European Journal of Obstetrics & Gynecology and Reproductive Biology 143 (2009) 116–120

Contents lists available at ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology

journal homepage: www.elsevier.com/locate/ejogrb

A score-based method to improve the quality of emergency gynaecological ultrasound examination

L.J. Salomon [a,*], M. Nassar [b], J.P. Bernard [a], Y. Ville [a], A. Fauconnier [b]

Société Française pour l'Amélioration des Pratiques Echographiques (SFAPE)

[a] Service de Gynécologie - Obstétrique, Hôpital Necker-Enfants Malades, AP-HP. Faculté de Médecine, Université Paris Descartes, Paris. France
[b] Service de Gynécologie - Obstétrique. Centre Hospitalier Intercommunal de Poissy-St Germain, 10 rue Champ Gaillard, Poissy, France

ARTICLE INFO

Article history:
Received 23 June 2008
Received in revised form 30 September 2008
Accepted 22 December 2008

Keywords:
Ultrasound
Emergency
Quality
Gynaecology
Resident

ABSTRACT

Objectives: The quality of ultrasound (US) examination in the setting of a gynaecological emergency consultation is variable. Our aims were to develop and evaluate the feasibility of an image-based scoring method for the evaluation of standardised gynaecological images in this context.
Study design: Ultrasound images of the uterus (2 planes), ovaries (1 plane of each side) and Morrison's pouch were obtained by specialist registrars as part of their assessment of gynaecological emergencies during a three-month period. Twenty images of each of these 5 planes were randomly selected and anonymised. These 100 images were analysed by two reviewers. Each image was scored according to 23 criteria yielding a maximum score of 23 points for the entire ultrasound examination. Training was then offered with special emphasis on quality criteria. Following training, another set of 100 images obtained by the same specialist registrars was scored. Scores before and after training were compared. Inter- and intra-reviewer reproducibility were analysed using intra-class correlation, adjusted Kappa and Bland–Altman plot.
Results: The mean (±SD) scores were 10.22(±2.75) and 16.1(±3.35) before and after theoretical training respectively ($p < 10^{-4}$). The intra-class correlation coefficient and the mean difference in score were 0.973 [0.957;0.990] and 0.02 [−1.98;1.94], and 0.952 [0.894;1] and −0.1 [−2.44;2.24] for inter- and intra-reviewer reproducibility respectively. Kappa values were above 0.8 for all but 3 criteria.
Conclusion: A quality control policy for gynaecological emergencies based on image scoring is feasible and allows for good inter- and intra-reviewer reproducibility. A policy of this nature is likely to improve the quality of emergency gynaecological ultrasound examination.
© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The use of ultrasound technology in obstetric emergencies is well-established [1]. Transvaginal sonography (TVS) has an established role in the detection of pathologies responsible for gynaecological emergencies [2–4]. It is also of great value in diagnosing ectopic pregnancies (EP) [5,6] and in the management of early pregnancy complications [7]. Use of sonography in an emergency gynaecology unit may increase the rate of pathology detected [8]. Furthermore it has been suggested recently that the availability of transvaginal sonography at the time of initial assessment of patients could improve diagnostic accuracy and reduces unnecessary admissions and follow-up examinations [7]. However, the scanning skills of the initial sonographer may be highly variable [9] and the ultrasound examination protocol for gynaecological emergencies is not standardised yet. There is, therefore, an emerging role for training and quality assurance programmes in gynaecological ultrasound examination [10]. Image scoring systems have been used for quality control of nuchal translucency and biometric measurements at the first and second trimester ultrasound examinations respectively [11,12].

The aims of this study were (i) to propose criteria for standardised assessment of ultrasound images in the context of gynaecological emergencies, (ii) to evaluate the feasibility of an image-based scoring method and (iii) to evaluate the impact of training and quality control on the quality of the examinations. Analysis of the outcome in relation to the quality of the examinations will be examined in future studies.

2. Material and methods

2.1. Settings

This study was undertaken in a teaching hospital with a dedicated area for assessing gynaecological emergencies. Every patient was managed by a specialist registrar with 2–4 years

* Corresponding author. Tel.: +33 1 44 49 40 30.
E-mail address: laurentsalomon@gmail.com (L.J. Salomon).

0301-2115/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ejogrb.2008.12.003

Table 1
Criteria for score-based objective evaluation.

Criteria	Sagittal view of the uterus	Transversal view of the uterus	View of the ovary (one image for each side)	Morrison pouch view
1	Uterine cervix visible	Endometrial midline echo is horizontal	Side stated (left or right)	Liver visible
2	Uterine fundus visible	One interstitial tube visible	Follicle(s) visible	Kidney visible
3	Endometrial midline echo visible	The other interstitial tube visible	Iliac vein visible	Ovoid section of the kidney
4	Endocervix visible	Visible space below the posterior uterine wall.	Long axis of the ovary $<30^{\circ}$ with the horizontal line	
5	Uterus occupying more than half of the total image size	Uterus occupying more than half of the total image size	Ovary occupying more than a quarter of the total image size	

experience in gynaecological emergencies. All patients underwent general physical and pelvic examinations. It is our policy to systematically perform a gynaecological ultrasound examination and to record images in cases with pelvic pain and/or bleeding. All patients gave informed consent for the use of ultrasound data in subsequent studies. There was no institutional review board (IRB) approval since this study did not modify routine emergency care and it was not required for this type of study in our institution.

2.2. Development of the standardised images and scoring system

To develop a scoring method, we used previous studies [7,9,13,14] and performed a comprehensive systematic review with a Medline search using the following MESH terms: emergencies, female, ovarian cysts, ovarian diseases, torsion, ultrasonography, pelvic inflammatory disease, salpingitis, pregnancy, ectopic pregnancy, haemoperitoneum, genital disease, pelvic pain. Only studies where the diagnostic values of the signs (i.e. sensitivity and/or specificity) were reported were included. Reviews and case reports were not included.

Five standardised images were considered mandatory, even in the absence of abnormalities. Four of them are to be obtained preferably with a transvaginal probe: (i) sagittal view of the uterus, (ii) transversal view of the uterus, (iii) and (iv) view of the ovary (one image for each side). The fifth is best obtained with the transabdominal probe: (v) view of Morrison's pouch. Three or more optional images to describe specific abnormalities found at examination can be added to the standard set: one for intra-uterine abnormality, one for extra-uterine abnormality, one for any other abnormal finding. These optional images were not evaluated in this study.

Between September 2006 and March 2007, we reviewed all consecutive cases of patients referred for pain and/or vaginal bleeding in the emergency unit of the gynaecological department. The sonographic examination was performed by residents in all cases on patients with an empty bladder. Both transabdominal and transvaginal examination were performed and images were recorded on paper in the medical file for each patient. During the first study period (September–November 2006), residents were unaware of the subsequent study at the time the images were recorded and stored. Registrars were asked to evaluate and take images of Morrison's pouch, the uterus in different planes as well as ovaries. There was no restriction on the number of images. In December 2006, theoretical instruction was given to all residents in the department. Particular emphasis was placed on the quality of ultrasound images and on the criteria images should meet. Theoretical instruction included theoretical and practical courses with image reviews. During the second study period (January–March 2007), residents were aware of this ongoing quality control programme and each was given a written protocol including the imaging requirements and the score.

All examinations were performed with no time constraint using the same probe and ultrasound machine [3.5–5 MHz curvilinear abdominal transducer and 7 MHz vaginal transducer- General Electric Voluson 730 Expert- GE Medical System Europe-78 Buc France], with cineloop facility.

Objective scoring was performed according to the pre-defined criteria summarized in Table 1. These criteria were specific for each type of image and were agreed upon by expert authors (LJS, JPB, YV, AF) on the basis of established standards and a comprehensive literature review. The optional images were not scored and evaluated in this preliminary study. Each correct criterion scored one point yielding a maximum score of 23 points for the entire ultrasound examination. Fig. 1 shows ultrasound images meeting our quality criteria. Besides these five basic images (including one image for each ovary), residents were asked to take any necessary additional images to illustrate any abnormal finding.

A preliminary study was performed on ten gynaecological ultrasound examinations. The observed score (mean $\pm$ SD) was 10.5 ± 2.6. Based on this preliminary result, we calculated that we would require at least two groups of 18 complete examinations to pick up a 2.5 point increase in the score following training.

2.3. Evaluation protocol

A total of 100 images consisting of 20 images for each of the standardised ultrasound planes (20 complete examinations) were selected from the first study period (September–November 2006) using random selection among cases of patients referred for pain and/or vaginal bleeding in the emergency unit of the gynaecological department.

Another set of 100 images consisting of 20 images for each of the standardised ultrasound planes (20 complete examinations) was randomly selected during the second study period (January–March 2007).

The 200 images were then projected on a wide screen during one session, to two different reviewers (A and B) for individual and independent evaluation using our objective scoring method. To assess intra-rater variability, 10 complete examinations were scored twice by one of the reviewers.

We aimed to examine improvement in the "process of care" only, and did not assess whether any perceived change has resulted in any improvement in patient care. This will be prospectively evaluated in another study.

2.4. Statistical analysis

Normality of the distribution of scores was tested using Shapiro–Wilk's *W*-test. The difference in mean scores attributed by each reviewer before and after training was estimated using a Student-*t*-test. Inter-rater variability was tested as follows: the mean scores attributed by each reviewer were compared using a paired *t*-test. Adjusted kappa coefficients [15] for each individual criterion were computed in order to test for the reproducibility of each independent criterion. The intra-class correlation of scores given by the two reviewers was calculated and any difference in scoring was assessed using the Bland–Altman method and plot [16]. Intra-reviewer variability was assessed using the same methods.

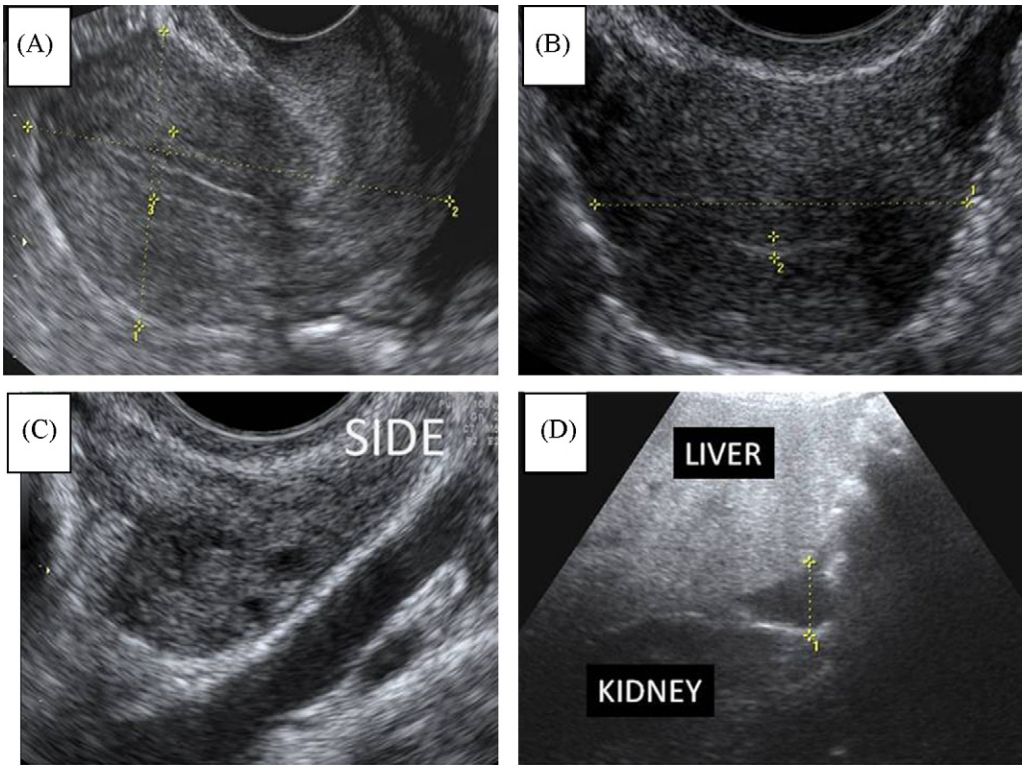

Fig. 1. Ultrasound images meeting our quality criteria. (A, sagittal view of the uterus; B, transversal view; C, ovary; D, Morrison's pouch).

Statistical analyses were performed using Statistica 6.0 (StatSoft, 0K 74104 USA) and Excel 2000 (Microsoft Seattle, USA). For all tests, a value of $p < 0.05$ was considered statistically significant. Adjusted kappa values below 0.6, between 0.6 and 0.8, and above 0.8 were taken as representing poor, moderate and good agreement respectively [17].

3. Results

The mean (±SD) scores were 10.22 (±2.75) and 16.1 (±3.35) before and after theoretical training respectively. There was a significant increase in scores following training ($p < 10^{-4}$) as illustrated in Fig. 2.

3.1. Inter-reviewer variability

There was no difference in mean scores attributed by each reviewer (13.15 ± 4.4 and 13.17 ± 4.1 for reviewer A and B respectively, $p > 0.05$). Adjusted kappa values between reviewers corresponding to each individual criterion were calculated for the 23 criteria assessed by the two reviewers. Values were below 0.6, between 0.6 and 0.8 and above 0.8 for 0, 3 and 20 criteria respectively. All results are shown in Table 2.

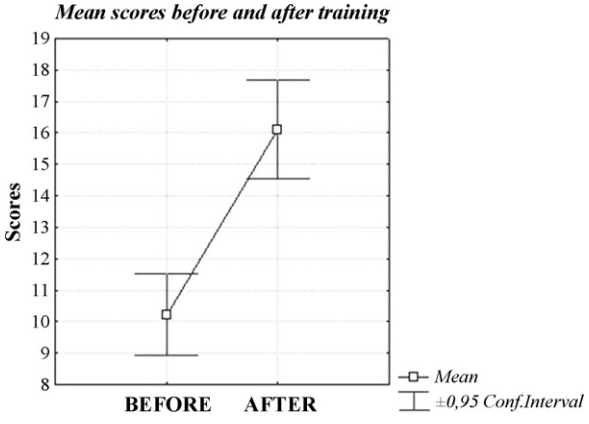

Fig. 2. Mean scores before and after training. There was a significant increase in scores ($p < 10^{-4}$).

The intra-class correlation coefficient was 0.973 [0.957;0.990] and the mean difference in score was 0.02 [−1.98;1.94]. The stability of this mean difference is illustrated in the Bland and Altman plot (Fig. 3).

3.2. Intra-reviewer variability

There was no difference in mean scores attributed by the same reviewer during the two scoring sessions (16.6 ± 3.7 and 16.7 ± 3.7 for first and second scoring respectively, $p > 0.05$). Adjusted kappa values within reviewer corresponding to each individual criterion were above 0.8 for all criteria. The intra-class correlation coefficient was 0.952 [0.894;1] and the mean difference in score was −0.1 [−2.44;2.24].

4. Discussion

We propose here a standardisation of the gynaecological emergency ultrasound examination that makes scoring and quality control feasible. Implementation of this system in our centre significantly improved the quality level of ultrasound examination in the gynaecological department.

To our knowledge this is the first study reporting a method for assessing the quality of emergency gynaecological ultrasound examination. The method is based on the use of standardised images designed to achieve the following goals: (i) to provide a selection of definite informative images that will summarize the dynamic process of the examination; (ii) the selected images should provide an ultrasound-based evaluation of the patient's pelvic anatomy and any anomalies encountered and (iii) the proposed set of images should meet most described ultrasound

Table 2
Kappa coefficients for each individual criteria for inter- and intra-rater reliability.

Criteria	κ	Sagittal view of the uterus	Transversal view of the uterus	Ovarian view (right/left)	Morrison pouch view
1		Uterine cervix visible	Endometrial midline echo is horizontal	Side stated (left or right)	Liver visible
	Inter-	0.80	0.93	1/1	1
	Intra-	0.80	1	1/1	1
2		Uterine fundus visible	One interstitial tube visible	Follicle(s) visible	Kidney visible
	Inter-	0.90	0.78	0.9/0.9	1
	Intra-	0.80	0.78	1/1	1
3		Endometrial midline echo visible	The other interstitial tube visible	Iliac vein visible	Ovoid section of the kidney
	Inter-	1	0.78	0.75/0.85	0.89
	Intra-	1	0.78	1/1	1
4		Endocervix visible	Visible space below the posterior uterine wall.	Long axis of the ovary $<30^{\circ}$ with the horizontal line	
	Inter-	0.80	0.93	0.85/0.85	
	Intra-	0.80	1	1/0.80	
5		Uterus occupying more than half of the total mage size	Uterus occupying more than half of the total image size	Ovary occupying more than a quarter of the total image size	
	Inter-	0.80	0.93	0.95/0.90	
	Intra-	1	1	0.80/0.80	

patterns known to have a diagnostic value. Quality assessment was based on image review with criteria-based scoring. The criteria were simple and chosen in order to improve the reproducibility of the sonographer's examination. This objective evaluation can be compared to that reported by Herman et al. [11] for first trimester nuchal translucency measurement, and the use of criteria for quality assessment of obstetrical biometrical planes as well [12].

Our study has nonetheless several limitations. First of all, development of the standardised images and criteria was agreed on the basis of a review of the literature: it was not a multi-centre expert consensus. Second, a qualitative review required the definition of criteria which may seem arbitrary, and the choice of these criteria may also introduce a bias in quality assessment if a systematic mistake for one criterion is masked by a good overall score. Although we demonstrated good reliability and reproducibility for this criteria-based score, it could be argued that the results might have been different using a different system or using different rules to evaluate agreement between reviewers. Moreover, each criterion was given equal weight: the scoring system is not constructed according to the clinical utility of the sonograms. We could have given different weights in order to emphasize certain quality aspects. Last, the study concerns an improvement in the care process but does not assess whether any perceived change has resulted in improved patient care. This is being prospectively evaluated and will be reported in another study. Moreover there is no evidence in the study that the perceived improvement in the quality of scanning was not due to training, but rather due to increased experience of the residents/registrars over time, for example. However, because the registrars already had between 2 and 4 years of experience, it is unlikely that the improvement in the quality was due to the extra experience over four months only.

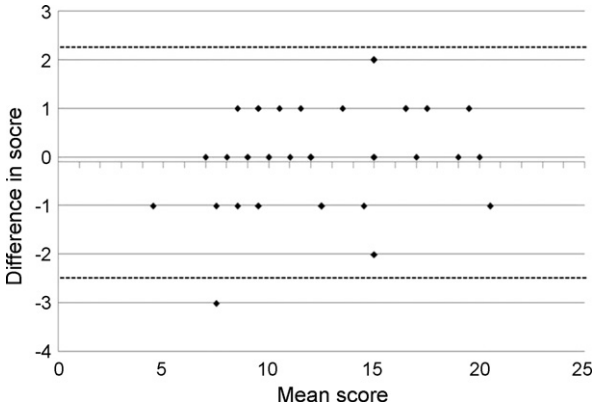

Fig. 3. Bland and Altman plot for the agreement in scoring between reviewer A and B.

The criteria and images we chose deserve explanation. Early pregnancy-related complications are one of the main reasons for consultation in the gynaecological emergency department. In a prospective study reporting ultrasound examination at assessment of 1000 emergency gynaecology patients, it appears that more than 50% of patients were pregnant, among whom 75% had an intra-uterine pregnancy [7]. Therefore, it appears mandatory to obtain good quality ultrasound images of the uterine cavity in order to demonstrate or rule out an intra-uterine pregnancy [18,19]. Jurkovic et al. stated that appropriate diagnosis of ectopic pregnancies should be based on a good understanding of pelvic anatomy and the pathophysiology of ectopic pregnancies, combined with a systematic examination of the key morphological features [13]. Accordingly, our criteria for these two images required an adequate visualisation of the uterus, from cervix to fundus, with the midline echo visible. We also asked for visualisation of the interstitial part of the tube on the transversal view, in order not to overlook interstitial or cornual pregnancies [13]. Above all, the two images we required meant the uterus had to be scanned in two different planes. Adequate visualisation of the space below the posterior wall of the uterus and assessment of Morrison's pouch should make it easy to diagnose any large accumulation of fluid in the pelvic or abdominal cavity [6,7,13,20–24]. The existence of free fluid at Morrison's pouch level is reported to be correlated with haemoperitoneum of more than 400 ml [25,26]. Right upper quadrant ultrasound examination should therefore be recommended as a part of routine gynaecological ultrasound examination [26]. Appropriate quantification of intraperitoneal fluid at US is important because fluid in the Pouch of Douglas may be present in up to 80% of EP whether ruptured or not [20,27]. We demonstrated in a previous study that use of a semi-quantitative estimate of haemoperitoneum volume, using transvaginal US examination of the uterus in a sagittal plane (adapted from [28]), showed an excellent correlation with the volume of haemoperitoneum measured during surgery. Last,

adequate ultrasound examination of the ovary permits diagnosis of ovarian cyst, which is a frequent reason for emergency consultation [7]. Three or more optional images to describe specific abnormalities found at examination can be added to the standard set: one for intra-uterine abnormality, one for extra-uterine abnormality, and one for any other abnormal finding. These optional images were not evaluated in this study.

This score-based method brought a degree of objectivity into image-quality evaluation and allowed for good inter- and intra-reviewer reproducibility with kappa coefficients above 0.8 in most cases (Table 2). Timor-Tritsch et al. demonstrated that sonography significantly decreases evaluation and disposition time for emergency department patients in early pregnancy. Many studies have emphasized the utility of ultrasound examination at emergency consultation, however, the ultrasound protocols are highly variable and comparison between studies is not easy [7,29]. This study should allow the definition of clear ultrasound standards and assessment of the standard of care provided in various reports. Although our results would need to be reproduced with a larger group of operators and examiners, this scoring method could be used to standardise policies for ultrasound examination, and its impact on ongoing quality control on patient outcomes should be evaluated in future studies. Studies have already demonstrated that emergency physicians were able to perform sonographic examinations safely and accurately with relatively limited training [30]. Besides its potential for audit and quality control, this image scoring method could also be useful during the training process. It could be a useful tool to evaluate the performance of trainees (since it would allow identification and correction of specific weaknesses) and to monitor the progress of prospective trainees in gynaecological ultrasonography. In this study, we demonstrated that there was a significant increase in scores following training ($p < 10^{-4}$).

To conclude, we have developed a scoring method for assessment of emergency gynaecological ultrasound examination. This score appears feasible with good inter- and intra-reviewer reproducibility. The true contribution of this approach should now be tested on a larger scale and the influence of an ongoing audit process of this nature should be evaluated. Although we were able to demonstrate an improvement in the quality of the examinations, no claim for increased utility of ultrasound examination can be made on the basis of this preliminary study. Further studies are needed to prove whether or not this improvement in the quality of the US procedure may shorten the time for diagnosis and repeated examination, or limit the risks of misdiagnosis.

References

[1] Lazebnik N, Lazebnik RS. The role of ultrasound in pregnancy-related emergencies. Radiol Clin North Am 2004;42(2):315–27.
[2] Taipale P, Tarjanne H, Ylostalo P. Transvaginal sonography in suspected pelvic inflammatory disease. Ultrasound Obstet Gynecol 1995;6:430–4.
[3] Ubaldi F, Wisanto A, Camus M, Tournaye H, Clasen K, Devroey P. The role of transvaginal ultrasonography in the detection of pelvic pathologies in the infertility workup. Hum Reprod 1998;13(2):330–3.
[4] Taipale P, Tarjanne H, Heinonen UM. The diagnostic value of transvaginal sonography in the diagnosis of endometrial malignancy in women with peri- and postmenopausal bleeding. Acta Obstet Gynecol Scand 1994;73(10):819–23.
[5] Bigrigg MA, Read MD. Management of women referred to early pregnancy assessment unit: care and cost effectiveness. BMJ 1991;302(6776):577–9.
[6] Shalev E, Yarom I, Bustan M, Weiner E, Ben-Shlomo I. Transvaginal sonography as the ultimate diagnostic tool for the management of ectopic pregnancy: experience with 840 cases. Fertil Steril 1998;69(1):62–5.
[7] Haider Z, Condous G, Khalid A, et al. Impact of the availability of sonography in the acute gynecology unit. Ultrasound Obstet Gynecol 2006;28(2):207–13.
[8] Mikkelsen AL, Felding C. Laparoscopy and ultrasound examination in women with acute pelvic pain. Gynecol Obstet Invest 1990;30:162–4.
[9] Adhikari S, Blaivas M, Lyon M. Diagnosis and management of ectopic pregnancy using bedside transvaginal ultrasonography in the ED: a 2-year experience. Am J Emerg Med 2007;25(6):591–6.
[10] AIUM. AIUM Practice Guideline for the performance of an antepartum obstetric ultrasound examination. J Ultrasound Med 2003;22(10):1116–25.
[11] Herman A, Maymon R, Dreazen E, Caspi E, Bukovsky I, Weinraub Z. Nuchal translucency audit: a novel image-scoring method. Ultrasound Obstet Gynecol 1998;12(6):398–403.
[12] Salomon LJ, Bernard JP, Duyme M, Doris B, Mas N, Ville Y. Feasibility and reproducibility of an image scoring method for quality control of fetal biometry in the second trimester. Ultrasound Obstet Gynecol 2006;27(1):34–40.
[13] Jurkovic D, Mavrelos D. Catch me if you scan: ultrasound diagnosis of ectopic pregnancy. Ultrasound Obstet Gynecol 2007;30(1):1–7.
[14] Lambert MJ, Villa M. Gynecological ultrasound in emergency medicine. Emerg Med Clin North Am 2004;22(3):683–96.
[15] Bennett E, Albert R, Goldstein A. Communications through limited response questioning. Public Opinion Q 1954;18:303–8.
[16] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–10.
[17] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
[18] Mol BW, Hajenius PJ, Engelsbel S, et al. Can noninvasive diagnostic tools predict tubal rupture or active bleeding in patients with tubal pregnancy? Fertil Steril 1999;71(1):167–73.
[19] Ankum WM. Diagnosing suspected ectopic pregnancy. Hcg monitoring and transvaginal ultrasound lead the way. BMJ 2000;321(7271):1235–6.
[20] Frates MC, Brown DL, Doubilet PM, Hornstein MD. Tubal rupture in patients with ectopic pregnancy: diagnosis with transvaginal US. Radiology 1994;191(3):769–72.
[21] Chen PC, Sickler GK, Dubinsky TJ, Maklad N, Jacobi RL, Weaver JE. Sonographic detection of echogenic fluid and correlation with culdocentesis in the evaluation of ectopic pregnancy. Am J Roentgenol 1998;170(5):1299–302.
[22] Sickler GK, Chen PC, Dubinsky TJ, Maklad N. Free echogenic pelvic fluid: correlation with hemoperitoneum. J Ultrasound Med 1998;17(7):431–5.
[23] Stengel D, Bauwens K, Sehouli J, et al. Emergency ultrasound-based algorithms for diagnosing blunt abdominal trauma.. Cochrane Database Syst Rev 2005;18(2):CD004446.
[24] Ma OJ, Gaddis G, Steele MT, Cowan D, Kaltenbronn K. Prospective analysis of the effect of physician experience with the FAST examination in reducing the use of CT scans. Emerg Med Australas 2005;17(1):24–30.
[25] Branney SW, Wolfe RE, Moore EE, et al. Quantitative sensitivity of ultrasound in detecting free intraperitoneal fluid. J Trauma 1995;39(2):375–80.
[26] Rodgerson JD, Heegaard WG, Plummer D, Hicks J, Clinton J, Sterner S. Emergency department right upper quadrant ultrasound is associated with a reduced time to diagnosis and treatment of ruptured ectopic pregnancies. Acad Emerg Med 2001;8(4):331–6.
[27] DiMarchi JM, Kosaka TS, Hale RW. What is the significance of the human chorionic gonadotropin value in ectopic pregnancy? Obstet Gynecol 1989;74:851–5.
[28] Dart R, McLean SA, Dart L. Isolated fluid in the cul-de-sac: how well does it predict ectopic pregnancy? Am J Emerg Med 2002;20(1):1–4.
[29] Kirk E, Condous G, Van Calster B, Van Huffel S, Timmerman D, Bourne T. Rationalizing the follow-up of pregnancies of unknown location. Hum Reprod 2007;22(6):1744–50.
[30] Shih CH. Effect of emergency physician-performed pelvic sonography on length of stay in the emergency department. Ann Emerg Med 1997;29(3):348–51. discussion 352.