplasty is recommended for the fibrous component of the septum without resection of the muscular part due to the high risk of bleeding [16].

It is our belief that MRI is recommended for evaluation of double uterine cavity abnormalities on HSG. It provides accurate characterization, which is essential for treatment decision and prediction of pregnancy outcomes.

The current study demonstrated that HSG is accurate (97.7%) in predicting bilateral tubal patency with a false negative rate of 1.6% (2/128). One case had unilateral tubal obstruction at laparoscopy; we attribute our false negative results to pelvic diffusion of contrast media from the only patent tube, which was misinterpreted as bilateral tubal patency. The other case had bilateral tubal obstruction at laparoscopy; contrast intravasation into uterine and ovarian veins during HSG was mistaken for

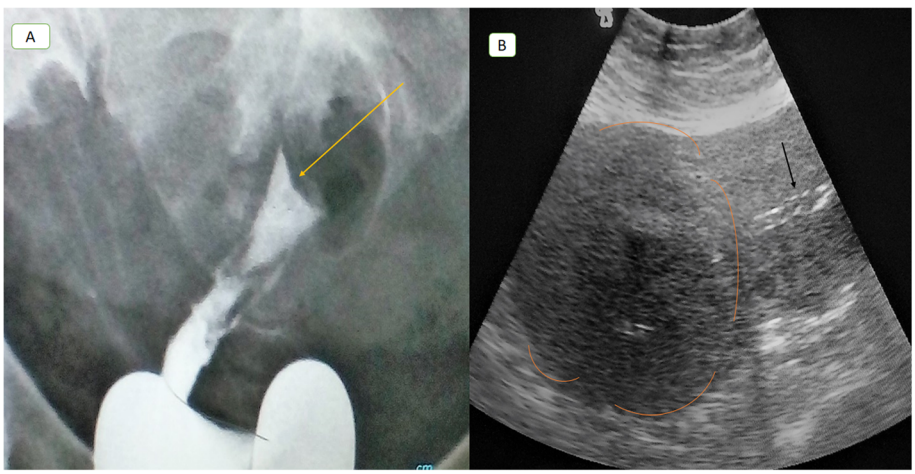

Fig. 3 Female aged 31 years old, nullipara, with fundal fibroid, hysterosalpingography first film (**a**) showed abnormal uterine contour (displaced uterine cavity downward and to the right with filling defect at the fundus) (arrow in **a**) with non-visualized both tubes and no immediate spill of contrast. All radiologist agreed to the diagnosis of fundal fibroid with bilateral tubal block, an abdominal ultrasound image of the same patient illustrating bulky uterus with a large hypoechoic fundal fibroid (circular lines in **b**); it displaces the endometrial line downward (arrow in **b**)

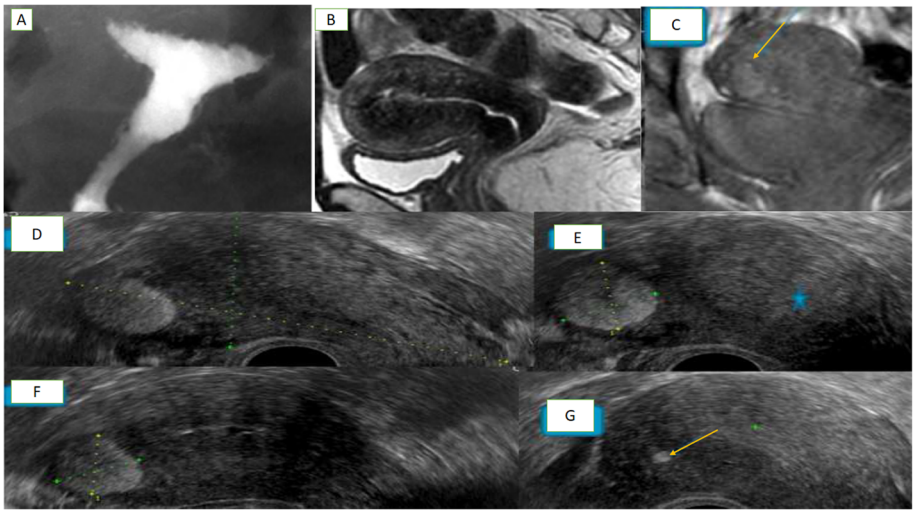

Fig. 4 A 44-year-old female with diffuse and focal adenomyosis. **a** HSG shows irregularity of the uterine contour with small outpouchings of contrast material, findings that represent diffuse adenomyosis. **b** Sagittal T2-weighted MR image shows thickening of the junctional zone to more than 1 cm with numerous foci of a hyperintense signal. Coronal T1 (**c**) showed focal adenomyoma (arrow in **c**), rounded lesions with a hyperintense signal not suppressed in fat suppression sequence (not shown). Sagittal US (**d**–**g**) showing bulky uterus with an ill-defined endomyometrial junction, heterogonous myometrium with echogenic linear striation (star in **e**), with a hyperechoic rounded fundal lesion (nicely corporates focal adenomyoma seen at MRI), and another similar smaller one in the anterior myometrium (arrow in **g**)

tubal filling. In our opinion, fluoroscopic observation of dynamic filling and spillage can reduce such diagnostic errors.

Differentiation of proximal tubal occlusion from transient spasm may be challenging in some cases. In the current study, HSG falsely diagnosed bilateral proximal tubal occlusion in three cases. In our opinion, if a proximal tubal occlusion is suspected at HSG, tubal spasm should be considered as the possible cause (if there are tapering and smooth cornual margin). A spasmolytic agent should be administered to relieve tubal spasm then we can inject additional 1–2 ml of contrast media until the tube fills and spills if not; proximal tubal obstruction is suggested (cornual margin is pointed or blunt and irregular), at that time, selective trans-cervical tubal recanalization under fluoroscopic guidance could be recommended.

The reported accuracy differed between studies concerning tubal obstruction (level and side of obstruction) [19–22]. It was more accurate in distal than proximal

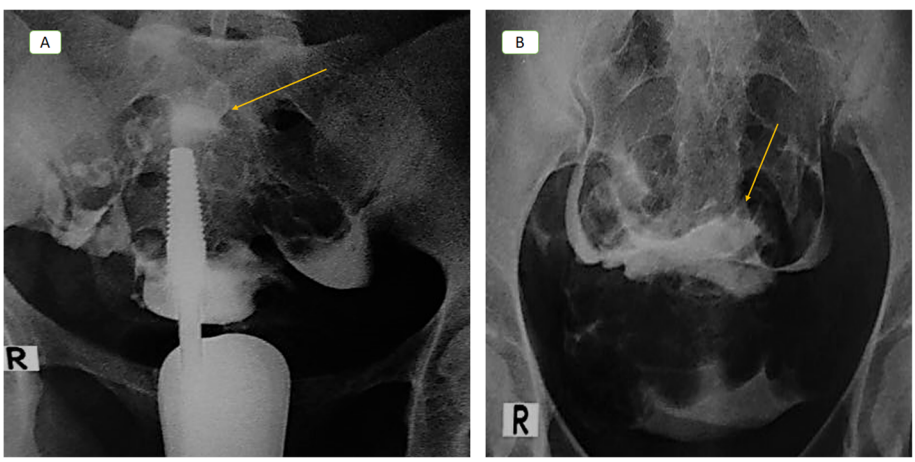

Fig. 5 Female aged 26 years old, G3 P0 A3 with uterine and pelvic adhesions in hysterosalpingography. **a** The first film showed a relatively reduced size of the uterine cavity with haziness of its outlines (arrow in a) and immediate spill of contrast from both tubes. The second film (**b**) showed loculation of the contrast in the central part of the pelvis (arrow in **b**); only expert radiologist correctly diagnosed the case, blindness to patient clinical data of previous three dilations and curettage for miscarriage may contribute to the other two radiologist misdiagnosis; results were confirmed by laparoscopy

Table 4 Consistency (intraobserver variability) between the two readings for each radiologist. Level of intraobserver agreement expressed in ICC (as equivalent of the overall k) (*P* value < 0.01)

HSG diagnosis	ICC	95% confidence interval	Intraobserver agreement (%)
Uterine cavity			
(normal vs. abnormal)	0.95	(0.87–0.96)	97%
congenital anomalies of uterine shape	0.81	(0.71–0.96)	92%
Intraluminal filling defect			
Sub mucous fibroid/polyp	0.92	(0.85–0.95)	93%
Uterine adhesion	0.25	(0.10–0.35)	80%
Abnormal uterine contour	0.59	(0.55–0.59)	89%
Tubes			
(normal vs. abnormal)	0.93	(0.87–0.96)	95%
Bilateral tubal patency	0.87	(0.71–0.87)	89%
Unilateral obstruction	0.66	(0.65–0.82)	81%
bilateral obstruction	0.74	(0.65–0.80)	92%
Peritoneal abnormalities			
No peritoneal abnormalities	0.61	(0.60–0.68)	88%
Pelvic adhesions	0.18	(0.10–0.20)	76%
Peritubal adhesions	0.23	(0.20–041)	81%

tubal obstruction [19] and in bilateral than unilateral tubal obstruction [22].

Tvarijonavičienė E and coworkers [22] reported that the choice of laparoscopy as a "gold standard" procedure in diagnosing tubal occlusion is questionable (the 3-year cumulative pregnancy rate in patients diagnosed to have bilateral tubal occlusion at LS was 2%). It indicates that laparoscopy is not real reference standard, but it is the best we have. Waheed S et al. [6] reported that HSG is significantly superior in predicting tubal patency (68% of patients had bilateral tubal patency on HSG, compared to 28% on laparoscopy). B. Berker et al. [21] recommended laparoscopy in patients with tubal obstruction (unilateral or bilateral) on HSG; they reported a change in the management plan from artificial reproductive technologies (ART) to intrauterine insemination in 12 patients with bilateral tubal obstruction on HSG but confirmed tubal patency by laparoscopy. Recently, fertiloscopy was analyzed as a procedure of choice for tubal status evaluation [23], but further studies are needed to assess the accuracy of this procedure.

Interestingly, HSG is specific (90.2%) in diagnosing peritubal adhesions in our study by using more than one of the reported radiographic signs with a false positive rate of 8.6% (11/128). Valentini et al. [24] reported accuracy (89.2%) with false positive rate 11.7% and false negative rate 9%. They recommended shortening of the 6-month interval between HSG and laparoscopy in cases

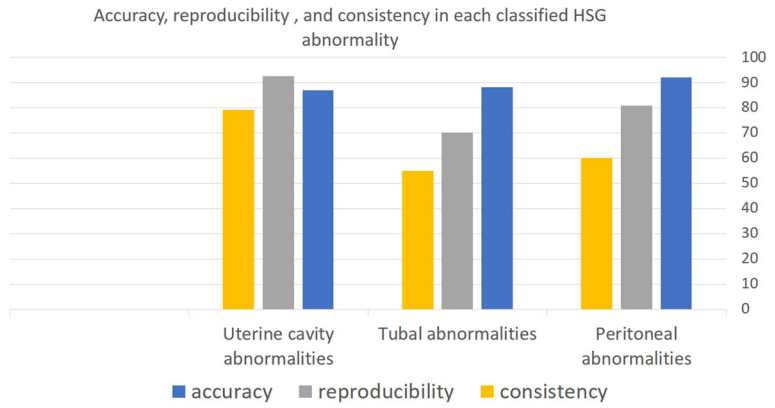

Fig. 6 Illustrates the association between accuracy, reproducibility, and consistency in each HSG classified abnormalities

with a radiographic diagnosis of peritubal adhesions; at least radiographic suspicion can simplify the laparoscopic procedure and obviating diagnostic laparoscopy before surgical intervention. We recommend HSG before surgical intervention in peritubal adhesions.

Our results are in line with previously published findings [24], which emphasize the good accuracy of HSG in the diagnosis of pelvic adhesions. Lower accuracy has been reported [6, 25–27]. Goynumer G et al. [25] described high PPV of HSG in cases with severe pelvic disease; however, due to its low NPV, suspicious and even normal HSG should undergo diagnostic laparoscopy [25]. Waheed S et al. [6] reported that laparoscopy is more accurate than HSG in diagnosing pelvic adhesions; both laparoscopy and HSG are complementary. S Tanahatoe et al. [28] concluded that laparoscopy may reveal normal findings or abnormalities not requiring ART in patients with abnormal HSG even with bilateral pathology.

The American Society for Reproductive Medicine (ASRM) states that HSG is the standard test for tubal patency. Laparoscopy is indicated for women with risk factors for peritoneal disease (pelvic pain, moderate or severe endometriosis, previous pelvic infection, or surgery) or an abnormal HSG or US who do not require assisted reproductive technique, e.g., severe male factor infertility. Laparoscopy is not recommended for routine infertility work up in the absence of pelvic pathology or another specific indication (i.e., severe dysmenorrhea) [29].

Based on our study and results of previous literature [6, 20, 21, 30, 31], we can conclude that HSG is a valuable imaging modality for evaluating the genital tract of an infertile female; invasive procedures like laparoscopy and hysteroscopy are not indicated in patients with normal HSG findings and patients may be managed conservatively. Patients who had a pelvic disease or showing tubal obstruction on HSG needs further evaluation by laparoscopy.

We should emphasize that HSG performance could be influenced by the faulty technique and artifacts (e.g., inadvertent insertion of the cannula, insufficient uterine pressure due to vaginal reflux, premature ending of the procedure). In addition, there is an unavoidable possibility of objectiveness involved in the interpretation of results. Minimization of misdiagnosis can be achieved by understanding technique limitations, being thoughtful and highly consistent in interpretations (we observed that consultant and fellow radiologists are more consistent than resident, which may be attributed to their better visual memory to remember cases rather than increased learning curve in between reading rounds). Finally, HSG should be interpreted in association with the findings of clinical examination and symptomatology, as well as other imaging modalities (e.g., transvaginal echography).

Strengths of our study include using uniform radiographic signs among three radiologists with different experience and comparison to the gold standard test. We included suboptimal HSG examinations to decrease concordance between readers and to represent day-to-day practice.

Limitation of our study includes its retrospective nature; the radiologist was given HSG snapshot series and did not look at the real-time HSG while it was being done to be able to observe dynamic filling and spillage. Although, no limitation to either group, which will not affect inter-reader variability. Now, most clinicians prefer not to be present at HSG time and read post hoc films later to make clinical decisions. The second reading in our study might be affected by training effect and detection bias as a consequence of the first reading; to overcome this, we ensured a gap of at least 3 months between the two readings. The duration between HSG and hysteroscopy or laparoscopy may possibly contribute to bias. In the future, we encourage future research to develop a guideline with exact definitions of what should be judged as HSG.

Conclusion

HSG has high validity in negative results; it can minimize the use of invasive procedures. Laparoscopy is recommended in patients who had a pelvic disease or showing tubal obstruction on HSG.

Acknowledgements

Not applicable.

Authors' contributions

Both authors (SA and HA) contributed equally in data analysis, methodology, and study design. SA was a major contributor in writing the manuscript. Both authors approved the final manuscript.

Funding

No disclosure of funding received for this work.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Assiut Medical Ethical Review Board (NO. 17100340). The need for consent was waived by an IRB of our university (Assiut university hospital) due to retrospective nature of the study.

Consent for publication

All patients included in this research gave written informed consent to publish the data contained within this study. If the patient was less than 16 years old, deceased, or unconscious when consent for publication was requested, written informed consent for the publication of this data was given by their parent or legal guardian.

Competing interests

The authors declare that they have no competing interests.

Author details
[1]Radiodiagnosis, South Egypt Cancer Institute, Assiut University, el Azhar street, Assiut, Egypt. [2]Obstetrics and Gynecology, Women Health Hospital, Assiut University, Assiut, Egypt.

Received: 13 August 2019 Accepted: 10 October 2019
Published online: 29 November 2019

References

1. Guler A, Turan A, Gökalp Ö, Neşe Ç, Çeliker FB (2016) Inter-observer variability of radiologists and gynecologists in hysterosalpingogram evaluation. J Turkiye Klinikleri Jinekoloji Obstetrik 26(1)
2. Brown SE, Coddington CC, Schnorr J, Toner JP, Gibbons W, Oehninger S (2000) Evaluation of outpatient hysteroscopy, saline infusion Hyster sonography, and hysterosalpingography in infertile women: a prospective, randomized study. Fertil Steril 74(5):1029–1034
3. Soares SR, dos MM R, Camargos AF (2000) Diagnostic accuracy of sonohysterography, transvaginal sonography, and hysterosalpingography in patients with uterine cavity diseases. Fertil Steril 73:406–411
4. Panchal S, Nagori C (2014) Imaging techniques for assessment of tubal status. J Hum Reprod Sci 7(1):2–12
5. Glastein IZ, Sleeper LA, Lavy Y, Simon A, Adoni A, Palti Z et al (1997) Observer variability in the diagnosis and management of the hysterosalpingogram. Fertil Steril 67(2):233–237
6. Waheed, S., Mazhar, R., Khan, N. H., & Rafi, M. The Comparison of Hysterosalpingography and laparoscopy in predicting fertility. Annals of King Edward Medical University, 13(3), 202
7. Renbaum L, Ufberg D, Sammel M, Zhou L, Jabara S, Barnhart K (2002) Reliability of clinicians versus radiologists for detecting abnormalities on hysterosalpingogram films. Fertil Steril 78(3):614–618
8. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
9. Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 33:613–619
10. Taskin EA, Berker B, Ozmen B, Sönmezer M, Atabekoğlu C (2011) Comparison of hysterosalpingography and hysteroscopy in the evaluation of the uterine cavity in patients undergoing assisted reproductive techniques. Fertil Steril 96:349–352
11. Nigam A, Saxena P, Mishra A (2015) Comparison of hysterosalpingography and combined laparohysteroscopy for the evaluation of primary infertility. Kathmandu Univ Med J (KUMJ) 13:281–285
12. Shakya B (2009) Hysterosalpingography vs hysteroscopy in the detection of intrauterine pathology in infertility. J Nepal Health Res Counc 7(1):6–9
13. Clever O, Esther A, Ebuka O, Anthony M (2015) Relative assessment of abnormalities patterns in hysterosalpingography, diagnostic laparoscopy and hysteroscopy with infertility cases in women in Nigeria. Int J Obstet Gynecol 3:081–089
14. Vahdat M, Sariri E, Kashanian M, Najmi Z, Mobasseri A, Marashi M et al (2016) Can combination of hysterosalpingography and ultrasound replace hysteroscopy in diagnosis of uterine malformations in infertile women? Med J Islam Repub Iran 30:352
15. Acholonu UC Jr, Silberzweig J, Stein DE, Keltz M (2011) Hysterosalpingography versus sonohysterography for intrauterine abnormalities. JSLS 15:471–474
16. Vickramarajah S, Stewart V, van Ree K, Hemingway AP, Crofton ME, Bharwani N (2017) Subfertility: what the radiologist needs to know. Radiographics 37(5):1587–1602
17. Ahmadi F, Siahbazi S, Akhbari F, Eslami B, Vosough A (2013) Hysterosalpingography finding in intra uterine adhesion (asherman's syndrome): a pictorial essay. Int J Fertil Steril 7(3):155–160
18. Chauhan MB, Lakra P, Nanda S, Malik R, Malhotra V (2013) Hysterosalpingography vs hysteroscopy: role in assessment of uterine factor during infertility workup. J S Asian Fed Obstet Gynecol 5:116–119
19. Farideh G, Fereshteh R (2017) Comparison of hysterosalpingography and laparoscopy in evaluation of female infertility. Med Res Arch 5:6 [S.l.], ISSN 2375–1924
20. Lavy Y, Lev-Sagie A, Holtzer H, Revel A, Hurwitz A (2004) Should laparoscopy be a mandatory component of the infertility evaluation in infertile women with normal hysterosalpingogram or suspected unilateral distal tubal pathology? Eur J Obstet Gynecol Reprod Biol 114(1):64–68
21. Berker B, Şükür YE, Aytaç R, Atabekoğlu CS, Sönmezer M, Özmen B (2015) Infertility workup: to what degree does laparoscopy change the management strategy based on hysterosalpingography findings? J Obstet Gynecol Res 41(11):1785–1790
22. Tvarijonavičienė E, Nadišauskienė RJ (2008) The value of hysterosalpingography in the diagnosis of tubal pathology among infertile patients. Medicina (Kaunas) 44(6):439–448
23. Watrelot A, Nisolle M, Chelli H, Hocke C, Rongieres C, Racinet C (2003) Is laparoscopy still the gold standard in infertility assessment? A comparison of fertiloscopy *versus* laparoscopy in infertility. Results of an international multicentre prospective trial: the 'FLY' (Fertiloscopy-LaparoscopY) study. Hum Reprod 18(4):834–839
24. Valentini AL, Muzii L, Marana R, Catalano GF, Summaria V, Felici F et al (2000) Improvement of Hysterosalpingography accuracy in the diagnosis of peritubal adhesions. AJR Am J Roentgenol 175(4):1173–1176
25. Goynumer G, Yetim G, Gokcen O, Karaaslan I, Wetherilt L, Durukan B (2008) Hysterosalpingography, laparoscopy or both in the diagnosis of tubal disease in infertility. WJOLS 1(2):23–26
26. Foroozanfard F, Sadat Z (2013) Diagnostic value of hysterosalpingography and laparoscopy for tubal patency in infertile women. Nurs Midwifery Stud 2(2):188–192
27. Ngowa JDK, Kasia JM, Georges NG-T, Nkongo V, Sone C, Fongang E (2015) Comparison of hysterosalpingograms with laparoscopy in the diagnostic of tubal factor of female infertility at the Yaoundé General Hospital, Cameroon. Pan Afr Med J 22:264. https://doi.org/10.11604/pamj.2015.22.264.8028
28. Tanahatoe S, Lambalk C, McDonnell J, Dekker J, Mijatovic V, Hompes P (2008) Diagnostic laparoscopy is needed after abnormal hysterosalpingography to prevent over-treatment with IVF. Reprod BioMed Online 16(3):410–415
29. Practice Committee of the American Society for Reproductive Medicine (2015) Diagnostic evaluation of the infertile female: a committee opinion. Fertil Steril 103(6):e44–e50
30. Perquin DA, Dorr PJ, de Craen AJ, Helmerhorst FM (2006) Routine use of hysterosalpingography prior to laparoscopy in the fertility workup: a multicentre randomized controlled trial. Hum Reprod 21:1227–1231
31. Fatum M, Laufer N, Simon A (2002) Should diagnostic laparoscopy be performed after normal hysterosalpingography in treating infertility suspected to be of unknown origin? Hum Reprod 17:1–3

Publisher's Note