Accuracy of imaging modalities for adnexal torsion: a systematic review and meta-analysis

B Wattar, M Rimmer, E Rogozinska, M Macmillian, KS Khan, BH Al Wattar

West Suffolk NHS Foundation Trust, Bury St Edmunds, Suffolk, UK
MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, UK
MRC Clinical Trials Unit, University College London, London, UK
MRC Centre for Regenerative Medicine, The University of Edinburgh, Edinburgh, UK
Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain
Warwick Medical School, University of Warwick, Coventry, UK

Correspondence: BH Al Wattar, Warwick Medical School, Medical School Building, Coventry, CV4 7HL, UK. Email: dr.basselwa@gmail.com

Accepted 5 June 2020.

Background Adnexal torsion (AT), a serious gynaecological emergency, often presents with non-specific symptoms leading to delayed diagnosis.

Objective To compare the test accuracy of ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI) to diagnose AT.

Search strategy We searched EMBASE, MEDLINE and Cochrane CENTRAL until December 2019.

Selection criteria Studies reporting on the accuracy of any imaging modality (Index Test) in female patients (paediatric and adult) suspected of AT compared with surgical diagnosis and/or standard clinical/radiological follow-up period until resolution of symptoms (Reference Standard).

Data collection and analysis We assessed study quality using QUADAS-2. We conducted test accuracy meta-analysis using a univariate model or a hierarchical model.

Main results We screened 3836 citations, included 18 studies (1654 women, 665 cases), and included 15 in the meta-analyses. Ultrasound pooled sensitivity (n = 12,1187 women) was 0.79 (95% CI 0.63–0.92) and specificity was 0.76 (95% CI 0.54–0.93), with negative and positive likelihood ratios of 0.29 (95% CI 0.13–0.66) and 4.35 (95% CI 2.03–9.32), respectively. Using Doppler with ultrasound (n = 7, 845 women) yielded similar sensitivity (0.80, 95% CI 0.67–0.93) and specificity (0.88, 95% CI 0.72–1.00). For MRI (n = 3, 99 women), the pooled sensitivity was 0.81 (95% CI 0.63–0.91) and specificity was 0.91 (95% CI 0.80–0.96). A meta-analysis for CT was not possible with two case-control studies and one cohort study (n = 3, 232 women). Its sensitivity range was 0.74–0.95 and specificity was 0.80–0.90.

Conclusions Ultrasound has good performance as a first-line diagnostic test for suspected AT. Magnetic resonance imaging could offer improved specificity to investigate complex ovarian morphology, but more evidence is needed.

Keywords Adnexa, computed tomography, Doppler, magnetic resonance imaging, meta-analysis, ovary, test accuracy, torsion, ultrasound.

Tweetable abstract To investigate adnexal torsion, ultrasound is a good first-line diagnostic test with a pooled sensitivity of 0.79 and specificity of 0.76.

Introduction

Adnexal torsion (AT) is a serious gynaecological emergency that involves a partial or complete twisting of the infundibulopelvic vascular pedicle. It acutely compromises the vascular supply of the ovary and the adjacent fallopian tube eliciting ischaemia, tissue necrosis, reduced ovarian follicular reserve, subfertility and early menopause.1 Its prevalence is unclear, but it is estimated to affect 2–7% of women undergoing surgery for acute pelvic pain.2 Most affected women present with non-specific symptoms such as abdominal pain, vomiting and fever leading to delayed diagnosis and increased risk of emergency oophorectomy.3 As such, establishing a prompt diagnosis is key to enable early surgical untwisting and restoration of the compromised vascular supply.

To aid its diagnosis, numerous imaging modalities have been used and evaluated in the literature.4 Ultrasound is commonly used to evaluate ovarian pathology because of its safety, availability and affordability. However, several...
factors could limit its accuracy to diagnose AT such as operators experience, machine quality, pregnancy and presence of complex ovarian morphology. Doppler is often used to highlight the compromised vascular supply to the adnexa; however, its added diagnostic value remains imprecise. Both computed tomography (CT) and magnetic resonance imaging (MRI) have been used to evaluate complex ovarian morphology; however, their use to diagnose AT could be hampered by the variations in diagnostic criteria and the experience of the assessor. Test accuracy for these modalities is not precisely known, thus increasing variations in practice and hindering effective policy-making.

We aimed to compare the test accuracies of the various imaging modalities used to diagnose AT by conducting a systematic review and meta-analysis.

Methods

We conducted a systematic review using an established methodology for test accuracy research and a prospectively registered protocol (CRD42018112048). We reported findings of our review as per established guidelines. Patients were not involved in the design and conduct of this review. We searched the COMET database and did not identify any relevant core outcome sets on the topic of interest.

Literature search

We searched the major electronic databases (EMBASE, MEDLINE and Cochrane CENTRAL) for primary diagnostic accuracy studies for adnexal torsion from inception until December 2019. We performed complementary searches in ClinicalTrials.gov, Google Scholar and Scopus to capture any relevant additional citations. We did not employ any search filters or language restrictions. We used MeSH terms (ovarian, ovary, tube, fallopian, twisted, torsion, adnexa, adnexal, adnexe) and combined them using the Boolean operators AND/OR to produce a sensitive search. We searched the bibliographies of potentially relevant articles to identify any additional citations not captured by our search.

Study selection and data extraction

We performed the study selection and inclusion process in two stages. First, two reviewers (BW and MPR) screened the titles and abstracts of potentially relevant articles. In the second stage, we assessed relevant articles in full against our inclusion criteria before inclusion. We included all primary studies reporting on the diagnostic accuracy of any imaging modality (Index Test) used in female patients (paediatric and adult) presenting with symptoms suggestive of AT (acute/sub-acute abdominal/pelvic pain, fever, nausea, vomiting, pelvic mass) compared with surgical diagnosis and/or standard clinical/radiological follow-up period until resolution of symptoms (Reference Standard) in no preferential order. We excluded studies reporting only on fetal/neonatal adnexal torsion or on isolated tubal torsion. We also excluded reviews, case reports and case series. Studies that identified their population by ‘asymptomatic ovarian mass’ were also excluded because this can overestimate the diagnostic accuracy. Any disagreements were resolved in consensus with a third reviewer (BHA). Studies that were of case–control design were included in our systematic review but not in the meta-analysis.

We extracted data in duplicate onto a piloted electronic data extraction sheet. We collected data on population characteristics, description of the index and reference tests, diagnostic criteria used, treatment algorithm in each study, and the duration of follow up.

Quality assessment of included studies

Two reviewers (BW and MPR) independently assessed the risk of bias and applicability of the included studies using the QUADAS-2 in four domains: patient selection, conduct of the index test, conduct of the reference standard and patient flow. We considered a study to be of high quality if it used a patient spectrum matching the review question, enrolled a consecutive or random sample of patients, used the index test as first-line imaging with a pre-defined benchmark for a positive test, all participants had surgical confirmation within 48 hours as reference standard, and the majority of recruited participants were included in analyses. The following were considered to be inappropriate patient spectra that introduced bias: cohorts limited to only paediatric, pregnant or non-pregnant women, studies involving women with asymptomatic pelvic mass, and studies with inappropriate exclusions. Lack of blinding to index test results upon the interpretation of the results of the reference standard was not considered to pose a high risk of bias.

Data synthesis

We constructed 2 x 2 tables for each imaging modality and calculated sensitivity, specificity and likelihood ratios for positive and negative test results with 95% CIs. We pooled the accuracy parameters using a hierarchical model (random effect) when a sufficient number of studies (at least four) were available. When fewer than four studies were available, we used a univariate model. We investigated heterogeneity visually from forest plots of sensitivity and specificity estimates. We considered the use of Doppler to be a potential effect-modifier in studies evaluating the use of ultrasound and investigated it using a meta-regression. We performed subgroup analyses to evaluate the effect of potential confounders (e.g. population age, pubertal status). We did not assess the publication bias because of the small number of studies included for each imaging
modality. We conducted our analysis using RevMan version 5.3, Open Meta-analyst software version 12.11.14, and Stata version 14 (StataCorp, College Station, TX, USA, 2015).

Funding
No funding was received directly to support this work.

Results
Characteristics of included studies
We identified 3836 potentially relevant citations; of these, 124 were reviewed in full against our inclusion criteria and 18 were included reporting on 1654 women (Figure 1). Most studies (15/18, 83%) were cohorts (14 retrospectives

![Image](image_url)
and one prospective) but three were retrospective case–control studies (3/18, 17%), one reporting on CT, one on ultrasound, and one on ultrasound and CT. The median sample size was 71 (range 29–323) with 665 confirmed cases of AT (665/1654, 40%). There were four studies from the USA (4/18, 22%), four from Israel (4/18, 22%), three from Korea (3/18, 17%), two from France (2/18, 11%) and one from each of India, Iran, China, Canada, and Saudi Arabia (see Supplementary material, Table S1). Two-thirds of studies used surgical exploration as the Reference Standard (12/18, 67%), while six used a mixture of surgical exploration and clinical follow up (6/18, 33%). Three studies reported on each of CT9,12,13 and MRI14–16 (3/18, 17%).

Fourteen studies reported on the accuracy of ultrasound (14/18, 44%), of these nine included the use of Doppler (9/14, 64%) and five included only adults (5/14, 36%) whereas the remaining included a mixture of paediatric and adult patients or did not report on age of participants. Ten ultrasound studies only used surgical exploration as a Reference test (10/14, 71%) whereas the remaining four used a mixture of surgical and clinical follow up.

**Quality of included studies**

The overall quality of included studies was moderate with two-thirds of included studies showing a high risk of bias for patient selection and applicability (Figure 2). The conduct and the applicability of the index and the reference tests were thought to be adequate in the majority of studies with only four showing a high risk of bias (4/18, 22%) for the index test. Seven studies showed a high risk of bias for

**Figure 2.** Quality of included studies on the diagnostic accuracy of imaging modalities in women with suspected adnexal torsion.
patient flow and timing of testing in the study (7/18, 39%) and six studies showed no risk of bias for these items (6/18, 33%) (Figure 2).

Test accuracy meta-analysis

The pooled sensitivity and specificity for ultrasound (12 studies, 1187 women)\textsuperscript{16–27} were 0.79 (95% CI 0.63–0.92) and 0.76 (95% CI 0.54–0.93) with a negative and positive likelihood ratio of 0.29 (95% CI 0.13–0.66) and 4.35 (95% CI 2.03–9.32), respectively. Visual inspection of heterogeneity showed greater variability in the sensitivity than the specificity measures (Figure 3). We evaluated the additional use of Doppler with ultrasound in a meta-regression (seven studies, 845 women)\textsuperscript{18–20,22–24,26} that showed slight improvements in sensitivity (0.80, 95% CI 0.67–0.93) and specificity (0.88, 95% CI 0.72–1.00), though not statistically significant (joint model, \(P\)-value = 0.7). We also conducted subgroup analyses in studies using surgical exploration only as Reference test (\(n = 9\), sensitivity 0.81, 95% CI 0.61–0.94, specificity 0.73, 95% CI 0.42–0.94)\textsuperscript{18–24,26,27} and in those reporting on adults only (\(n = 3\), sensitivity 0.84, 95% CI 0.34–0.98, specificity 0.78, 95% CI 0.42–0.94).\textsuperscript{19,20,27} Both subgroups showed similar estimates to the whole population.

Test accuracy meta-analysis for MRI (3 studies, 99 women)\textsuperscript{14–16} showed pooled sensitivity of 0.81 (95% CI 0.63–0.91) and specificity of 0.91 (95% CI 0.80–0.96) (Figure 3). With two case–control studies and one cohort study (\(n = 3, 232\) women), a meta-analysis for CT was not possible. It had a reported sensitivity ranging from 0.74 to 0.95, and specificity from 0.80 to 0.90. Figure 4 illustrates the scatter of the accuracy parameters for all reported imaging modalities across the included studies.

Discussion

Main findings

Our findings support an overall good performance for ultrasound as a first-line diagnostic tool for AT. Evaluating the ovarian vascular blood flow using Doppler slightly improved the diagnostic accuracy of ultrasound, though

![Figure 3. Estimates of sensitivity and specificity for (A) ultrasound, (B) magnetic resonance imaging and (C) computed tomography scan to diagnose suspected adnexal torsion.](image-url)
Figure 4. Scatter plot illustrating the accuracy of the various imaging modalities for diagnosing suspected adnexal torsion.

Strengths and limitations
We conducted our review using a standard methodology for diagnostic accuracy reviews, registered our protocol prospectively, and reported according to established guidelines. We adopted a pragmatic search strategy and inclusion criteria including all suspected cases of AT to offer the most comprehensive patient spectrum for evidence synthesis. We considered the potential effect of Doppler on the accuracy of ultrasound using a meta-regression and performed sub-group analyses where possible.

Our findings are not without limitations. Overall, our pooled estimates suffered from heterogeneity probably as the result of variations in the characteristics of the women included (such as age and reproductive status) in our meta-analysis, so we interpret the findings with caution. Our inclusion criteria are pragmatic and comprehensive to capture the whole literature on the diagnosis of AT. However, we acknowledge the increased heterogeneity and the potential effect of several confounders such as variations in age, reproductive status, operator experience and sequential testing. Most studies included a mixed population of paediatric and adult female patients, which limited our ability to adjust for important factors such as ultrasound route (transabdominal versus transvaginal) and the underlying ovarian pathology (e.g. dermoid cysts). Adjustment for such factors would only be possible using an Individual Patient Data meta-analysis, which was not feasible in our review. Still, we believe our review to offer the most comprehensive evidence synthesis at present to advise current clinical practice.

Interpretation
Establishing an accurate diagnosis in women with suspected AT remains a clinical challenge because of the non-specific presentation and the varied differential diagnosis. Several ovarian pathologies could produce similar radiological signs (including ovarian oedema, unilateral enlargement, midline shift) as well as overlap with an acute AT (e.g. teratoma, endometrioma, haemorrhagic cyst) complicating the radiological diagnosis. As a gynaecological emergency, rapid diagnosis of AT is crucial to optimise the outcomes of affected women and advise any planned surgical intervention (e.g. laparoscopy for smaller masses versus laparotomy for large complex torsion). Our estimates support the role of ultrasound as a reliable first-line diagnostic tool for AT. Certainly, several emergency departments now offer rapid-access ultrasound to aid the diagnosis in women with non-specific abdominal pain, which seems to optimise the diagnosis and management process. Our findings depict relatively wide confidence intervals for the accuracy of ultrasound to diagnose AT. Therefore, clinicians should consider the diagnostic limitations of ultrasound, especially when faced with complex ovarian morphology such as very large cysts, complex masses or paediatric cases, which this was not statistically significant with overlapping confidence intervals. Assessment of CT and MRI was limited by the number of available studies on those two modalities. Overall, MRI seemed to offer higher specificity, which could be of value when investigating ambiguous adnexal masses with high suspicion of torsion, but more studies are needed to define the role of MRI in the diagnostic pathway of AT. Data pooling was not possible for CT, though its reported range was consistent with that of ultrasound.
might increase the rate of false-negative findings. Given the established limitations of ultrasound, clinicians should correlate the clinical, biochemical and radiological findings before deciding to operate on symptomatic women. Such practice is key specifically when planning the management of particular patient groups (e.g., prepubertal girls and pregnant women) to aid the decision-making for the surgical route of choice (e.g., laparotomy for large complex masses) and the surgical approach (oophorectomy versus conservative surgery).2

The role of MRI in investigating larger and more complex ovarian morphology is well established.29–32 However, considering its higher cost and limited availability, reserving its use as a second-line diagnostic tool seems reasonable within the context of our findings. We were unable to identify uniform diagnostic criteria to establish an ultrasonographic diagnosis of AT because of the varied reporting across included studies. This was also the case for reported diagnostic radiological features on CT and MRI. Certain features seem to be more suggestive of AT (e.g., ovarian oedema >5 cm, twisted pedicles on colour Doppler, free fluid in the pelvis and the whirlpool sign).17,19,33 However, future consensus work is needed to evaluate the accuracy of unified diagnostic criteria that correlate with the clinical presentation.

Establishing a well-defined care pathway for women presenting with acute abdominal/pelvic pain shared across multiple disciplines is key for efficient diagnosis and management of AT.34 Currently, care for affected women is heterogeneous, often tailored by the attending clinician and their speciality of interest (emergency medicine, general surgery, urology, gynaecology) increasing the chance of delayed diagnosis and treatment. Developing and evaluating standardised care pathways with rapid access to imaging services is needed to improve the long-term outcomes of women with AT.

Conclusion

Ultrasound has good performance as a first-line diagnostic test for women with suspected AT. Magnetic resonance imaging could offer improved specificity to investigate complex ovarian morphology, but more evidence is needed.

Disclosure of interests

All authors have nothing to disclose. BHAW holds a personal Lectureship from the UK National Health Institute of Research. KSK is Distinguished Investigator at the University of Granada with a grant awarded by the Beatriz Galindo Programme (senior modality) of the Spanish Ministry of Science, Innovation and Universities. Completed disclosure of interests forms are available to view online as supporting information.

Contribution to authorship

BW and MR conducted the search, data extraction and first draft of the manuscript. BHAW and ER conducted the statistical analysis. MM and KSK contributed to data interpretation and the final editing of the manuscript.

Funding

None.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Characteristics of included studies on the diagnostic accuracy of imaging modalities in women with suspected adnexal torsion.

References


