MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women: a test-accuracy study and economic evaluation

Khalid S Khan,[1] Konstantinos Tryposkiadis,[2] Seema A Tirlapur,[1] Lee J Middleton,[2] Andrew J Sutton,[3,4] Lee Priest,[2] Elizabeth Ball,[5] Moji Balogun,[6] Anju Sahdev,[7] Tracy Roberts,[3] Judy Birch,[8] Jane P Daniels[2,9*†] and Jonathan J Deeks[2†] on behalf of the MEDAL study group

[1]Women's Health Research Unit, Queen Mary University of London, London, UK
[2]Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK
[3]Health Economics Unit, University of Birmingham, Birmingham, UK
[4]Leeds Institute of Health Sciences and NIHR Diagnostic Evidence Co-operative Leeds, University of Leeds, Leeds, UK
[5]Obstetrics and Gynaecology, Barts Health NHS Trust, London, UK
[6]Radiology Department, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK
[7]Radiology Department, Barts Health NHS Trust, London, UK
[8]Pelvic Pain Support Network, Poole, UK
[9]Nottingham Clinical Trials Unit, University of Nottingham, Nottingham, UK

*Corresponding author Jane.Daniels@nottingham.ac.uk
†Joint last author

Declared competing interests of authors: Elizabeth Ball declares UK travel reimbursement from Shire Medical (Lexington, MA, USA) outside the submitted work. Jonathan J Deeks was Deputy Chairperson of the National Institute for Health Research Health Technology Assessment (HTA) Commissioning Board (2011–16) and the HTA Efficient Study Designs Board (2016).

Published July 2018
DOI: 10.3310/hta22400

Scientific summary

MRI vs. laparoscopy for CPP diagnosis

Health Technology Assessment 2018; Vol. 22: No. 40
DOI: 10.3310/hta22400

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Chronic pelvic pain (CPP) may be defined as pain in the pelvic and lower abdominal region, that lasts for ≥ 6 months. Idiopathic CPP is defined as CPP with an uncertain or unknown structural cause, after the exclusion of any other recognisable gynaecological pathology. Conditions such as endometriosis, fibroids, adenomyosis, cystic ovaries, adhesions and pelvic inflammatory disease are potential gynaecological causes, although CPP can have origins in the bowel, bladder, pelvic bones, muscles and fascia.

Symptoms experienced in CPP are variable and non-specific, so establishing a differential diagnosis can be hard. The Royal College of Obstetricians and Gynaecologists guidelines on the management of CPP [Royal College of Obstetricians and Gynaecologists. *Chronic Pelvic Pain, Initial Management (Green-top Guideline No. 41)*. London: Royal College of Obstetricians and Gynaecologists; 2012] provide a number of suggested initial investigations, including history, microbiological screening and vaginal examination, all based on low-quality evidence. If no cause of the pain is found, a diagnostic laparoscopy is often performed, although many conditions associated with CPP are not amenable to either laparoscopic diagnosis or treatment.

The aim of the Magnetic resonance imaging to Establish Diagnosis Against Laparoscopy (MEDAL) study was to assess if magnetic resonance imaging (MRI) can add value in the decision to perform a laparoscopy in women presenting with CPP. The important question was whether or not a normal MRI scan has a high enough negative predictive value to replace and avoid laparoscopy.

Objectives

1. To estimate the accuracy of post-MRI diagnoses with respect to (1) the absence of any observed condition or cause (i.e. idiopathic CPP); and (2) the main gynaecological conditions or causes of CPP as target conditions, using both post-laparoscopy diagnoses and expert independent panel (EIP) consensus (with and without incorporating MRI findings) as reference standards.
2. To determine the added value of MRI in a care pathway involving baseline history, clinical examination and ultrasound before performing laparoscopy.
3. To quantify the impact that pre-laparoscopy MRI scans can have on decision-making, with respect to triaging for therapeutic laparoscopy.
4. To perform a decision-analytic model-based economic evaluation determining the cost-effectiveness of MRI compared with laparoscopy.

Methods

We performed a comparative test-accuracy study with panel consensus for determining the reference standard.

Participants and setting

We recruited women aged ≥ 16 years, who had been referred to a gynaecologist, with CPP of at least 6 months' duration, and in whom there was an indication for diagnostic laparoscopy, from 26 UK hospitals. Women were excluded if they had had a hysterectomy, were pregnant, were unable to give written informed consent, definitely had a clinical indication for a MRI scan or if they had a previously established cause of CPP. Information concerning menstrual, obstetric and contraceptive history, pelvic examination and ultrasound scans was collected at baseline. Patient-completed questionnaires captured

domains such as (1) the location and severity of pain; (2) neuropathic, bowel-, urinary- and endometriosis-associated pain symptoms; (3) personality characteristics; (4) coping strategies; (5) depression; (6) prior sexual and physical abuse; and (7) general quality of life. Six months after diagnostic laparoscopy, data regarding treatments offered were collected and patient questionnaires were repeated.

Index tests

Diagnostic laparoscopy and MRI scans were the index tests. The MRI standard protocol comprised T1 and T2 axial, T2 sagittal, T2 coronal and T1-fast-spin (FS) axial, sagittal and coronal sequences. The observations were captured on a standardised report form and verified by a second radiologist who was blind to the initial MRI report. The diagnostic laparoscopy was performed under general anaesthetic in accordance with the gynaecologist's standard practice. When clinically indicated, additional therapeutic and diagnostic procedures were performed at the time of surgery.

To avoid verification bias, MRI was performed before diagnostic laparoscopy, and treating gynaecologists and investigators were kept blind to the MRI reports and images. To model the impact of using the MRI scan report, an independent gynaecologist reviewed each case. This gynaecologist was randomly selected from another hospital recruiting to the MEDAL study.

At several stages in the study path for each patient, the treating gynaecologist and the independent gynaecologist reviewed the diagnostic evidence available. To establish a possible diagnosis for each target condition (including idiopathic CPP), the gynaecologists indicated their level of certainty that the condition was causing the pelvic pain on a numerical rating scale, ranking from 0, meaning no chance, to 10, meaning certain.

Reference standard

The analysis considered two sets of reference standard. The first set identified was to assess the accuracy of MRI for observing a condition; the second was to assess the accuracy of MRI for identifying the condition(s) causing pelvic pain. The first set used findings on laparoscopy for the reference standard; the second used consensus from an EIP. The consensus process, the target condition definitions and the data presentation were standardised prior to the panel meetings, and reproducibility was assessed. The first-stage reference diagnosis was based on patient history and reported symptoms, clinical examination, ultrasound, laparoscopy and follow-up. For the second-stage reference diagnosis, the MRI scan report was provided. The panel members selected all conditions for which they were at least 50% certain that these were the underlying cause.

Sample size

A sample size of 250 women was chosen to address the primary research question of how many women could have avoided laparoscopy if MRI was routinely used in practice for diagnosing women with CPP. The study had over 90% power (at $p = 0.05$) to detect a reduction of 10% in the number of laparoscopies needed (i.e. from 100% down to 90%). With a prevalence of $\geq 30\%$ of idiopathic CPP, this sample size had the ability to reliably rule out a sensitivity or specificity of $< 70\%$ if the 'true' sensitivity or specificity was $> 80\%$.

Analysis

The prevalence of all gynaecological and non-gynaecological conditions in relation to whether or not they were believed to be causing pain (as determined by the reference standard established by consensus by the EIP) was calculated along with 95% confidence intervals (CIs) using binomial exact methods. The kappa value, alongside the 95% CIs, indicated the agreement between the three individual members of the EIP in their diagnoses.

The primary analysis involved calculations of sensitivity and specificity (and the associated 95% CIs) for the presence or absence of each structural gynaecological cause of pain; data were taken from the binary yes/no responses of the reference standard before and after the MRI data were revealed. Receiver operator

characteristic curves were also constructed for each test for each condition, using the certainty estimates provided by the treating gynaecologists and the independent gynaecologists, and the binary responses of the reference standard. The area under the receiver operating characteristic (AUROC) curve and 95% CI was estimated for each test. Differences between the AUROC estimates of each of the four diagnoses (pre and post laparoscopy by treating gynaecologist, and pre and post MRI by independent gynaecologist) were then calculated using a non-parametric approach for correlated data and are presented along with the 95% CIs. The same analysis was done to compare the responses after the MRI data were revealed to the EIP.

Cost-effectiveness evaluation

The cost-effectiveness for a range of scenarios for MRI and laparoscopy using different sensitivity and specificity values, as well as do-nothing (no MRI or laparoscopy) scenarios, was examined for women with undiagnosed CPP.

A decision-analytic model with a 6-month time horizon was developed in consultation with the clinical and methodological team. Two pathways for laparoscopy and MRI were constructed, with the cut-off value on the receiver operating characteristic (ROC) curve used to inform the sensitivity and specificity of each test being treated as a decision variable. Data collected in the MEDAL study to inform the prevalence, quality-adjusted life-year (QALY) and test-accuracy data were used to parameterise the economic model. Cost data were obtained from standard NHS sources.

A cost–utility analysis was undertaken based on the primary outcome of cost per QALY. The secondary outcomes of the cost per patient correctly diagnosed and the cost per patient appropriately treated were also considered as part of the wider cost-effectiveness analyses. The perspective of the analysis was the health-care provider perspective (the UK NHS) in a secondary care (hospital) setting. Given the 6-month time horizon, no discounting was applied. The results in this study are described using the incremental cost-effectiveness ratio (ICER). A range of one-way sensitivity analyses were undertaken to gain further insights into the impact of reasonable changes on key parameters in the model. A probabilistic sensitivity analysis was also conducted.

Results

Main findings of the test-accuracy study

The diagnostic study demonstrated that, compared with the reference standard diagnosis verified with laparoscopy, MRI had high specificity but poor sensitivity for observing deep-infiltrating endometriosis (3%), endometrioma (33%), adhesions (19%) and ovarian cysts (20%). Sensitivity was higher for detecting these conditions as a cause of pelvic pain, as categorised by the EIP, but not high enough to be useful. MRI correctly identified 56% (95% CI 48% to 64%) of women judged to have idiopathic CPP, but missed 46% (95% CI 37% to 55%) of women considered to have a gynaecological structural cause of CPP.

Magnetic resonance imaging added significant value in the care pathway for determining the cause of pain over and above the information gained by history, gynaecological examination and pelvic ultrasound (before performing diagnostic laparoscopy) in identifying deep-infiltrating endometriosis ($p = 0.006$) and endometrioma ($p = 0.02$), but not for other gynaecological structural causes or for identifying idiopathic CPP ($p = 0.08$).

Laparoscopy was significantly more accurate than MRI in diagnosing idiopathic CPP ($p < 0.0001$), superficial peritoneal endometriosis ($p < 0.0001$), deep-infiltrating endometriosis ($p < 0.0001$), and endometrioma of the ovary ($p = 0.02$) as the cause of pelvic pain. The accuracy of laparoscopy appeared able to rule in these diagnoses. Laparoscopy was not able to make accurate diagnoses that adhesions were the cause of pelvic pain, nor to diagnose adenomyosis.

Using MRI to identify women who require therapeutic laparoscopy would lead to 369 women in a cohort of 1000 women receiving laparoscopy unnecessarily, and 136 women who required laparoscopy not receiving it. A strategy of all women receiving laparoscopy would lead to 557 women who do not need laparoscopic treatment receiving it.

Main findings of the economic evaluation

Feeding these diagnostic performance data into a decision-analytic model to inform cost-effectiveness demonstrated that, for the majority of cut-off values used in the analysis, laparoscopy was found to be more effective, in terms of QALYs gained, than MRI. Laparoscopy was found to be increasingly cost-effective for longer time horizons. Extensive sensitivity analysis shows that, compared with laparoscopy, MRI is never a cost-effective option.

Conclusions

With respect to specific gynaecological conditions, MRI was accurate for endometrioma, adhesions, ovarian cysts and fibroids when compared with observations at laparoscopy. When compared with assignment of a cause for CPP symptoms by an expert panel, MRI was accurate for endometrioma, adhesions and fibroids. For determining the cause of pain, MRI added value in the care pathway over and above the information gained by history, gynaecological examination and pelvic ultrasound (before performing diagnostic laparoscopy) for the target conditions of deep-infiltrating endometriosis and endometrioma. The evaluation of non-gynaecological conditions suffered from a lack of agreement over verification of diagnosis by an expert panel, and this limits the useful interpretation of the data.

As a result of the extensive assumptions that were required to complete the economic analysis, the findings of the economic evaluation are insufficient on their own to inform practice. This study has shown that diagnostic laparoscopy is more cost-effective than MRI, although the importance of extending the time horizon and the cut-off point for the sensitivity and specificity of the tests has been demonstrated.

Strengths and limitations

Construction of an appropriate study designed to produce valid, reliable and generalisable results to guide decision-making proved challenging from the outset. The challenges included, among others, the variation in gynaecological practice, the multiplicity of target conditions with respect to the cause of CPP, the incongruence between observed conditions and attribution of the cause to CPP symptoms, the lack of consensus on radiology standards and the reporting of pelvic MRI for benign gynaecological conditions, the need to blind MRI results to avoid bias in the assessment of their value and the lack of objective reference standards for the various target conditions.

There is little evidence as to how diagnostic consensus panels should be operationalised. We piloted and then used a standardised approach to minimise variation in delineation of the reference standard, involving independent gynaecologists who provided the reference diagnosis for each case presented based on the consensus achieved through reviewing the structured summaries or full accounts of all the data collected. Understanding the difference between the presence of a condition and a target condition being the cause of pain proved to be the main challenge. During the course of the panel assessments, we undertook formal repeatability testing and found that the inter-rater agreement for assignment of gynaecological conditions as causes of CPP was good or very good, though it was poor or moderate for non-gynaecological conditions.

The strength of this model-based economic analysis is that it has utilised contemporary data from the MEDAL study and considered the cut-off point on the ROC curve to inform the optimum sensitivity and specificity for each test. However, it was limited to a 6-month time horizon for the data collection and does not capture the full long-term effects of successful testing and treatment for women with CPP. This means that the full patient benefit of laparoscopy and MRI has perhaps not been considered.

Implications for health care

Magnetic resonance imaging was dominated by laparoscopy in differential diagnosis of women presenting to gynaecology clinics with CPP. It did not add value to information, already gained from history, examination and ultrasound, about idiopathic CPP and various gynaecological conditions before considering a laparoscopy. To realise the little value that MRI did show for a few conditions, such as endometrioma and deep-infiltrating endometriosis, a value judgement about cost-effectiveness would have to be made, as the diagnostic gain from performing a MRI scan comes tagged with additional costs.

Recommendations for further research

The evaluation of tests for differential diagnosis in care pathways targeting various conditions for treatment requires methodological development (e.g. in areas of consensus science with respect to panel member numbers, sample size of cases for reliability). For CPP, the development of diagnostic prediction models incorporating biomarkers into baseline data may improve the use of diagnostic laparoscopy, but research will first be required to determine objective diagnostic criteria for the various target conditions and valid biomarkers for endometriosis.

Trial registration

This trial is registered as ISRCTN13028601.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/22/50. The contractual start date was in October 2011. The draft report began editorial review in August 2016 and was accepted for publication in November 2017. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).