HEALTH TECHNOLOGY ASSESSMENT

VOLUME 22 ISSUE 44 AUGUST 2018
ISSN 1366-5278

Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis

Marie Westwood, Bram Ramaekers, Shona Lang, Sabine Grimm, Sohan Deshpande, Shelley de Kock, Nigel Armstrong, Manuela Joore and Jos Kleijnen

National Institute for Health Research

DOI 10.3310/hta22440

Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis

Marie Westwood,[1]* Bram Ramaekers,[2] Shona Lang,[1] Sabine Grimm,[2] Sohan Deshpande,[1] Shelley de Kock,[1] Nigel Armstrong,[1] Manuela Joore[2] and Jos Kleijnen[3]

[1]Kleijnen Systematic Reviews Ltd, York, UK
[2]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre and CAPHRI, School for Public Health and Primary Care, Department of Health Services Research, Maastricht University, the Netherlands
[3]School for Public Health and Primary Care, Care and Public Health Research Institute (CAPHRI), Maastricht University, the Netherlands

*Corresponding author

Declared competing interests of authors: none

Published August 2018
DOI: 10.3310/hta22440

This report should be referenced as follows:

Westwood M, Ramaekers B, Lang S, Grimm S, Deshpande S, de Kock S, *et al*. Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2018;**22**(44).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

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 16/30/02. The contractual start date was in December 2016. The draft report began editorial review in May 2017 and was accepted for publication in October 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.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

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

Abstract

Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis

Marie Westwood,[1]* Bram Ramaekers,[2] Shona Lang,[1] Sabine Grimm,[2] Sohan Deshpande,[1] Shelley de Kock,[1] Nigel Armstrong,[1] Manuela Joore[2] and Jos Kleijnen[3]

[1]Kleijnen Systematic Reviews Ltd, York, UK
[2]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre and CAPHRI, School for Public Health and Primary Care, Department of Health Services Research, Maastricht University, the Netherlands
[3]School for Public Health and Primary Care, Care and Public Health Research Institute (CAPHRI), Maastricht University, the Netherlands

*Corresponding author marie@systematic-reviews.com

Background: Ovarian cancer is the sixth most common cancer in UK women and can be difficult to diagnose, particularly in the early stages. Risk-scoring can help to guide referral to specialist centres.

Objectives: To assess the clinical and cost-effectiveness of risk scores to guide referral decisions for women with suspected ovarian cancer in secondary care.

Methods: Twenty-one databases, including MEDLINE and EMBASE, were searched from inception to November 2016. Review methods followed published guidelines. The meta-analysis using weighted averages and random-effects modelling was used to estimate summary sensitivity and specificity with 95% confidence intervals (CIs). The cost-effectiveness analysis considered the long-term costs and quality-adjusted life-years (QALYs) associated with different risk-scoring methods, and subsequent care pathways. Modelling comprised a decision tree and a Markov model. The decision tree was used to model short-term outcomes and the Markov model was used to estimate the long-term costs and QALYs associated with treatment and progression.

Results: Fifty-one diagnostic cohort studies were included in the systematic review. The Risk of Ovarian Malignancy Algorithm (ROMA) score did not offer any advantage over the Risk of Malignancy Index 1 (RMI 1). Patients with borderline tumours or non-ovarian primaries appeared to account for disproportionately high numbers of false-negative, low-risk ROMA scores. (Confidential information has been removed.) To achieve similar levels of sensitivity to the Assessment of Different NEoplasias in the adneXa (ADNEX) model and the International Ovarian Tumour Analysis (IOTA) group's simple ultrasound rules, a very low RMI 1 decision threshold (25) would be needed; the summary sensitivity and specificity estimates for the RMI 1 at this threshold were 94.9% (95% CI 91.5% to 97.2%) and 51.1% (95% CI 47.0% to 55.2%), respectively. In the base-case analysis, RMI 1 (threshold of 250) was the least effective [16.926 life-years (LYs), 13.820 QALYs] and the second cheapest (£5669). The IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) were the cheapest (£5667) and the second most effective [16.954 LYs, 13.841 QALYs], dominating RMI 1. The ADNEX model (threshold of 10%), costing £5699, was the most effective (16.957 LYs, 13.843 QALYs), and compared with the IOTA group's simple ultrasound rules, resulted in an incremental cost-effectiveness ratio of £15,304

per QALY gained. At thresholds of up to £15,304 per QALY gained, the IOTA group's simple ultrasound rules are cost-effective; the ADNEX model (threshold of 10%) is cost-effective for higher thresholds.

Limitations: Information on the downstream clinical consequences of risk-scoring was limited.

Conclusions: Both the ADNEX model and the IOTA group's simple ultrasound rules may offer increased sensitivity relative to current practice (RMI 1); that is, more women with malignant tumours would be referred to a specialist multidisciplinary team, although more women with benign tumours would also be referred. The cost-effectiveness model supports prioritisation of sensitivity over specificity. Further research is needed on the clinical consequences of risk-scoring.

Study registration: This study is registered as PROSPERO CRD42016053326.

Funding details: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables

List of figures

List of boxes

Glossary

Cost-effectiveness analysis An economic analysis that converts effects into health terms and describes the costs for additional health gain.

Decision modelling A mathematical construct that allows the comparison of the relationship between costs and outcomes of alternative health-care interventions.

False negative An incorrect negative test result – the number of diseased persons with a negative test result.

False positive An incorrect positive test result – the number of non-diseased persons with a positive test result.

Incremental cost-effectiveness ratio The difference in the mean costs of two interventions in the population of interest divided by the difference in the mean outcomes in the population of interest.

Index test The test of which the performance is being evaluated.

Markov model An analytic method particularly suited to modelling repeated events or the progression of a chronic disease over time.

Meta-analysis A statistical technique used to combine the results of two or more studies and obtain a combined estimate of effect.

Metaregression A statistical technique used to explore the relationship between the study characteristics and the study results.

Negative predictive value The probability of non-disease among persons with a negative test result.

Opportunity cost The cost of forgone outcomes that could have been achieved through alternative investments.

Positive predictive value The probability of disease among persons with a positive test result.

Probabilistic sensitivity analysis A method of quantifying the uncertainty in a mathematical model, such as a cost-effectiveness model.

Publication bias The bias arising from the preferential publication of studies with statistically significant results.

Regression analysis A statistical method for estimating relationships among variables.

Quality-adjusted life-year A measure of health gain, used in economic evaluations, in which survival duration is weighted or adjusted by the patient's quality of life during the survival period.

Quality of life An individual's emotional, social and physical well-being, and their ability to perform the ordinary tasks of living.

Receiver operating characteristic curve A graph that illustrates the trade-offs between sensitivity and specificity, which result from varying the diagnostic threshold.

Reference standard The best currently available method for diagnosing the target condition. The index test is compared against this to allow for the calculation of estimates of accuracy.

Sensitivity The proportion of people with the target disorder who have a positive test result.

Specificity The proportion of people without the target disorder who have a negative test result.

True negative A correct negative test result – the number of non-diseased persons with a negative test result.

True positive A correct positive test result – the number of diseased persons with a positive test result.

List of abbreviations

ACOG	American Congress of Obstetricians and Gynecologists
ADNEX	Assessment of Different NEoplasias in the adneXa
AFP	alpha-fetoprotein
AOC	advanced-stage ovarian cancer
apo A-1	apolipoprotein A-1
beta-hCG	beta-human chorionic gonadotrophin
CA125	cancer antigen 125
CCT	controlled clinical trial
CE	Conformité Européenne
CEIA	chemiluminescent enzyme immunoassay
CI	confidence interval
CRC	colorectal cancer
CRD	Centre for Reviews and Dissemination
CT	computed tomography
DAR	diagnostic appraisal review
EFSUMB	European Federation of Societies for Ultrasound in Medicine and Biology
EIA	enzyme immunoassay
FIGO	International Federation of Gynecology and Obstetrics
FN	false negative
FP	false positive
FSH	follicle-stimulating hormone
GP	general practitioner
HE4	human epididymis protein 4
HR	hazard ratio
HTA	Health Technology Assessment
IBS	irritable bowel syndrome
ICER	incremental cost-effectiveness ratio
ICON	International Collaborative Ovarian Neoplasm
IOTA	International Ovarian Tumour Analysis
IU	International Unit
LY	life-year
MDT	multidisciplinary team
MIA	multivariate index assay
MIA2G	multivariate index assay, second generation
MMS	multimodal screening
MRI	magnetic resonance imaging
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
PET	positron emission tomography
PET-CT	positron emission tomography computed tomography
PI	predictive index
PROBAST	Prediction model study Risk Of Bias Assessment Tool
PSA	probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
QUADAS-2	quality assessment of diagnostic accuracy studies 2
RCOG	Royal College of Obstetrics and Gynaecology
RCT	randomised controlled trial
RMI 1	Risk of Malignancy Index 1
ROCkeTS	Refining Ovarian Cancer Test accuracy Scores
ROMA	Risk of Ovarian Malignancy Algorithm
SIGN	Scottish Intercollegiate Guidelines Network
SMDT	specialist multidisciplinary team
TA	Technology Appraisal

TN	true negative	TVS	transvaginal sonography
TP	true positive	USS	ultrasound screening
TRF	transferrin		

Note

This monograph is based on the Technology Assessment Report produced for the National Institute for Health and Care Excellence (NICE). The full report contained a considerable number of data that were deemed confidential. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of confidential data removed and replaced by the statement 'confidential information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain English summary

Ovarian cancer is the sixth most common cancer in UK women and is more likely to be treated successfully if found early and treated by specialist teams. However, early-stage ovarian cancer can be difficult to diagnose. Symptoms, such as feeling bloated, feeling full early or having a poor appetite, abdominal or pelvic pain, and needing to urinate more often or more urgently can be early warning signs of ovarian cancer, but can also be caused by other conditions (e.g. fibroids, endometriosis and infections).

It is important to find tests that can predict which women are more likely to have ovarian cancer so that they can be referred to a specialist centre as quickly as possible.

This assessment considered how best to combine information from blood tests, ultrasound and clinical examinations (signs and symptoms reported by the patient and menopausal status), in order to decide when a woman is more likely to have ovarian cancer and should therefore be referred to a specialist centre for further investigations (including biopsy or surgery) and treatment.

A total of 51 studies of a variety of tools used to predict ovarian cancer in women who had a mass that was visible on ultrasound were included in the study. Two tools, one based on features seen by ultrasound (the International Ovarian Tumour Analysis simple ultrasound rules) and one that combined morphological features seen on ultrasound, a tumour marker and clinical information [the Assessment of Different NEoplasias in the adneXa (ADNEX) model], identified a higher proportion of those women with cancer than the method that is currently recommended [the Risk of Malignancy Index (RMI 1)]. This means that if the RMI 1 were replaced by either of these tools, more women with ovarian cancer would be referred to a specialist centre; however, more women with benign (non-cancerous) lumps would also be referred.

Health economic analyses indicated that the ADNEX model (threshold 10%), may be cost-effective compared with alternative tools to predict ovarian cancer.

Scientific summary

Background

Current guidance [National Collaborating Centre for Cancer. *Ovarian Cancer: The Recognition and Initial Management of Ovarian Cancer*. Clinical guideline (CG122). Manchester: National Institute for Health and Care Excellence; 2011] recommends that serum cancer antigen 125 (CA125) levels should be measured in secondary care, in all people with suspected ovarian cancer for whom serum CA125 levels have not already been measured in primary care. CG122 specifically recommends the calculation of a Risk Malignancy Index 1 (RMI 1) score, which includes CA125 levels, morphological features seen on ultrasound and menopausal status, with referral to a specialist multidisciplinary team (SMDT) for people with a RMI 1 score of ≥ 250. An evaluation of current evidence is needed to assess the clinical utility and cost-effectiveness of alternative methods of risk-scoring.

Objectives

The overall objective of this assessment was to summarise the evidence on the clinical effectiveness and cost-effectiveness of using alternative risk scores that include CA125 levels, human epididymis protein 4 (HE4) levels or morphological features seen on ultrasound {Risk of Ovarian Malignancy Algorithm [ROMA], International Ovarian Tumour Analysis [IOTA] group's simple ultrasound rules the IOTA group's Assessment of Different NEoplasias in the adneXa [ADNEX] model, Overa [multivariate index assay, second generation (MIA2G)], and RMI 1 at thresholds other than 250} to guide referral decisions for women with suspected ovarian cancer in secondary care. The following research questions were defined:

- What is the accuracy of alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 and CA125 levels and morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), in which the target condition is histologically confirmed ovarian cancer?
- What are the effects of using alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 and CA125 levels and morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), on clinical management decisions and clinical outcomes?
- What is the cost-effectiveness of alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 and CA125 levels and morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), when routinely used in secondary care to guide decisions about referral to a SMDT for women with suspected ovarian cancer?

Methods

Assessment of clinical effectiveness

Twenty-one databases, including MEDLINE and EMBASE, research registers and conference proceedings, were searched from inception to November 2016. Search results were screened for relevance independently by two reviewers. A full-text inclusion assessment, data extraction and a quality assessment were conducted by one reviewer and checked by a second. Study quality was assessed using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) tool and PROBAST (Prediction model study Risk Of Bias Assessment Tool). A meta-analysis using weighted averages and random-effects modelling was used to estimate summary sensitivity and specificity with 95% confidence intervals (CIs). Analyses were conducted separately for each assay, threshold and target condition (all malignancy, ovarian cancer and borderline cancer) for which data were available.

Assessment of cost-effectiveness

The base-case analysis included seven risk scores:

1. Risk of Malignancy Index RMI 1 score (at a threshold of 250)
2. Risk of Ovarian Malignancy Algorithm score using Abbott Diagnostics' ARCHITECT CA125 and HE4 assays (Abbott Diagnostics, Abbott Park, IL, USA)
3. Risk of Ovarian Malignancy Algorithm using Roche Diagnostics' Elecsys CA125 and HE4 assays (Roche Diagnostics, Rotkreuz, Switzerland)
4. Overa (MIA2G; Vermillion, Inc., Austin, TX, USA; at a threshold of 5 units)
5. International Ovarian Tumour Analysis Simple Rules (inconclusive, assumed to be malignant)
6. International Ovarian Tumour Analysis ADNEX model (at a threshold of 10%)
7. Risk of Malignancy Index (threshold of 200).

This assessment used the economic model from CG122 as a starting point to develop a de novo model adapted to better fit the scope of the current assessment; consistent with the CG122 model, the population age was assumed to be 40 years.

In the de novo health economic model, the mean expected costs and quality-adjusted life-years (QALYs) were calculated for each alternative risk assessment strategy. These long-term consequences were estimated based on the accuracy of the different strategies to detect ovarian cancer, followed by referral to a SMDT and treatment in tertiary care, or no tertiary referral. It was also taken into account that a small proportion of patients with pelvic masses are diagnosed with colorectal cancer (consistent with CG122).

A decision tree and a Markov model were developed. The decision tree was used to model the short-term outcomes. It was assumed that patients who are found to have a high risk of malignancy [i.e. who receive a high-risk test result (either true or false positive)] are referred to a SMDT, and patients who receive a low-risk test result (either true or false negative) are not referred to a SMDT.

Results

Assessment of clinical effectiveness

Fifty-one diagnostic cohort studies (65 publications and one unpublished interim report) were included in the systematic review. Sixteen studies were identified for the ROMA score, 18 for the IOTA group's simple ultrasound rules, seven for the IOTA group's ADNEX model, three for Overa (MIA2G) and 10 for different thresholds of the RMI 1; some studies assessed more than one risk score. The main potential sources of bias in the included studies related to patient flow (not all patients were included in the analysis) and the applicability of the index text (test performed before referral, retrospective application of variables, use of experienced ultrasound practitioners and risk score-specific pre-study training).

The ROMA score, using the Abbott Diagnostics' ARCHITECT or Roche Diagnostics' Elecsys tumour marker assays, did not offer any clear performance advantage over the RMI 1. The only ROMA score study ($n = 213$ participants) using the Abbott Diagnostics ARCHITECT assay, which included all participants in the analysis, reported similar sensitivity and specificity estimates for the ROMA score and the RMI 1 at a decision threshold of 200, 75% (95% CI 60.4% to 86.4%) versus 77.1% (95% CI 62.7% to 88.0%), and 87.9% (95% CI 81.9% to 92.4%) versus 81.8% (95% CI 75.1% to 87.4%), respectively. By contrast, when participants with borderline tumours and/or those with malignancies other than epithelial ovarian cancer were excluded from the analyses (two studies, $n = 1172$ participants), the summary specificity estimate for the ROMA score (53.3%, 95% CI 50.0% to 56.7%) was significantly lower than that for the RMI 1 score at a decision threshold of 200 (80.3%, 95% CI 77.5% to 82.9%), and the summary sensitivity estimates were similar and higher, at 96.4% (95% CI 93.6% to 98.2%) and 93.4% (95% CI 90.0% to 95.9%). The only study to report a direct comparison of the ROMA score, using Roche Diagnostics' Elecsys tumour marker assays and the RMI 1 score at a decision threshold of 200, included all study participants in the analysis,

irrespective of final histological diagnosis, but classified participants with borderline tumours as disease negative. In this study, the sensitivity estimate for the ROMA score appeared to be slightly higher than that for the RMI 1 score, at 83.8% (95% CI 73.4% to 91.3%) versus 78.4% (95% CI 67.3% to 87.1%), respectively, and the specificity estimate for the ROMA score appeared to be slightly lower than that for the RMI 1 score, at 68.8% (95% CI 61.6% to 75.4%) versus 79.6% (95% CI 73.1% to 85.1%), respectively, but neither difference was statistically significant. The summary estimates of sensitivity and specificity for the ROMA score, using Roche Diagnostics' Elecsys tumour marker assays at the manufacturer's recommended thresholds, were derived from non-comparative accuracy studies in which all participants were included in the analysis (with the target condition being all malignancy) were 79.1% (95% CI 74.2% to 83.5%) and 79.1% (95% CI 76.3% to 81.6%), respectively (two studies, $n = 1252$ participants). In studies in which the manufacturer's recommended cut-off points were used, the performance of the ROMA score did not differ significantly between premenopausal women and postmenopausal women. Limited data indicated that patients with borderline tumours and those with non-ovarian primaries accounted for disproportionately high numbers of those with false-negative, low-risk ROMA scores. There were no studies evaluating the ROMA score using CA125 and HE4 assays on the Fujirebio Diagnostics' LUMIPULSE® G automated chemiluminescent enzyme immunoassay system (Fujirebio Diagnostics, Gothenburg, Sweden).

The summary estimates of sensitivity, derived from direct comparison studies that included all study participants in their analyses [two studies, $n =$ (confidential information has been removed)], were significantly higher for both the ADNEX model, at 96% (95% CI 94.5% to 97.1%), and IOTA group's simple ultrasound rules, at 92.8% (95% CI 90.9% to 94.3%), than for the RMI 1 score at a decision threshold of 200: 66% (95% CI 62.9% to 69%) (confidential information has been removed). Conversely, the summary estimates of specificity, for both the ADNEX model, at 67% (95% CI 64.2% to 69.6%), and the IOTA group's simple ultrasound rules, at 71.6% (95% CI 68.9% to 74.1%), were significantly lower than those for the RMI 1 score at a decision threshold of 200: 89% (95% CI 87% to 90.7%) (confidential information has been removed). In order to achieve similar levels of sensitivity to those provided by the ADNEX model and the IOTA group's simple ultrasound rules, a very low RMI 1 score decision threshold (25) would be needed; the summary sensitivity and specificity estimates for the RMI 1 score at this threshold were 94.9% (95% CI 91.5% to 97.2%) and 51.1 (95% CI 47.0% and 55.2%), respectively.

No studies were identified that directly compared Overa (MIA2G) to the RMI 1.

Studies evaluating the RMI 1 score at different thresholds indicated no significant difference in performance between thresholds of 200 and 250.

Assessment of cost-effectiveness

In the base-case analysis, the RMI 1 with a threshold of 250 was the least effective [16.926 life-years (LYs), 13.820 QALYs] and the second cheapest (£5669). The IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant), was the cheapest (£5667) and the second most effective (16.954 LYs, 13.841 QALYs), and thereby dominated the RMI 1 (at both the 200 and 250 thresholds). The IOTA group's ADNEX model (threshold of 10%), with a cost of £5699, was the most effective (16.957 LYs, 13.843 QALYs), and compared with the IOTA group's simple ultrasound rules, resulted in an incremental cost-effectiveness ratio of £15,304 per QALY gained. The remaining risk scores [ROMA using Abbott Diagnostics' ARCHITECT, ROMA using Roche Diagnostics' Elecsys and Overa (MIA2G) by Vermillion] were dominated. As a result, the incremental analysis indicated that, up to thresholds of £15,304 per QALY gained, the IOTA group's simple ultrasound rules are cost-effective, whereas the IOTA group's ADNEX model (threshold of 10%) is cost-effective for higher thresholds. Consequently, at willingness-to-pay thresholds of both £20,000 and £30,000 per QALY, the RMI 1, at a threshold of 250, had a probability of being cost-effective of 1%. For the IOTA group's simple ultrasound rules and the IOTA group's ADNEX model (threshold of 10%), this was 39% and 60%, respectively (at the £20,000 threshold), and 23% and 75%, respectively (at the £30,000 threshold). The probabilities for the other risk scores were < 1% for these thresholds.

The sensitivity and scenario analyses indicated that the hazard ratio for SMDT referral versus no SMDT referral (for patients with ovarian cancer) was the most influential parameter in the model, and that the results were reasonably robust. Most scenario analyses indicated that at thresholds of £20,000 and £30,000 per QALY gained, the IOTA group's ADNEX model (threshold of 10%) remained the cost-effective strategy. In two scenario analyses, the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) was considered to be cost-effective at a threshold of £20,000 and/or £30,000 per QALY gained. For the scenario comparing the optimal sensitivity RMI 1 threshold, which was found to be 25 (at all thresholds of £2890 per QALY gained or higher), the RMI 1 was still dominated.

For the premenopausal and postmenopausal subgroups, the IOTA group's ADNEX model (threshold of 10%) remained cost-effective at thresholds of £20,000 and £30,000 per QALY gained.

Conclusions

Implications for service provision

There is evidence to suggest that using either the ADNEX model or the IOTA group's simple ultrasound rules to assess the risk of malignancy in women with an adnexal mass may offer increased sensitivity relative to current practice (the RMI 1 at a decision threshold of 250 or 200); that is, a higher proportion of those women who have a malignant tumour would be referred to a SMDT. A similar sensitivity could be achieved with the RMI 1 by using a very low decision threshold (25); however, this is associated with a lower specificity and a greater number of unnecessary referrals than those achievable using either the ADNEX model or the IOTA group's simple ultrasound rules. The limited available evidence suggests that the ROMA score does not offer any clear performance advantage over the RMI 1. Although Overa (MIA2G) appears to have higher sensitivity than the ROMA score, there are no data to support a direct comparison between Overa (MIA2G) and the RMI 1.

Overall, the cost-effectiveness model provides evidence to strongly prioritise sensitivity over specificity. As a result, the IOTA group's ADNEX model (threshold of 10%), which had the highest sensitivity (96.3%), was considered to be cost-effective.

Suggested research priorities

Further studies or analyses of the IOTA data set are needed to understand the role of menopausal status and other potentially relevant factors, such as family history of ovarian cancer, in the performance of both the IOTA and the ADNEX tests. Large diagnostic cohort studies are needed to fully evaluate the performance of the ROMA score (using different manufacturers' tumour marker assays) and of Overa (MIA2G), compared with the RMI 1, at a decision threshold of 250 or 200. These studies should be conducted in a population that includes the full spectrum of differential diagnoses likely to be present in those referred to secondary care for the investigation of an adnexal mass. Further studies are required to explore the distribution of histological diagnoses among patients with false-negative, low-risk classifications. A more complete exploration of the types of patients who are likely to be misclassified as being at a low risk of having ovarian cancer using the various risk-scoring options available, as well as an investigation of the downstream clinical consequences for these patients, is required.

Study registration

This study is registered as PROSPERO CRD42016053326.

Funding

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

Chapter 1 Objective

The overall objective of this assessment was to summarise the evidence on the clinical effectiveness and cost-effectiveness of using alternative risk scores, which includes measuring the levels of cancer antigen 125 (CA125) and human epididymis protein 4 (HE4) or morphological features seen on ultrasound (detailed in *Chapter 2*, *Intervention technologies*), to guide referral decisions for people with suspected ovarian cancer in secondary care. The current National Institute for Health and Care Excellence (NICE) guidance (CG122)[1] recommends that the levels of serum CA125 should be measured in secondary care, in all people with suspected ovarian cancer for whom serum CA125 levels have not already been measured in primary care. CA125 levels are a component of secondary care investigation and are not used in isolation; the CG122 specifically recommends the calculation of a Risk of Malignancy Index 1 (RMI 1) score, which includes the measurement of CA125 levels, and referral to a specialist gynaecological oncology multidisciplinary team (MDT) for people with a RMI 1 score of ≥ 250. The CG122 does not currently include any recommendations on HE4 testing or alternative methods of risk-scoring. An evaluation of current evidence was needed to assess the clinical utility and cost-effectiveness of alternative methods of risk-scoring. The following research questions were defined to address the objectives of this assessment:

- What are the performance characteristics of alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 or CA125 levels or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), for which the target condition is histologically confirmed ovarian cancer?
- What are the effects of using alternative risk scores (including alternative RMI 1 score thresholds), which include measuring HE4 or CA125 levels or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), on clinical management decisions and clinical outcomes?
- What is the cost-effectiveness [incremental cost per quality-adjusted life-year (QALY)] of alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 or CA125 levels or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), when routinely used, in secondary care, to guide decisions about referral to a specialist multidisciplinary team (SMDT), for women with suspected ovarian cancer?

Chapter 2 Background and definition of the decision problem(s)

Population

The primary indication for this assessment was optimisation of the routine secondary care assessment of women with suspected ovarian cancer, to decide whether or not a patient should be referred to a SMDT. The assessment was conducted in the context of an update to the current guidance (CG122).[1] The relevant population was women of any age, including premenopausal and postmenopausal women, who had been referred to secondary care for the investigation of suspected ovarian cancer. This assessment includes data from women of any age, but no cost-effectiveness modelling was undertaken for the population aged < 18 years owing to a lack of data on the performance of risk scores in this age group. Women with a previous history of ovarian cancer who were being monitored for possible recurrence, and those referred directly from primary care to a SMDT, were outside the scope of this assessment.

Target condition

The target condition for this assessment was ovarian cancer. Ovarian cancer is a term describing a group of cancers arising from cells in or near the ovaries. Ovarian cancers can be classified based on tissue type (epithelial ovarian tumours, sex cord–stromal tumours and germ cell tumours), with epithelial carcinomas being the most common type (90%) of primary ovarian cancers; non-epithelial ovarian cancers are more common in premenopausal women.[2] The target conditions covered by the CG122 were epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma and borderline ovarian cancer;[1] excluded target conditions were pseudomyxoma peritonei, relapsed ovarian, fallopian tube or peritoneal cancer, germ cell tumour of the ovary and sex cord–stromal tumours of the ovary. This assessment was not limited to any particular type of ovarian cancer.

Ovarian cancers are staged using the four-stage International Federation of Gynecology and Obstetrics (FIGO) system:[3]

1. stage I – confined to the organ of origin (ovaries or fallopian tubes)
2. stage II – invasion of the surrounding organs or tissues [pelvic extension or primary peritoneal cancer (below the pelvic brim)]
3. stage III – spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes
4. stage IV – distant metastases, excluding peritoneal (e.g. lungs, liver, spleen).

Ovarian cancer can also be graded based on how differentiated cells appear:

- grade 1 – well differentiated
- grade 2 – moderately differentiated
- grade 3 – poorly differentiated/undifferentiated.

Ovarian cancer is the sixth most common cancer in women in the UK (as of 2013), accounting for 4% of all new cases of cancer in females.[4,5] In 2013, there were 7284 new cases of ovarian cancer in women in the UK, giving an age-standardised incidence rate of 23.3 per 100,000.[4,5] Ovarian cancer accounts for around 5% of cancer deaths in women in the UK; 2014 statistics recorded 4100 ovarian cancer deaths.[6] The incidence of ovarian cancer is strongly related to age, with 2011–13 data indicating

that approximately half (53%) of new cases of ovarian cancer were diagnosed in women > 65 years of age.[4,5] Ovarian cancer mortality is also strongly related to age at diagnosis.[6]

Data from the Office for National Statistics, published by Cancer Research UK,[7] indicate that, although ovarian cancer incidence rates have increased overall since the 1970s, the UK age-standardised incidence rates decreased by 6% in the decade between 2002–4 and 2011–13. However, it remains the case that a high proportion of women (58%) are diagnosed at an advanced stage (stage III or IV), and 21% have metastases at diagnosis.[8] Ovarian cancer survival is strongly related to stage at diagnosis; 2012 data[6] showed that the 1-year and 5-year survival rates for women diagnosed at stage I were 97% and 90%, respectively, versus 53% and 4%, respectively, for women diagnosed at stage IV. Improving early diagnosis is therefore a priority, and variation in the performance of testing strategies for the detection of different stages of ovarian cancer should be considered. The majority of studies about ovarian cancer diagnosis concern epithelial carcinomas; however, there is some evidence to indicate that the diagnostic performance of tumour markers and risk scores may vary between tumours of different tissue types;[9] the possible effects of tumour tissue type on estimates of test performance should also be considered.

It has been suggested that CA125 results should be interpreted cautiously in premenopausal women because of the high rate of false-positive (FP) diagnoses resulting from various non-malignant conditions (e.g. fibroids, endometriosis, adenomyosis, pelvic infection).[10] It is therefore important to consider the effects of the menopausal status of women on the performance of testing strategies, either by stratification of data from test accuracy studies or by including menopausal status in risk models (as in the RMI 1).

Intervention technologies

Serum tumour markers are used in the secondary care investigation of people with suspected ovarian cancer; these are not considered to be 'stand-alone' diagnostic tests, but are used in conjunction with other tests, signs and symptoms to assess the risk of malignancy. An estimate of an individual's risk of malignancy can inform decisions about specialist referral, further testing and treatment. It is anticipated that these risk assessment tools will be used in secondary care, for people in whom ultrasound imaging suggests confined disease or a low volume of disease outside the pelvis (stages I–IIIb).

An optimised risk assessment that reduces the number of women with ovarian cancer who are not referred for further specialist care [i.e. those with a 'false-negative' (FN) risk assessment] has the potential to improve prognosis, be cost-saving in terms of unnecessary further investigations and reduce associated anxiety. Prognosis may be adversely affected by a failure to refer women to a SMDT and specialist surgery. In particular, it is likely that women who are believed to have a benign explanation for any pelvic mass will be operated on in secondary care. If they actually have ovarian cancer, then the prognosis might be worse than if they had been operated on by a specialist gynaecological oncology surgeon. Indeed, there is evidence of up to a 45% difference in the median overall survival between a set of regional centres in the UK and the UK as a whole.[11]

The current standard assessment (RMI 1) has been reported as having poor sensitivity – approximately 63% at an operating threshold of 200.[12] If referral decisions are based on the RMI 1 score at this threshold, there remains the potential for significant numbers of people with ovarian cancer to remain unreferred, and to experience consequential delays in diagnosis and detrimental effects on prognosis. A systematic review of studies comparing HE4, CA125 and the Risk of Ovarian Malignancy Algorithm (ROMA) score reported similar overall sensitivity estimates for HE4 and CA125 (76% and 79%, respectively) and a higher sensitivity (85%) for the ROMA score.[9] Sensitivity estimates were lower for early-stage cancer (55% for both HE4 and CA125 levels, and 74% for the ROMA score).[9] Risk scores with higher sensitivity are needed to facilitate prompt referral of the appropriate patient group.

The Risk of Ovarian Malignancy Algorithm score

The ROMA score uses serum HE4 and serum CA125 levels, along with menopausal status, to generate an individualised estimate of the risk that a person has ovarian cancer. Initially, a predictive index (PI) value is calculated using a formula that differs depending on whether the woman is premenopausal or postmenopausal (*Equations 1* and *2* in *Box 1*). This PI value can then be used to calculate the ROMA score (*Equation 3* in *Box 1*).[13] The ROMA score is intended for use in women who present with an adnexal mass (i.e. following ultrasound examination). Manufacturers of HE4 assays recommend the use of these assays, in the context of a ROMA score, in combination with a specific CA125 assay or assays; if a CA125 level has been obtained in primary care, using a different assay, this will need to be repeated in secondary care before a ROMA score can be calculated.

Cut-off values for the ROMA scores are used to classify individuals as having a low or high risk of developing epithelial ovarian cancer. Recommended thresholds can differ depending on the tumour marker assays used, as described below.

There are currently three commercial HE4 assays for use with automated immunoassay analysers that are available for use in the UK NHS; a summary of the key technical characteristics of these assays is provided below (*Table 1*).

BOX 1 Risk of Ovarian Malignancy Algorithm equations

Premenopausal women

$$PI = -12.0 + 2.38 \times \ln(HE4) + 0.0626 \times \ln(CA125). \quad (1)$$

Postmenopausal women

$$PI = -8.09 + 1.04 \times \ln(HE4) + 0.732 \times \ln(CA125), \quad (2)$$

$$\text{ROMA score } (\%) = \exp(PI)/[1 + \exp(PI)] \times 100\%. \quad (3)$$

(CA125), serum concertation of CA125 in U/ml; (HE4), serum concentration of HE4 in pmol/l; ln, natural logarithm.

TABLE 1 Technical characteristics of serum HE4 assays available to the NHS

Name of assay (manufacturers' details)	Company	Detection		Assay time
		Limit	Range	
ARCHITECT HE4 (Abbott Diagnostics, Abbott Park, IL, USA)	Abbott Diagnostics	15 pmol/l	20–1500 pmol/l	28 minutes[a]
LUMIPULSE® G HE4 (Fujirebio Diagnostics, Gothenburg, Sweden)	Fujirebio Diagnostics	3.5 pmol/l	20–1500 pmol/l	35 minutes[b]
Elecsys® HE4 (Roche Diagnostics, Rotkreuz, Switzerland)	Roche Diagnostics	15 pmol/l	15–1500 pmol/l	18 minutes[c]

a Time is for how long it took the analyser to complete the sample analysis once initiated.
b Using the LUMIPULSE G1200 instrument; time shown is for one sample; time for all 42 results is 55 minutes.
c Report time is dependent on whether or not other tests are carried out on the same sample, but typically takes < 30 minutes.

The ARCHITECT human epididymis protein 4 assay (Abbott Diagnostics)

The ARCHITECT HE4 assay (Abbott Diagnostics, Abbott Park, IL, USA) is a chemiluminescent microparticle immunoassay for the quantitative determination of HE4 levels in human serum. The assay is designed for use on an immunoassay analyser, specifically the ARCHITECT i2000SR or the ARCHITECT i1000SR analysers. Additional materials required to run the assay are the ARCHITECT HE4 assay software file, ARCHITECT HE4 calibrators, ARCHITECT HE4 controls, ARCHITECT multiassay manual diluent, ARCHITECT pre-trigger solution, ARCHITECT trigger solution, ARCHITECT wash buffer, ARCHITECT reaction vessels, ARCHITECT sample cups, ARCHITECT septum and ARCHITECT replacement caps.

The results of the assay are intended to be used in conjunction with the ARCHITECT CA125 II assay, as an aid in estimating the risk of epithelial ovarian cancer in women presenting with a pelvic mass who will undergo surgical intervention. The company recommends that the HE4 assay results are used in the calculation of the ROMA scores, using the following cut-off values for ROMA scores, based on obtaining a specificity of 75%:

- in premenopausal patients, a ROMA value of $\geq 7.4\%$ indicates a high risk of finding epithelial ovarian cancer
- in premenopausal patients, a ROMA value of $< 7.4\%$ indicates a low risk of finding epithelial ovarian cancer
- in postmenopausal patients, a ROMA value of $\geq 25.3\%$ indicates a high risk of finding epithelial ovarian cancer
- in postmenopausal patients, a ROMA value of $< 25.3\%$ indicates a low risk of finding epithelial ovarian cancer.

These results must be interpreted in conjunction with other methods and clinical data (e.g. symptoms and medical history), in accordance with standard clinical management guidelines. The company states that additional testing should be done if the HE4 results are inconsistent with the clinical evidence.

LUMIPULSE G human epididymis protein 4 (Fujirebio Diagnostics)

The LUMIPULSE G HE4 (Fujirebio Diagnostics, Gothenburg, Sweden) is a chemiluminescent enzyme immunoassay (CEIA) for the quantitative measurement of HE4 levels in human serum. The assay is designed for use on the LUMIPULSE G system (either the LUMIPULSE G1200 or the LUMIPULSE G600 immunoassay analysers). Samples are run using immunoreaction cartridges, which contain reagents and into which samples are added. Further materials required for the assay are LUMIPULSE G HE4 calibrators, LUMIPULSE G substrate solution, LUMIPULSE G wash solution, LUMIPULSE G specimen diluent I, sampling tips for the LUMIPULSE system, soda lime for the LUMIPULSE system and LUMIPULSE G dilution cartridges.

The assay is intended for use in conjunction with CA125 levels (measured using the LUMIPULSE G CA125 II assay) as an aid in estimating the risk of epithelial ovarian cancer in premenopausal and postmenopausal women presenting with a pelvic mass who will undergo surgical intervention.

The company recommends that the HE4 results are used in the calculation of the ROMA scores, and suggests the following cut-off values for the ROMA scores, based on obtaining a specificity of 75%:

- in premenopausal patients, a ROMA value of $\geq 13.1\%$ indicates a high risk of finding epithelial ovarian cancer
- in premenopausal patients, a ROMA value of $< 13.1\%$ indicates a low risk of finding epithelial ovarian cancer
- in postmenopausal patients, a ROMA value of $\geq 27.7\%$ indicates a high risk of finding epithelial ovarian cancer
- in postmenopausal patients, a ROMA value of $< 27.7\%$ indicates a low risk of finding epithelial ovarian cancer.

Results should be interpreted in conjunction with further methods and clinical data (e.g. clinical findings, age, family history and imaging results), in accordance with standard clinical management guidelines.

A further HE4 assay is also available from Fujirebio Diagnostics: the HE4 enzyme immunoassay (EIA), a manual, enzyme immunometric assay for the quantitative determination of HE4 in human serum. Clinical experts commented that manual kits would be unlikely to be used in routine practice in the NHS; therefore, this assay has not been included in the scope of this assessment.

Elecsys human epididymis protein 4 immunoassay (Roche Diagnostics)

The Elecsys HE4 (Roche Diagnostics, Rotkreuz, Switzerland) is an immunoassay that uses Roche Diagnostics' electrochemiluminescence detection technology to quantify HE4 levels. The assay uses anti-HE4 mouse monoclonal antibodies to capture HE4 in a serum sample and label it with a ruthenium complex. The application of a voltage to the samples then induces chemiluminescent emissions, which are measured by a photomultiplier.

The assay is designed for use on an immunoassay analyser, specifically the following analysers: modular analytics E170, cobas e 411, cobas e 601/e 602 and cobas e 801. Additional materials required for the HE4 assay are the HE4 CalSet (Roche Diagnostics, Rotkreuz, Switzerland), PreciControl HE4 (Roche Diagnostics, Rotkreuz, Switzerland) and Diluent MultiAssay (Roche Diagnostics, Rotkreuz, Switzerland). Further materials are also required for the general running of analysers, such as wash and cleaning solutions and disposable consumables.

The assay is intended to be used in conjunction with the Elecsys CA125 II assay as an aid in estimating the risk of epithelial ovarian cancer in premenopausal and postmenopausal people with a pelvic mass. The company recommends that the HE4 and CA125 assay results are used in the calculation of the ROMA scores. The company suggests the following cut-off values for the ROMA scores, based on obtaining a specificity of 75%:

- in premenopausal patients, a ROMA value of $\geq 11.4\%$ indicates a high risk of finding epithelial ovarian cancer
- in premenopausal patients, a ROMA value of $< 11.4\%$ indicates a low risk of finding epithelial ovarian cancer
- in postmenopausal patients, a ROMA value of $\geq 29.9\%$ indicates a high risk of finding epithelial ovarian cancer
- in postmenopausal patients, a ROMA value of $< 29.9\%$ indicates a low risk of finding epithelial ovarian cancer.

The company states that additional testing should be done if the HE4 results are inconsistent with the clinical evidence.

Simple Rules ultrasound classification system (International Ovarian Tumour Analysis group)

Simple Rules is a morphological scoring system based on the presence of ultrasound features (described as rules) to characterise an ovarian mass as benign or malignant, and was developed by the International Ovarian Tumour Analysis (IOTA) group. The system uses a morphological scoring system based on the presence of ultrasound features to characterise an ovarian mass as benign or malignant, and requires the use of transvaginal sonography (TVS), which may be supplemented with abdominal ultrasound for larger masses. There are five 'rules' describing the features of malignant tumours (M-rules) and five rules that describe benign tumours [B-rules (see *Table 2*)].[14,15] Because the use of the Simple Rules system requires specialist training in interpreting real-time ultrasound images in relation to these rules, it is assumed that using the Simple Rules system in the specified population will require a secondary care ultrasound examination (i.e. a repeat examination in which the ultrasound has been conducted in primary care).

TABLE 2 Simple Rules ultrasound classification system (IOTA group)

M-rules (rules for predicting a malignant tumour)	B-rules (rules for predicting a benign tumour)
• Irregular solid tumour • Ascites present • Four or more papillary structures • Irregular multilocular solid tumour with a largest diameter of ≥ 100 mm • Very strong blood flow (colour score of 4)	• Unilocular • Solid components present, with largest solid component having a largest diameter of < 7 mm • Acoustic shadows present • Smooth multilocular tumour with a largest diameter of < 100 mm • No blood flow (colour score of 1)

The M-rules and B-rules can be combined to aid classification:

- If any M-rules (and no B-rules) apply, the mass is classified as malignant.
- If any B-rules (and no M-rules) apply, the mass is classified as benign.
- If both M-rule and B-rule (or neither) apply, the mass is unclassifiable, and the IOTA group states that there are then a number of options:
    - classify the mass as malignant
    - refer the patient to an expert ultrasound operator for a second opinion
    - use alternative imaging techniques
    - use the Simple Rules risk model[16] to calculate risk of malignancy using the morphological features seen on ultrasound described in the Simple Rules model.

No specific make or model of ultrasound device is required for the model inputs. A transvaginal probe is required, and the image must be of sufficient quality to allow the ultrasound features specified by the model to be seen. The IOTA group states that the approach to evaluating masses required by the classification system is not more time-consuming than a standard ultrasound scan.

The IOTA group organises 1-day courses that teach the techniques for classifying masses required by the system, with participants assessed by a multiple-choice test. An online training tool, which will be freely accessible to NHS practitioners, is also being developed. In addition to this training, the IOTA group also recommends that practitioners should have completed 300 gynaecological scans. Software is not required to run the Simple Rules model.

The Simple Rules model is not recommended for use with women who are pregnant. Physiological changes during pregnancy can alter the appearance of ovarian masses, which can affect the classification made using the Simple Rules model, and the model has not been validated in this group.

The Assessment of Different NEoplasias in the adneXa model (International Ovarian Tumour Analysis group)

The Assessment of Different NEoplasias in the adneXa (ADNEX) model was developed by the IOTA group to aid preoperative discrimination between benign, borderline, stage I invasive, stages II–IV invasive and secondary metastatic ovarian tumours in women with an ovarian (including paraovarian and tubal) mass.[17] The model uses nine predictors, three clinical variables [age, serum CA125 level and type of referral centre (oncology or other)] and six ultrasound variables (maximal lesion diameter, proportion of solid tissue, > 10 cyst locules, number of papillary projections, acoustic shadows and ascites). The IOTA group have produced iPhone, Android and web applications for calculating the ADNEX risk score (www.iotagroup.org/adnexmodel/). Guidance has also been published on the application of the ADNEX model in clinical practice, and the selection of risk cut-off values for risk stratification and choice of clinical management.[18] The IOTA group notes that, as with other diagnostic prediction models (other IOTA group models, ROMA scores, the RMI 1), the ADNEX model cannot be applied to women with conservatively treated adnexal tumours.

Overa (multivariate index assay, second generation)

The Overa [multivariate index assay, second generation (MIA2G); Vermillion, Inc., Austin, TX, USA] assay is a Conformité Européenne (CE)-marked qualitative serum test that combines the results of five immunoassays into a single numeric result [i.e. the Overa (MIA2G) risk score]. The five biomarkers included in the test are:

1. follicle-stimulating hormone (FSH)
2. HE4
3. apolipoprotein A-1 (apo A-1)
4. transferrin (TRF)
5. CA125.

The levels of these biomarkers present in serum are determined using immunoassays run on the Roche Diagnostics' cobas® 6000 system (Roche Diagnostics, Rotkreuz, Switzerland). The Overa (MIA2G) risk score is generated by the company's OvaCalc software, with the results ranging between 0.0 and 10.0. A risk score of < 5.0 is indicative of a low probability of malignancy and a score of ≥ 5.0 indicates a high probability of malignancy.

The assay is indicated for use in people > 18 years with a pelvic mass in whom surgery may be considered. It is intended for use as part of a preoperative assessment to help decide if a person presenting with a pelvic mass has a high risk or a low risk of ovarian malignancy.

The company states that the test results must be interpreted in conjunction with an independent clinical and imaging evaluation, and that the test is not intended for use in screening or as a stand-alone assay.

The Risk of Malignancy Index 1

The RMI 1, used at thresholds other than those currently recommended in the NICE clinical guidelines (see *Comparator*), was considered as an alternative intervention technology.

Comparator

The comparator for this assessment is the RMI 1, using the referral thresholds that best reflect current UK clinical practice (≥ 250), recommended in NICE clinical guideline CG122.[1] The RMI 1 score uses three components (measured serum CA125 levels, ultrasound imaging and menopausal status) to calculate a risk score:

$$\text{RMI 1 score} = U \times M \times \text{CA125}, \quad (4)$$

where:

- U is the ultrasound score – 1 point scored for the presence of each of the following features: multilocular cysts, solid areas, metastases, ascites, bilateral lesions. $U = 0$ (0 points), $U = 1$ (1 point) or $U = 3$ (2–5 points)
- M is the menopause score – $M = 1$ (premenopausal) or $M = 3$ (postmenopausal); a 'postmenopausal' woman is one who has had no period for more than 1 year or a woman aged > 50 years who has had a hysterectomy
- CA125 is the serum CA125 concentration – measured in international units (IU)/ml.

Notably, because the ultrasound score component of this equation is zero, if none of the specified features is present on an ultrasound scan, RMI 1 scores above zero are possible only if ultrasound scans identify features indicative of ovarian cancer.

The NICE clinical guideline CG122[1] recommends that people with a RMI 1 score of ≥ 250 should be referred to a specialist gynaecological oncology MDT. However, this guideline also includes a research recommendation that states that further research should be undertaken to determine the optimum RMI 1 threshold that should be applied in secondary care to guide the management of people with suspected ovarian cancer. The guideline notes that there was variation in the evidence base at that time with regard to the optimum RMI 1 threshold to use in secondary care, and that the value used will have implications for the management options considered, and the number of women who will be referred for specialist treatment.

The Scottish Intercollegiate Guidelines Network (SIGN)'s guideline on the management of epithelial ovarian cancer (SIGN 135)[19] recommends referring women with a RMI 1 score of > 200 to a gynaecological oncology MDT. In addition, the Royal College of Obstetrics and Gynaecology (RCOG)'s guideline on ovarian cysts in postmenopausal women recommends the use of 200 as a threshold to predict the likelihood of ovarian cancer, although it notes that the threshold of 250 is also acceptable; in the current literature,[10] a score of 200 is often used as a cut-off value.

Reference standard

Histopathology is the reference standard for assessing the accuracy of tests to identify people at a high risk of developing epithelial ovarian cancer. In addition to distinguishing between malignant and benign tumours, this testing can also determine the type of ovarian cancer present. Tissue samples used to confirm diagnosis can be obtained by biopsy or during surgery; however, for the population of interest (people in whom imaging suggests a confined disease or a low volume of disease outside the pelvis), it is expected that pre-surgery biopsy would not routinely occur. When tissue samples are not taken, clinical follow-up (ideally for a minimum of 12 months) may be required to determine the presence, or absence, of ovarian cancer.

Care pathway

Primary care assessment and criteria for referral to secondary care

The 2011 NICE clinical guideline CG122[1] provides recommendations about the assessment of people with suspected ovarian cancer in primary care.

These recommendations include information about signs and symptoms (e.g. abdominal bloating, feelings of satiety or loss of appetite, pelvic or abdominal pain, changes in bowel habit and urinary frequency/urge) as well as information about the use of CA125 testing.

More recent guidance about cancer diagnoses, the NICE guidance NG12,[20] published in 2015, reproduces the recommendations from the CG122[1] with no update.

The more recent (2013) guidance, from SIGN[19] provided recommendations covering similar topic areas.

Establishing a diagnosis in secondary care

The 2011 NICE clinical guideline CG122[1] also includes recommendations about testing following referral to secondary care. These recommendations cover the use of various blood tests [alpha-fetoprotein (AFP), beta-human chorionic gonadotrophin (beta-hCG) and CA125], risk scoring using the RMI 1 score, imaging (ultrasound and CT) and the role of the MDT.

Those secondary care recommendations that refer to CA125 consider its use in a clinical context, particularly in relation to the calculation of the RMI 1 score.[1]

The SIGN guideline (SIGN 135)[19] includes similar recommendations about the RMI 1 score and further imaging investigations.

The RCOG and the British Society for Gynaecological Endoscopy have produced a joint guideline about the management of suspected ovarian masses in premenopausal women. This guideline aimed to clarify when ovarian masses can be managed in a 'benign' gynaecological service and when referral to a gynaecological oncological service is needed.[10] The guideline notes the importance of thorough history-taking, including risk factors, and careful physical examination, including abdominal and vaginal examination and the determination of the presence or absence of local lymphadenopathy.

The Royal College of Radiologists iRefer radiological investigation guidelines tool[21] recommends that CT of the abdomen and pelvis has a role in identifying women who may benefit from chemotherapy or cytoreductive surgery. MRI of the abdomen and pelvis is recommended for specialised investigation when enhanced CT is contraindicated, or for problem-solving. PET-CT is indicated as a specialised investigation for difficult management situations.

Management of early (stage I) ovarian cancer

National Institute for Health and Care Excellence guideline CG122[1] includes recommendations about the overall management of women with suspected early (stage I) ovarian cancer, and NICE Technology Appraisal (TA) guidance TA55[22] provides recommendations about first-line chemotherapy regimens.

Management of advanced (stage II to IV) ovarian cancer

National Institute for Health and Care Excellence guideline CG122[1] includes recommendations about the management of women with advanced (stages II–IV) ovarian cancer, and NICE TA guidance (TA55 and TA284)[22,23] provides recommendations about first-line chemotherapy regimens.

Further recommendations about chemotherapy regimens for women with recurrent ovarian cancer can be found in NICE TA guidance documents TA389, TA381 and TA285.[24–26]

Summary of the decision problem

The current guidance, NICE clinical guideline CG122,[1] recommends that serum CA125 levels should be measured in secondary care in all women with suspected ovarian cancer in whom serum CA125 levels have not already been measured in primary care. CA125 levels can inform clinical decision-making in secondary care and are not used in isolation; CG122[1] specifically recommends the calculation of a RMI 1 score, which includes CA125 level. CG122 does not currently include any recommendations on HE4 levels, risk scores or testing algorithms (other than RMI 1 score). An update to the section of CG122[1] that deals with establishing a diagnosis in secondary care is planned in order to assess the potential role of alternative risk scores in assessing women with suspected ovarian cancer for possible referral to a SMDT and to consider the best way to incorporate tumour markers and other tests in the decision-making process.

This assessment systematically reviews the evidence about the comparative performance of alternative risk scores that include CA125 levels, HE4 levels or ultrasound (detailed in *Intervention technologies*) to guide referral decisions for women with suspected ovarian cancer in secondary care. The assessment focuses on direct comparisons between the interventions described and the RMI 1 score, using the referral threshold of ≥ 250 (current practice as indicated in CG122[1]). However, assessments of the accuracy of individual risk scores have also been included. Data were collected on the accuracy and comparative accuracy of different risk scores, alternative cut-off values and risk scores used in combination in order to determine the best way to incorporate tumour markers and ultrasound findings in the diagnostic process. Prediction-modelling studies have also been included, which report the development and validation of multivariable prediction models intended to be used to guide individual patient care.

Chapter 3 Assessment of clinical effectiveness

A systematic review was conducted to summarise the evidence on the clinical effectiveness of different risk scores, used as a triage step to guide referral decisions for women with suspected ovarian cancer in secondary care, compared with the RMI 1 score, as recommended in CG122.[1] Systematic review methods followed the principles outlined in the Centre for Reviews and Dissemination's (CRD's) guidance[27] for undertaking reviews in health care and NICE's diagnostics assessment programme manual.[28]

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Systematic review methods

Search strategies

Search strategies were based on the specified risk scores [the ROMA score, the IOTA group's simple ultrasound rules, the ADNEX score, Overa (MIA2G) score and the RMI 1 score] and the target condition (ovarian cancer), as recommended in the CRD's guidance[27] for undertaking reviews in health care and the Cochrane's handbook for diagnostic test accuracy reviews.[29,30]

Candidate search terms were identified from target references, browsing database thesauri (e.g. MEDLINE MeSH and EMBASE Emtree), and from existing reviews identified during the initial scoping searches. These scoping searches were used to generate test sets of target references, which informed the text-mining analysis of high-frequency subject indexing terms, using EndNote X6 [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] reference management software. Strategy development involved an iterative approach, testing candidate text and indexing terms across a sample of bibliographic databases and aiming to reach a satisfactory balance of sensitivity and specificity. Search strategies were developed specifically for each database.

No restrictions on language, publication status or date of publication were applied. Searches took into account generic and other product names for the intervention. The main EMBASE strategy for each search was independently peer reviewed by a second information specialist, using the Canadian Agency for Drugs and Technologies in Health's peer review checklist.[31] Identified references were downloaded in EndNote X6 software for further assessment and handling. References in retrieved articles were checked for additional studies. The final list of included papers were also checked on PubMed for retractions, errata and related citations.[32–35]

The following databases were searched for relevant studies:

- MEDLINE (via Ovid) – 1946 to week 2 November 2016
- MEDLINE In-Process & Other Non-Indexed Citations (via Ovid) – to 22 November 2016
- MEDLINE Daily Update (via Ovid) – to 22 November 2016
- MEDLINE Epub Ahead of Print (via Ovid) – to 22 November 2016
- EMBASE (via Ovid) – 1974 to 23 November 2016
- Cochrane Database of Systematic Reviews (via Wiley Online Library) – to issue 11 of 12, November 2016
- Cochrane Central Register of Controlled Trials (via Wiley Online Library) – to issue 10 of 12, October 2016
- Database of Abstracts of Reviews of Effects (via Wiley Online Library) – to issue 2 of 4, April 2015
- Health Technology Assessment (HTA) Database (via Wiley Online Library) – to issue 4 of 4, October 2016
- International Network of Agencies for HTA publications (via the internet: www.inahta.org/publications/) – to 25 November 2016

- National Institute for Health Research (NIHR) HTA programme (via the internet: www.nets.nihr.ac.uk/programmes/hta) – to 25 November 2016
- Aggressive Research Intelligence Facility database (via the internet: www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/index.aspx) – to 25 November 2016
- PROSPERO (international prospective register of systematic reviews; via the internet: www.crd.york.ac.uk/prospero/) – to 25 November 2016.

Completed and ongoing trials were identified by searches of the following resources:

- National Institutes of Health ClinicalTrials.gov (via the internet: www.clinicaltrials.gov/) – to 24 November 2016
- European Union Clinical Trials Register (via the internet: www.clinicaltrialsregister.eu/ctr-search/search) – to 25 November 2016
- World Health Organization's International Clinical Trials Registry Platform (via the internet: www.who.int/ictrp/en/) – 24 November 2016.

The following key conference proceedings were identified in consultation with clinical experts and were screened for the last 3 years:

- Radiological Society of North America
- American Society of Clinical Oncology Annual Conference
- Society of Gynecologic Oncology
- The National Cancer Research Institute
- European Society of Radiology.

Full search strategies are presented in *Appendix 1*.

Inclusion and exclusion criteria

Inclusion criteria for each of the clinical effectiveness questions are summarised in *Table 3*. Studies that fulfilled these criteria were eligible for inclusion in the review.

Inclusion screening and data extraction

Two reviewers (MW and SL or SD) independently screened the titles and abstracts of all reports identified by searches, and any discrepancies were discussed and resolved by consensus. Full copies of all studies deemed potentially relevant were obtained, and the same reviewers independently assessed these for inclusion; any disagreements were resolved by consensus. Details of the studies excluded at the full-paper screening stage are presented in *Appendix 2*.

When studies reported insufficient information (e.g. tumour marker assay details not specified, incomplete accuracy data), the authors were contacted by e-mail to request additional information.

Studies cited in materials provided by the manufacturers of HE4 assays, the manufacturer of the Overa (MIA2G) multiple-marker test and the IOTA group were first checked against the project reference database in EndNote X6; any studies not already identified by our searches were screened for inclusion, following the aforementioned process.

Data were extracted on the following: study design/details; participant characteristics (age, pre- or post-menopause, presenting symptoms, tumour marker levels and other risk factors, when these were reported); details of the risk score and its component tests [manufacturer, antibody, detection method (including analyser used), ultrasound method and definition of a positive risk score]; details of the reference standard (details of the methods used, when these were reported, definition of disease positive and details of the final histopathological diagnoses of study participants, when these were reported); and test performance outcome measures. Data were extracted by one reviewer, using a piloted, standard data extraction form, and checked by a second reviewer (MW and SL or SD); any disagreements were resolved by consensus.

TABLE 3 Inclusion criteria

Criteria	Question	
	What are the performance characteristics of alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 levels, CA125 levels or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice[1]), for which the target condition is histologically confirmed ovarian cancer?	**What are the effects of using alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 levels, CA125 levels or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice[1]), on clinical management decisions and clinical outcomes?**
Participants	Women of any age with suspected ovarian cancer, who have not previously been treated for ovarian cancer and are not currently receiving chemotherapy	
Setting	Secondary care[a]	
Interventions (index test)	Alternative methods of risk-scoring or RMI 1 used at thresholds other than 250, as described in *Chapter 2*, *Intervention technologies*[b]	
Comparators	RMI 1 score[c]	
Reference standard	Histological examination of surgically resected tissue sample[d]	NA
Outcomes	Diagnostic accuracy (the numbers of TP, FN, FP and TN test results), whereby the target condition is histologically confirmed ovarian cancer	Diagnosis of ovarian cancer confirmed by pathological examination of a biopsy, or prognostic outcomes for ovarian cancer (e.g. stage at diagnosis, differentiation status, suitability for surgical intervention/curative treatment, overall survival, progression-free survival)
Study design[c]	Diagnostic cohort studies directly comparing one or more interventions (index tests) with the comparator[e]	Prediction-modelling studies, randomised and non-RCTs

NA, not applicable; RCT, randomised controlled trial; TN, true negative; TP, true positive.
a Studies will be included if the setting is unclear, but the population is described as women with suspected ovarian cancer.
b Any data on the accuracy of risk scores used in combination or in sequence with one or more additional tests (e.g. RMI 1 score and HE4 levels, IOTA group's Simple Rules and CA125 levels) will also be included.
c Not applicable for prediction-modelling studies.
d Studies that use the histological examination of a biopsy sample or follow-up (ideally for a minimum of 12 months) of women with a risk score below the referral threshold, who do not have a pelvic mass requiring surgery as the reference standard, will also be included.
e Studies assessing the accuracy of individual risk scores will also be included.

Quality assessment

The methodological quality of included test accuracy studies was assessed using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) tool,[36] and the methodological quality of prediction model studies was assessed using the PROBAST (Prediction model study Risk Of Bias Assessment Tool).[37] Quality assessment was undertaken by one reviewer and checked by a second reviewer (MW and SL or SD); any disagreements were resolved by consensus or discussion with a third reviewer.

The results of the quality assessments are summarised in tables and graphs in the results of the systematic review (see *Study quality*) and examples of full quality assessments (QUADAS-2 and PROBAST) are provided in *Appendix 3*; full quality assessments for all included studies can be obtained from the authors.

Methods of analysis/synthesis

Sensitivity and specificity were calculated for each set of 2 × 2 data. All meta-analyses estimated separate pooled estimates of sensitivity and specificity using random-effects logistic regression.[24] The bivariate/hierarchical summary receiver operating characteristic model[38–40] could not be applied because the data sets were too small and/or homogeneous. Heterogeneity was assessed visually, using summary receiver operating characteristic plots or receiver operating characteristic space plots. Analyses were performed in MetaDisc (Hospital Universitario Ramón y Cajal, Madrid, Spain).[41]

The differences between the sensitivity and specificity estimates for different risk scores were described as statistically significant when the 95% confidence intervals (CIs) did not overlap.

Studies were grouped by risk score, manufacturer of the tumour marker assays (when appropriate), definition of disease positive (target condition) and menopausal status. Stratified results tables and forest plots were used to illustrate the variation of test performance by threshold.

Results of the assessment of clinical effectiveness

The searches of bibliographic databases identified 2456 records after deduplication. Following the initial screening of titles and abstracts, 241 publications were considered to be potentially relevant and ordered for full-paper screening; of these, 64 were included in the review.[17,42–103]

In addition, one set of slides from a conference presentation was provided, through NICE, by the manufacturer of Overa (MIA2G),[104] and an unpublished interim report of phase 5 of the IOTA study was provided (confidential information has been removed) personal communication: e-mail via Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project Lead, Kleijnen Systematic Reviews Ltd, 1 March 2017. All potentially relevant studies cited in other documents supplied by the test manufacturers had already been identified through other sources. *Figure 1* shows the flow of studies through the review process, and *Appendix 2* provides details, with reasons for exclusions, of all publications excluded at the full-paper screening stage. In total, there were 51 included studies, reported in 65 publications, and one unpublished interim report.

A total of 165 publications were excluded after full-text screening. Six articles could not be obtained,[105–110] and a further three ongoing studies, reported in four references, were identified as potentially relevant to future updates of this assessment.[111–114] Of particular note is Refining Ovarian Cancer Test accuracy Scores (ROCkeTS),[112,113] a large prospective Phase III study, which was funded by NIHR and which is due to report in 2019/2020. The ROCkeTS study is evaluating the clinical utility, as well as the accuracy, of the RMI 1, ROMA scores, IOTA group's simple ultrasound rules and other models and novel models not included in the scope of this assessment, and will consider the delivery of tests in the NHS (in which an imaging service is predominantly delivered by sonographers, rather than expert gynaecologists or radiologists). Trial registry entries for two additional diagnostic test accuracy studies were identified: one ongoing study is comparing the diagnostic performance or IOTA group's simple ultrasound rules with that of ultrasound pattern recognition in women undergoing surgery for adnexal mass (the reference standard is the histopathological diagnosis) and the estimated completion date is September 2017;[111] the second trial registry entry referred to a study assessing the diagnostic performance of a two-step triage process involving RMI 1 (threshold of 200) and IOTA group's simple ultrasound rules, which has been terminated without publication.[114]

The authors of 11 studies that were reported as conference abstracts with insufficient detail were contacted to determine whether or not the studies met our inclusion criteria, or when the outcomes were unclearly reported in the full paper;[45,53,60,83,84,90,94,115–118] four authors provided additional information that allowed the study to be included in this review.[83,84,90,94]

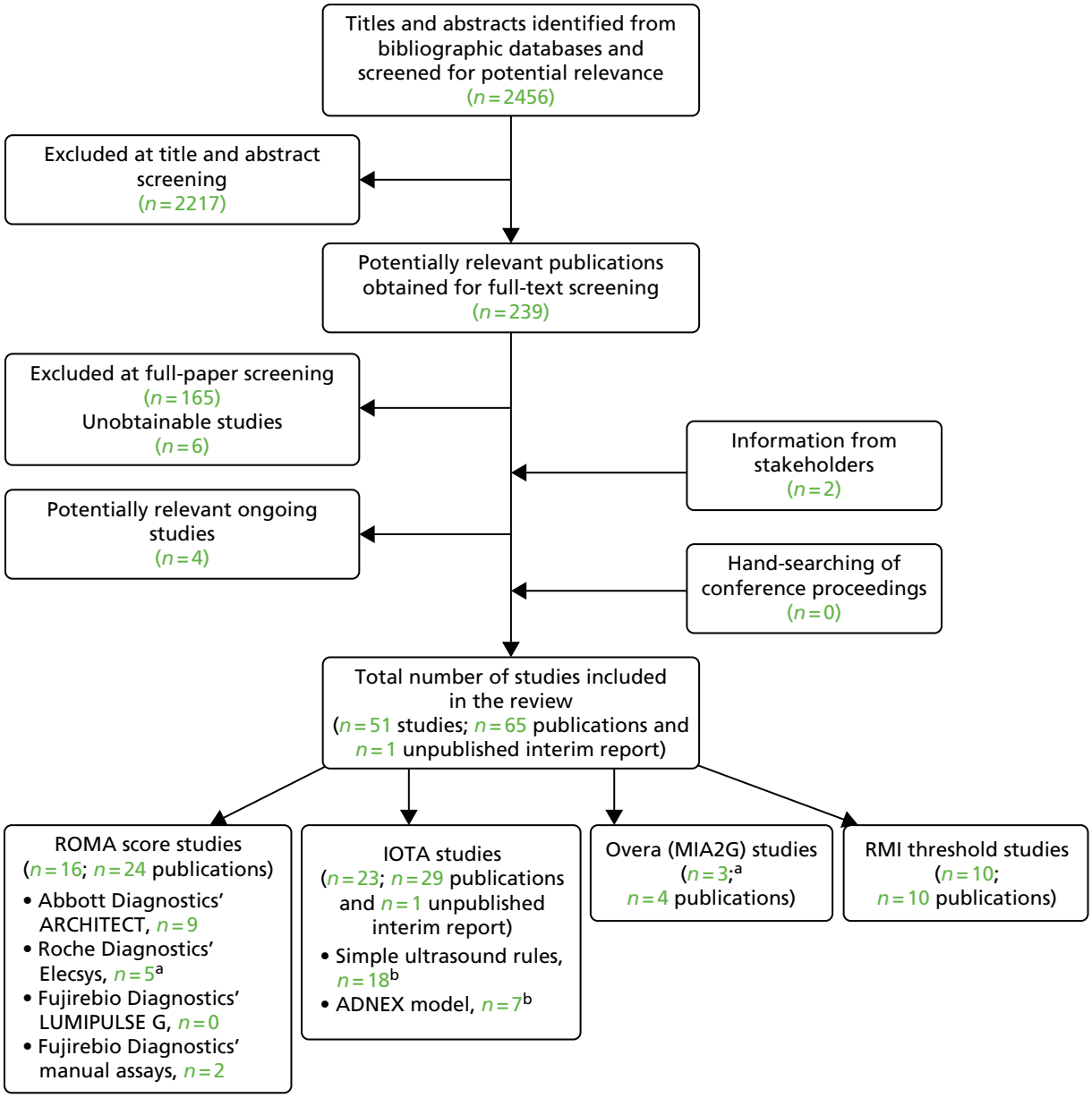

FIGURE 1 Flow of studies through the review process. a, One study reported data for both the ROMA score using Roche Diagnostics' Elecsys tumour marker assays and Overa (MIA2G); and b, two studies reported data for both the ADNEX model and IOTA group's simple ultrasound rules.

Overview of included studies

Details of the 51 included studies and their associated references are provided in *Table 4*. The following sections of this report cite studies using the primary publication and, when this is different, the publication (shown in bold in *Table 5*) in which the referenced data were reported.

All studies included in our systematic review were diagnostic cohort studies that reported data on the diagnostic accuracy of one or more ovarian cancer risk scores [the ROMA score, IOTA group's simple ultrasound roles, the ADNEX model or Overa (MIA2G)], or that provided data on the accuracy of the RMI 1 at different decision thresholds (including 250, as specified in the current NICE guidelines[1]). Although 10 studies reported an age range that included women aged < 18 years,[42,44,48,51,52,61,64,65,83,103] no study reported separate test performance data for this age group or indicated how many women were aged < 18 years. Sixteen studies reported data on the accuracy of the ROMA score,[81–83,86,89,90,94–99,101–104] five of which reported data to support a direct comparison of the ROMA score to the RMI 1 score, using a

TABLE 4 Details of included studies

Details	Country	*n*	Main target condition reported
ROMA score			
Abbott Diagnostics			
ARCHITECT			
Karlsen *et al.* (2012)[83]	Denmark	579	All ovarian malignancies, excluding borderline
Al Musalhi *et al.* (2016)[103]	Oman	213	All malignant tumours, including borderline
Chan *et al.* (2013)[82]	Multinational (Asia)	387	All epithelial ovarian malignancies, including borderline
Clemente *et al.* (2015)[90]	The Philippines	62	Ovarian malignancies (undefined – not clear whether or not borderline tumours were included)
Li *et al.* (2016)[96]	China	917	Ovarian malignancies (undefined – not clear whether or not borderline tumours were included)
• Moore *et al.* (2011)[101] • Moore *et al.* (2011)[87] • Moore *et al.* (2012)[88]	USA	450	All epithelial ovarian malignancies, including borderline
Novotny *et al.* (2012)[86]	Czech Republic	277	All malignant tumours, including borderline
Presl *et al.* (2012)[81]	Czech Republic	552	Ovarian malignancies (undefined – not clear whether or not borderline tumours were included)
Winarto *et al.* (2014)[99]	Indonesia	128	All epithelial ovarian malignancies, including borderline
Fujirebio Diagnostics			
Langhe *et al.* (2013)[94]	NR	377	All malignant tumours, including borderline
• Van Gorp *et al.* (2012)[98] • Kaijser *et al.* (2013)[91] • Kaijser *et al.* (2013)[92] • Kaijser *et al.* (2013)[56] • Kaijser *et al.* (2014)[100] • Van Gorp *et al.* (2011)[85]	Belgium	374	All malignant tumours, including borderline
Roche Diagnostics			
Janas *et al.* (2015)[97]	Poland	259	All malignant tumours, including borderline
Shulman *et al.* (2016)[104]	USA	993	All malignant tumours, including borderline
Xu *et al.* (2016)[95]	China	521	All epithelial ovarian malignancies, excluding borderline
Yanaranop *et al.* (2016)[89]	Thailand	260	All malignant tumours – borderline tumours classified as disease negative
• Zhang *et al.* (2015)[102] • Chen *et al.* (2015)[93]	China	612	All epithelial ovarian malignancies

TABLE 4 Details of included studies (*continued*)

Details	Country	*n*	Main target condition reported
Simple ultrasound rules (IOTA group)			
• Abdalla *et al.* (2013)[48] • Abdalla *et al.* (2013)[57]	Poland	87	All malignant tumours, including borderline
Alcázar *et al.* (2013)[52]	Spain	340	All malignant tumours, including borderline
Baker *et al.* (2013)[66]	UK	28	All ovarian malignancies
• Di Legge *et al.* (2012)[61] • Di Legge *et al.* (2012)[61]	Multinational (worldwide)	2445	All malignant tumours, including borderline
Fathallah *et al.* (2011)[63]	France	109	All malignant tumours, including borderline
IOTA[a]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
• Knafel *et al.* (2015)[49] • Knafel *et al.* (2013)[54]	Poland	226	All malignant tumours, including borderline
Meys *et al.* (2016)[44]	The Netherlands	326	All malignant tumours, including borderline
Murala *et al.* (2014)[60]	UK	51	All malignant tumours (undefined – not clear whether or not borderline tumours were included)
• Piovano *et al.* (2016)[58] • Piovano *et al.* (2015)[84]	Italy	391	All malignant tumours, including borderline
Ruiz de Gauna *et al.* (2015)[64]	Spain	154	All malignant tumours, including borderline
Sayasneh *et al.* (2013)[62]	UK	255	All malignant tumours, including borderline
Silvestre *et al.* (2015)[55]	Brazil	75	All malignant tumours, including borderline
Tantipalakorn *et al.* (2014)[51]	Thailand	319 (masses)	All malignant tumours, including borderline
Testa *et al.* (2014)[50]	Multinational (Europe)	2403	All malignant tumours, including borderline
• Timmerman *et al.* (2010)[65] • Ameye *et al.* (2012)[67]	Multinational (worldwide)	1938	All malignant tumours, including borderline
Tinnangwattana *et al.* (2015)[47] Tongsong *et al.* (2016)[59]	Thailand	94	All malignant tumours, including borderline
Weinberger *et al.* (2013)[53]	NR	347	All ovarian malignancies, including borderline
ADNEX model			
IOTA[a]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Joyeux *et al.* (2016)[43]	France	284	Ovarian malignancies, including borderline

continued

TABLE 4 Details of included studies (*continued*)

Details	Country	*n*	Main target condition reported
Meys *et al.* (2016)[44]	The Netherlands	326	All malignant tumours, including borderline
Moffatt *et al.* (2016)[45]	UK	81	Ovarian malignancies (undefined – not clear whether or not borderline tumours were included)
Sayasneh *et al.* (2016)[46]	UK and Italy	610	All malignant tumours, including borderline
Szubert *et al.* (2016)[42]	Poland and Spain	327	All ovarian malignancies, including borderline
Van Calster *et al.* (2014)[17]	Multinational (Europe)	2403	All malignant tumours, including borderline
Overa (MIA2G)			
Coleman *et al.* (2016)[70] Wolf *et al.* (2015)[69]	USA	493	All malignant tumours, including borderline
Shulman *et al.* (2016)[104]	USA	993	All malignant tumours, including borderline
Zhang *et al.* (2015)[68]	USA	305	All malignant tumours, including borderline
RMI 1 threshold variation			
Aktürk *et al.* (2011)[71]	Turkey	100	All ovarian malignancies, excluding borderline
Asif *et al.* (2004)[77]	Pakistan	100	All malignant tumours (undefined – not clear whether or not borderline tumours were included)
Davies *et al.* (1993)[79]	UK	124	All malignant tumours, including borderline
Jacobs *et al.* (1990)[78]	UK	139	All malignant tumours, including borderline
Lou *et al.* (2010)[73]	China	223	All malignant tumours, including borderline
Manjunath *et al.* (2001)[75]	India	148	All malignant tumours, excluding borderline
Morgante *et al.* (1999)[80]	Italy	124	All malignant tumours, including borderline
Tingulstad *et al.* (1996)[76]	Norway	173	All malignant tumours, including borderline
Ulusoy *et al.* (2007)[74]	Turkey	296	All malignant tumours, including borderline
Yamamoto *et al.* (2009)[72]	Japan	253	All ovarian malignancies, including borderline

NR, not reported.
a Frances Nixon, personal communication.
Note that some studies evaluated multiple risk scores.

TABLE 5 The QUADAS-2 results for accuracy studies of risk scores

	Risk of bias				Applicability		
Study (year of publication)	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Abdalla *et al.* (2013)[48]	?	+	?	+	–	–	–
Aktürk *et al.* (2011)[71]	?	+	?	?	+	+	+
Al Musalhi *et al.* (2016)[103]	?	+	?	?	+	+	–
Alcázar *et al.* (2013)[52]	?	+	?	+	+	+	+
Asif *et al.* (2004)[77]	+	?	?	?	+	+	+
Baker *et al.* (2013)[66]	–	?	?	–	+	–	?
Chan *et al.* (2013)[82]	+	+	+	–	+	+	+
Clemente *et al.* (2015)[90]	?	?	+	+	?	+	+
Coleman *et al.* (2016)[70]	+	+	+	+	+	+	+
Davies *et al.* (1993)[79]	+	+	–	–	+	+	?
Di Legge *et al.* (2012)[61]	+	+	?	+	–	–	?
Fathallah *et al.* (2011)[63]	+	+	+	–	?	+	+
IOTA5 (2017)[a]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Jacobs *et al.* (1990)[78]	+	+	–	–	+	+	?
Janas *et al.* (2015)[97]	?	+	?	+	?	+	+
Joyeux *et al.* (2016)[43]	?	?	?	+	+	+	+
Karlsen *et al.* (2012)[83]	?	+	?	–	+	+	+
Knafel *et al.* (2016)[49]	+	+	?	+	?	–	–
Langhe *et al.* (2013)[94]	?	+	?	–	?	–	?
Li *et al.* (2016)[96]	+	+	?	?	?	+	?
Lou *et al.* (2010)[73]	?	?	+	?	+	+	–
Manjunath *et al.* (2001)[75]	+	+	+	–	+	+	+
Meys *et al.* (2016)[44]	+	–	+	+	?	–	+
Moffatt *et al.* (2016)[45]	?	?	+	–	+	–	?
Moore *et al.* (2011)[101]	?	+	?	–	–	+	+
Morgante *et al.* (1999)[80]	+	?	?	?	+	+	?
Murala *et al.* (2014)[60]	+	–	?	–	+	?	?
Novotny *et al.* (2012)[86]	?	+	?	?	+	+	?
Piovano *et al.* (2016)[58]	+	+	+	+	?	+	–
Presl *et al.* (2012)[81]	?	?	?	?	+	?	?

continued

TABLE 5 The QUADAS-2 results for accuracy studies of risk scores (*continued*)

Study (year of publication)	Risk of bias				Applicability		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ruiz de Gauna *et al.* (2015)[64]	+	+	+	+	+	+	+
Sayasneh *et al.* (2013)[62]	+	+	?	+	–	+	–
Sayasneh *et al.* (2016)[46]	+	?	+	+	+	+	–
Shulman *et al.* (2016)[104]	?	?	?	?	+	+	–
Silvestre *et al.* (2015)[55]	+	+	+	+	+	+	–
Szubert *et al.* (2016)[42]	?	+	?	+	+	–	–
Tantipalakorn *et al.* (2014)[51]	?	+	?	–	+	+	–
Testa *et al.* (2014)[50]	+	+	+	+	–	–	–
Timmerman *et al.* (2010)[65]	+	+	+	–	–	–	–
Tingulstad *et al.* (1996)[76]	–	+	?	?	+	+	–
Tinnangwattana *et al.* (2015)[47]	+	+	+	+	+	+	–
Ulusoy *et al.* (2007)[74]	+	+	?	+	–	–	–
Van Calster *et al.* (2014)[17]	+	+	+	?	+	+	–
Van Gorp *et al.* (2012)[98]	+	+	?	–	?	–	+
Weinberger and Minar (2013)[53]	?	?	?	?	?	–	?
Winarto *et al.* (2014)[99]	?	+	?	+	?	+	+
Xu *et al.* (2016)[95]	–	+	?	+	?	+	+
Yamamoto *et al.* (2009)[72]	?	+	?	+	+	+	+
Yanaranop *et al.* (2016)[89]	?	+	+	+	+	+	+
Zhang *et al.* (2015)[68]	?	?	?	?	+	+	?
Zhang *et al.* (2015)[102]	–	+	?	–	?	+	+

+, low risk; –, high risk; ?, unclear risk.
a Frances Nixon, personal communication.

decision threshold of 200.[83,89,98,99,103] There were no studies that reported comparative accuracy data for the ROMA score versus the RMI 1, using a decision threshold of 250. Seventeen published studies reported data on the accuracy of the IOTA group's simple ultrasound rules,[44,47–53,55,58,60–66] six of which reported data to support a direct comparison of the IOTA group's simple ultrasound rules with the RMI 1 score, using a decision threshold of 200.[44,48,50,61,62,65] One study compared the IOTA group's simple ultrasound rules with the RMI 1 score, using a decision threshold of 250, but this study was reported only as a conference abstract and the results were incomplete.[60] Six published studies reported data on the accuracy of the ADNEX model,[17,42–46] one of which reported data to support a direct comparison of the ADNEX model with the RMI 1, using a decision threshold of 200.[44] The unpublished interim report (Frances Nixon, personal communication) provided data to support a direct comparison between the IOTA group's simple ultrasound rules, the ADNEX model and the RMI 1 at both decision thresholds (200 and 250). Three studies reported data on the accuracy of Overa (MIA2G),[68,70,104] one of which also provided comparative accuracy data for Overa (MIA2G) versus the ROMA score.[104] There were no studies comparing the accuracy of Overa (MIA2G) with the RMI 1, at any decision threshold. Finally, 10 studies provided data on the accuracy of the RMI 1 at different decision thresholds.[71–80]

No randomised controlled trials (RCTs) or controlled clinical trials (CCTs) were identified; no studies provided data on patient-relevant outcomes following different risk assessment strategies.

Approximately half of the included published studies (25/51) were conducted in Europe,[17,42–46,48–50,52,58,60,62–64,66,78–81,83,86,97,98,119] six of which were conducted solely in the UK[45,60,62,66,78,79] and a further two were multinational studies that included a UK centre.[17,42] The unpublished interim analysis (Frances Nixon, personal communication) (confidential information has been removed). There were two multinational, worldwide studies, both of which included UK centres.[61,65] Four studies were conducted in the USA,[68,70,101,104] 13 were conducted in Asia,[47,51,72,73,75,77,82,89,90,95,96,99,102] two were conducted in Turkey,[71,74] one was conducted in Oman[103] and one was conducted in Brazil.[55] Two studies, which were published only as conference abstracts, did not report information about geographic location.[53,94]

Seventeen published studies,[17,44,46,47,50,51,61,62,65,78,81,86,95–98,102] and the unpublished study for which an interim report was provided (confidential information has been removed) (Frances Nixon, personal communication), were publicly funded, and four studies reported receiving some funding from manufacturers (including a supply of test kits, reagents and analysers).[70,82,83,101] The remaining 29 included studies either did not report any information about funding[42,43,45,48,49,52,53,55,58,60,63,64,66,68,71–77,79,80,89,90,94,99,104] or stated that they were unfunded.[103]

All studies included women with an adnexal/ovarian mass; however, studies frequently reported analyses that excluded some women based on their final histopathological diagnosis (information that could not be known at the point of presentation); hence, only those studies that reported data for the target condition 'all malignant tumours, including borderline' could be considered to have evaluated risk scores in a population similar to those in whom these scores would be applied in practice. Full study details [inclusion and exclusion criteria, baseline characteristics of study participants and details of the risk score(s) (index test) evaluated are provided in *Appendix 4* (*Tables 34* and *35*)].

Study quality

All studies included in this systematic review were diagnostic cohort studies. The methodological quality of these studies was assessed using the QUADAS-2 tool (summarised in *Table 5* and *Figure 2*). One of these studies[17] reported the development and validation of the ADNEX model, in addition to the test accuracy results. This study was assessed using PROBAST, a tool specifically developed to assess the methodological quality of prediction-modelling studies, (*Table 6*) as well as the QUADAS-2. Examples of full QUADAS-2 and PROBAST assessments are provided in *Appendix 3*, and full assessments for each included study are available on request.

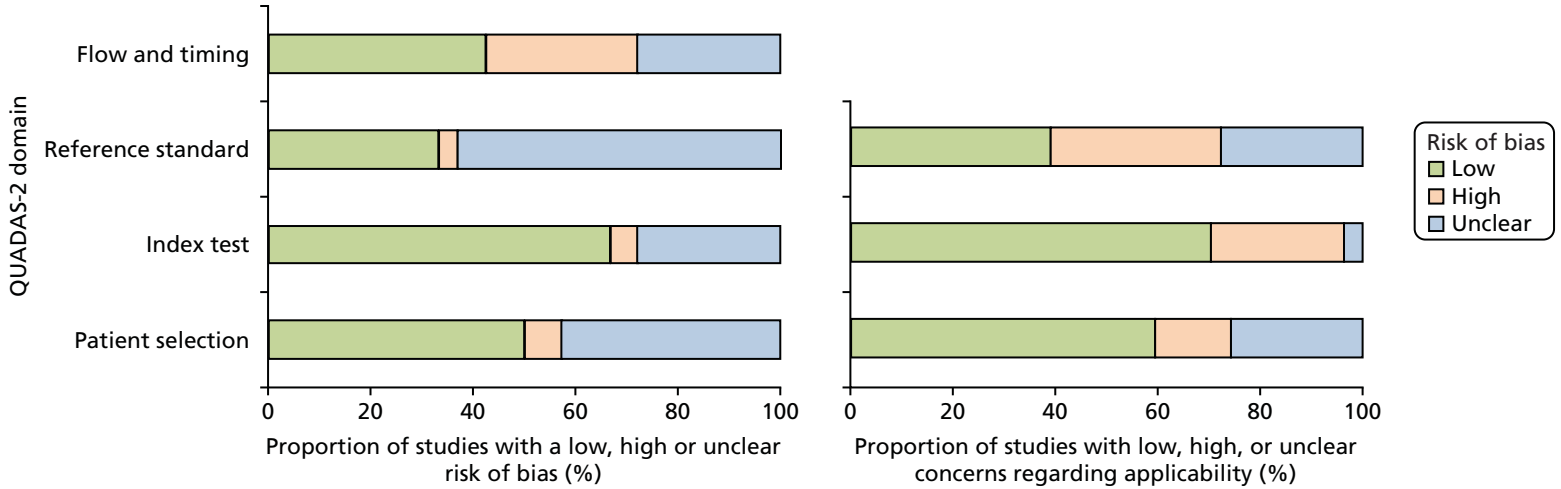

FIGURE 2 Summary of QUADAS-2 results for accuracy studies of risk scores.

TABLE 6 The PROBAST results for studies reporting the development and validation of risk scores

Study (year of publication)	Risk of bias									Applicability concerns						
	Participant selection		Predictors		Outcome		Analysis		Overall judgement	Participant selection		Predictors		Outcome		Overall judgement
	Development	Validation	Development	Validation	Development	Validation	Development	Validation		Development	Validation	Development	Validation	Development	Validation	
Van Calster *et al.* (2014)[17]	+	+	+	+	?	?	?	+	?	–	–	+	+	+	+	+

+, low risk; –, high risk; ?, unclear risk.

Eight studies were reported only as conference abstracts or meeting slides, with limited descriptions of the methods used,[45,53,60,66,68,90,94,104] and study methods were generally poorly reported. Thirty-seven studies (73%) were rated as having an 'unclear' risk of bias on at least one QUADAS-2 domain, and 24 studies (47%) were rated as being 'unclear' for applicability on at least one domain.

Two studies[64,70] were rated as having a 'low' risk of bias and 'low' concerns regarding applicability for all domains, and four further studies were rated low for all risk-of-bias domains.[47,50,55,58] In total, 11 studies (22%) were rated as having 'low' concerns regarding all applicability domains.[43,52,64,70–72,75,77,82,83,89]

Nineteen studies (37%) were rated as having a 'high' risk of bias on at least one QUADAS-2 domain, whereas 26 studies (51%) were rated as 'high' for applicability on at least one domain.

The main potential sources of bias across the included published studies concerned flow and timing. Fifteen studies (30%) were rated as having a 'high' risk of bias on the flow and timing domain. For most of these studies (13/15[45,47,51,60,63,65,66,78,82,94,98,101,102]), this was because not all included patients were included in the analysis. In five studies[51,75,78,79,83] the included patients did not all receive the same reference standard.

The main areas of concern regarding applicability were in relation to how the index test was applied and whether or not this could be considered to be representative of routine practice, and how the reference standard positive (target condition) was defined. Fourteen studies (28%) were rated as having 'high' concerns regarding the applicability of the index test; for six studies,[48,50,61,65,94,98] this was because all or part of the index test was performed before referral; in three studies,[45,53,66] the index test was applied retrospectively to existing patient data; and in seven studies,[42,44,49,50,53,65,74] the index test was performed by practitioners whose level of experience was judged to be higher than that likely to be routinely available in secondary care settings. Eighteen studies (35%) were rated as having 'high' concerns regarding the applicability of the reference standard because malignancy was defined as 'any malignant tumour', which could include non-ovarian cancers and metastases, whereas the scope of this assessment defined the target condition as ovarian cancer. However, it should be noted that, in order for a study to report risk score performance data for the specific target condition of ovarian cancer, study participants found to have non-ovarian cancers and metastases would need to be excluded from the analysis. Studies that excluded patients with non-ovarian cancers and metastases were rated as having a 'high' risk of bias on the flow and timing domain, because post hoc exclusion of these patients may result in overestimation of test performance. *Appendix 4*, *Table 36* lists the final histological diagnoses (where reported) of the study participants. These data illustrate the between-study variations in the definitions of disease positive used, which could include borderline, non-ovarian cancers, metastatic cancers and non-ovarian metastatic cancers. To take into account as much of this heterogeneity as possible, the results were analysed according to whether or not disease positive (target condition) was defined as 'ovarian malignancy' or 'any malignant tumour', and whether or not this definition included borderline tumours.

(Confidential information has been removed.)

Overall, more than half of the included studies were rated as having a high or unclear risk of bias for patient selection, the reference standard and flow and timing. More than half of the studies were rated as having a high level of, or unclear, concern for the applicability of the reference standard.

The PROBAST prediction score (see *Table 6*) for Van Calster *et al.*[17] indicated that there was a high risk of bias for the applicability of patient selection. The high risk of bias was attributable to the selection of women from a mixture of secondary and tertiary care centres, which is not a complete match for the scope of this assessment. However, the ADNEX model adjusts for study setting and, therefore, the overall concern regarding applicability is low. The overall risk of bias was judged to be unclear, as not all aspects of the model development were clearly described.

Clinical effectiveness of risk scores

No RCTs or CCTs were identified; no studies provided data on patient-relevant outcomes following different risk assessment strategies.

Diagnostic performance of the Risk of Ovarian Malignancy Algorithm score

Details of Risk of Ovarian Malignancy Algorithm studies

Sixteen diagnostic cohort studies,[81–83,86,89,90,94–99,101–104] reported in 24 publications,[56,81–83,85–104] provided data on the diagnostic performance of the ROMA score for identifying women who have an adnexal mass and are at a high risk of developing ovarian cancer. Nine studies[81–83,86,90,96,99,101,103] used a ROMA score based on Abbott Diagnostics' ARCHITECT tumour marker assays, of which six[81,82,86,96,99,103] evaluated a decision threshold for the ROMA score that was consistent with the manufacturer's recommendations. None of the included studies used the Fujirebio Diagnostics' LUMIPULSE G automated CEIA system. For information, two studies[94,98] that used a ROMA score based on manual Fujirebio Diagnostics' tumour marker EIAs (see *Appendix 5*, *Tables 41* and *42*) were included, both using the manufacturer's recommended decision threshold for the ROMA score; however, it should be noted that the manual assays are not specified interventions for this assessment. Finally, five studies[89,95,97,102,104] used a ROMA score based on Roche Diagnostics' Elecsys tumour marker assays, all of which used the manufacturer's recommended decision threshold for the ROMA score.

None of the ROMA score studies that used Abbott Diagnostics' ARCHITECT tumour marker assays was conducted in the UK; three studies[81,83,86] were conducted in European countries, four[82,90,96,99] were conducted in Asia, one[101] was conducted in the USA and one[103] was conducted in Oman. None of the ROMA score studies that used Roche Diagnostics' Elecsys tumour marker assays was conducted in the UK, and only one[97] was conducted in a European country. Three[89,93,95] of the remaining studies were conducted in Asia and one[104] was conducted in the USA.

This assessment is primarily concerned with providing a comparison between the RMI 1,[78] used with a decision threshold of 250 (current standard practice in the NHS[1]), and the specified alternative risk-scoring methods (see *Chapter 2*, *Intervention technologies*). No studies were identified that reported a direct comparison (both tests used to assess the same patient cohort) between the ROMA score and the RMI 1, used with a decision threshold of 250. Five studies reported direct comparisons between the ROMA score and the RMI 1, used with a decision threshold of 200; three studies used Abbott Diagnostics' ARCHITECT tumour marker assays;[83,99,103] one study used Roche Diagnostics' Elecsys assays;[89] and one study used Fujirebio Diagnostics' manual EIAs.[98] The following sections report all available data from direct comparison studies, as well as non-comparative data on the accuracy of the ROMA score, when decision thresholds that were consistent with the manufacturers' recommendations were used. Additional accuracy data for alternative decision thresholds are reported in *Appendix 5*, *Table 37*.

The target condition for this assessment is ovarian cancer, including conditions covered by the NICE clinical guideline CG122[1] (i.e. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma and borderline ovarian cancer). All studies in this section included women with one or more adnexal mass. The definition of reference standard positive 'ovarian cancer' varied between studies, with borderline tumours being most frequently classified as positive or excluded from analyses. In addition, some studies included patients with non-ovarian primary cancers/metastases to the ovary[97,98,103] and germ cell tumours.[103] When the target condition was described as 'all ovarian malignancy', those women whose postoperative histological diagnosis was identified as non-ovarian primary were excluded from the estimates of test performance. Conversely, when the target condition was described as 'all malignant tumours', women with a non-ovarian primary were not excluded and were classified as being disease positive; this could potentially include women with any tumour on the ovaries that has metastasised from another primary [e.g. colorectal cancer (CRC)], and/or women with an adnexal/pelvic mass that turns out to be non-ovarian (not clearly specified by the included studies). Full details of the final histopathological diagnoses of study women who had a malignant mass are reported in *Appendix 4*, *Table 36*.

Accuracy of the Risk of Ovarian Malignancy Algorithm score using Abbott Diagnostics' ARCHITECT tumour marker assays

Three[83,99,103] of the nine[81–83,86,90,96,99,101,103] ROMA score studies that used Abbott Diagnostics' ARCHITECT tumour marker assays, reported a direct comparison of the ROMA score with the RMI 1. Only one study included all participants the analysis, regardless of their final histopathological diagnosis (target condition: all malignant tumours, including borderline) and this study used different thresholds from those recommended by the manufacturer (13.1% in premenopausal women and 27.7% in postmenopausal women, as opposed to the manufacturer's recommendation of 7.4% and 25.3%).[103] One study was a retrospective study, which excluded women with histopathological diagnoses other than epithelial ovarian cancer.[99] A second study excluded from the analysis nine women (1%) with non-epithelial ovarian cancer, 69 women (6%) with non-ovarian cancers and 252 women (21%) with borderline tumours;[83] the distribution of positive and negative ROMA score results in these women was not reported.

The sensitivity estimate for the ROMA score was highest (96.4%, 95% CI 93.6% to 98.2%) when the analyses excluded women with borderline tumours and those with malignancies other than epithelial ovarian cancer, and lowest (75.0%, 95% CI 60.4% to 86.4%) when all women were included in the analysis, regardless of their final histopathological diagnosis (*Table 7*). Conversely, the specificity estimate for the ROMA score was highest (87.9%, 95% CI 81.9% to 92.4%) in the study that included all participants,[103] and lowest (53.3%, 95% CI 50.0% to 56.7%) when the analyses excluded women with borderline tumours and those with malignancies other than epithelial ovarian cancer (see *Table 7*). When women with borderline tumours and/or those with malignancies other than epithelial ovarian cancer were excluded from the analyses, the sensitivity estimates for the ROMA score were not significantly different from those for the RMI 1 (threshold 200), whereas the specificity estimates were significantly lower (see *Table 7*). In contrast, the study that included all participants in the analysis reported similar sensitivity and specificity estimates for the ROMA score and the RMI 1, with a sensitivity of 75% (95% CI 60.4% to 86.4%) versus 77.1% (95% CI 62.7% to 88.0%), and a specificity of 87.9% (95% CI 81.9% to 92.4%) versus 81.8% (95% CI 75.1% to 87.4%), respectively.[103] This study also reported lower sensitivity and higher specificity estimates, for both the ROMA score and the RMI 1, in premenopausal women than those in postmenopausal women (see *Table 7*).

One study reported test performance estimates calculated both with and without the inclusion of participants with borderline tumours.[99] Although the number of participants involved was small, these data indicated that around half of the FN risk scores were accounted for by women with borderline tumours, 3 out of 6 (50%) using the ROMA score and 7 out of 13 (54%) using the RMI 1 (threshold of 200).[99] Approximately 13% (17/128) of the women in this study had borderline tumours, whereas 39% (50/128) had malignant tumours [i.e. a higher proportion of women with borderline tumours had a negative ROMA score – 17.6% (3/17) – than was the case for women with malignant tumours – 6% (3/50)].[99] A similar pattern was observed for the RMI 1; the proportion of women with borderline tumours who had a negative RMI 1 was approximately 41% (7/17), compared with 12% (6/50) for those with malignant ovarian tumours.[99]

One additional study reported performance estimates for the ROMA score, excluding women with borderline tumours and those with non-ovarian malignancies, without a comparison with the RMI 1.[82] When data from this study were combined with the ROMA data from the two similar comparative accuracy studies,[83,99] the summary estimates of sensitivity did not change significantly [95.1%, 95% CI 92.4% to 97.1%, based on three studies (*Table 8*), vs. 96.4%, 95% CI 93.6% to 98.2%, based on two studies; see *Table 7*]. The summary estimate of specificity, based on all three studies (62.5%, 95% CI 59.7% to 65.3%; see *Table 8*), was higher than that derived from the two comparative accuracy studies alone (53.3%, 95% CI 50% to 56.7%; see *Table 7*). There were no additional studies that evaluated the performance of the ROMA score alone, and included all participants in the analysis (target condition: all malignant tumours, including borderline).

TABLE 7 Comparative accuracy of the ROMA score using Abbott Diagnostics' ARCHITECT tumour marker assays vs. the RMI

Study (year of publication)	Subgroup	ROMA threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)	RMI 1	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours, including borderline																	
Al Musalhi *et al.* (2016)[103]	All women	13.1%/27.7%	36	12	20	145	213	75.0 (60.4 to 86.4)	87.9 (81.9 to 92.4)	200	37	11	30	135	213	77.1 (62.7 to 88.0)	81.8 (75.1 to 87.4)
	Premenopausal women	13.1%	11	10	14	127	162	52.4 (29.8 to 74.3)	90.1 (83.9 to 94.5)	200	12	9	21	120	162	57.1 (34.0 to 78.2)	85.1 (78.1 to 90.5)
	Postmenopausal women	27.7%	25	2	5	19	51	92.6 (75.7 to 99.1)	79.2 (57.8 to 92.9)	200	22	2	9	18	51	91.7 (73.0 to 99.0)	66.7 (46.0 to 83.5)
Target condition: epithelial ovarian malignancies, including borderline																	
Winarto *et al.* (2014)[99]	All women	7.4%/25.3%	61	6	35	26	128	91.0 (81.5 to 96.6)	42.6 (30.0 to 55.9)	200	54	13	21	40	128	80.6 (69.1 to 89.2)	65.6 (52.3 to 77.3)
Target condition: epithelial ovarian malignancies, excluding borderline																	
Karlsen *et al.* (2012)[83]	All women	7.4%/25.3%	244	8	371	438	1061	96.8 (93.8 to 98.6)	54.1 (50.6 to 57.6)	200	238	14	150	659	1061	94.4 (90.9 to 96.9)	81.5 (78.6 to 84.1)
Winarto *et al.* (2014)[99]	All women	7.4%/25.3%	47	3	35	26	111	94.0 (83.5 to 98.7)	42.6 (30.0 to 55.9)	200	44	6	21	40	111	88.0 (75.7 to 95.5)	65.6 (52.3 to 77.3)
Summary estimates								96.4 (93.6 to 98.2)	53.3 (50.0 to 56.7)							93.4 (90.0 to 95.9)	80.3 (77.5 to 82.9)
Karlsen *et al.* (2012)[83]	Premenopausal women	7%	46	3	251	279	579	93.9 (83.1 to 98.7)	52.6 (48.3 to 57.0)	200	41	8	42	488	579	83.7 (70.3 to 92.7)	92.1 (89.4 to 94.2)
	Postmenopausal women	25.3%	198	5	120	159	482	97.5 (94.3 to 99.2)	57.0 (51.0 to 62.9)	200	196	7	108	171	482	96.6 (93.0 to 98.6)	61.3 (55.3 to 67.0)

TN, true negative; TP, true positive.

TABLE 8 Accuracy of the ROMA score using Abbott Diagnostics' ARCHITECT tumour marker assays at the manufacturer's recommended thresholds

Study (year of publication)	Subgroup	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: epithelial ovarian malignancies, including borderline									
Winarto *et al.* (2014)[99]	All women	7.4%/25.3%	61	6	35	26	128	91.0 (81.5 to 96.6)	42.6 (30.0 to 55.9)
Target condition: epithelial ovarian malignancies, excluding borderline									
Karlsen *et al.* (2012)[83]	All women	7.4%/25.3%	244	8	371	438	1061	96.8 (93.8 to 98.6)	54.1 (50.6 to 57.6)
Chan *et al.* (2013)[82]	All women	7.4%/25.3%	58	7	41	281	387	89.2 (79.1 to 95.6)	87.3 (83.1 to 90.7)
Winarto *et al.* (2014)[99]	All women	7.4%/25.3%	47	3	35	26	111	94.0 (83.5 to 98.7)	42.6 (30.0 to 55.9)
Summary estimates								**95.1 (92.4 to 97.1)**	**62.5 (59.7 to 65.3)**
Karlsen *et al.* (2012)[83]	Premenopausal women	7%	46	3	251	279	579	93.9 (83.1 to 98.7)	52.6 (48.3 to 57.0)
Chan *et al.* (2013)[82]	Premenopausal women	7%	18	4	34	235	291	81.8 (59.7 to 95.9)	87.4 (82.8 to 91.1)
Summary estimates								**90.1 (80.7 to 95.9)**	**64.3 (60.9 to 67.7)**
Karlsen *et al.* (2012)[83]	Postmenopausal women	25.3%	198	5	120	159	482	97.5 (94.3 to 99.2)	57.0 (51.0 to 62.9)
Chan *et al.* (2013)[82]	Postmenopausal women	25.3%	40	3	7	46	96	93.0 (80.9 to 98.5)	86.8 (56.3 to 67)
Summary estimates								**96.7 (93.7 to 98.6)**	**61.7 (56.3 to 67.0)**
Target condition: epithelial ovarian malignancies (stage III/IV) – borderline and stage I/II tumours excluded									
Chan *et al.* (2013)[82]	All women	7.4%/25.3%	35	3	41	281	360	92.1 (78.6 to 98.3)	87.3 (83.1 to 90.7)
	Premenopausal women	7%	10	2	34	235	281	83.3 (51.6 to 97.9)	87.4 (82.8 to 91.1)
	Postmenopausal women	25.3%	24	1	7	46	78	96.0 (79.6 to 99.9)	86.8 (74.7 to 94.5)

continued

TABLE 8 Accuracy of the ROMA score using Abbott Diagnostics' ARCHITECT tumour marker assays at the manufacturer's recommended thresholds (*continued*)

Study (year of publication)	Subgroup	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: epithelial ovarian malignancies (stage I/II) – borderline and stage III/IV tumours excluded									
Chan *et al.* (2013)[82]	All women	7.4%/25.3%	19	4	41	281	345	82.6 (61.2 to 95.0)	87.3 (83.1 to 90.7)
	Premenopausal women	7.4%	6	2	34	235	277	75.0 (34.9 to 96.8)	87.4 (82.8 to 91.1)
	Postmenopausal women	25.3%	12	2	7	46	67	85.7 (57.2 to 98.2)	86.8 (74.7 to 94.5)
Target condition: ovarian borderline tumours – higher-stage tumours excluded									
Chan *et al.* (2013)[82]	All women	7.4%/25.3%	9	7	41	281	338	56.3 (29.9 to 80.2)	87.3 (83.1 to 90.7)
	Premenopausal women	7.4%	6	2	34	235	277	75.0 (34.9 to 96.8)	87.4 (82.8 to 91.1)
	Postmenopausal women	25.3%	12	2	7	46	67	85.7 (57.2 to 98.2)	86.8 (74.7 to 94.5)

TN, true negative; TP, true positive.

One study[82] assessed the variation in the performance of the ROMA score with different stages of epithelial ovarian cancer (see *Table 8*). The sensitivity estimate was highest (92.1%, 95% CI 78.6% to 98.3%) when the target condition was stage III/IV epithelial ovarian cancer, and women with stage I/II and borderline disease were excluded from the analysis.[82] There was a small, but non-significant, fall in sensitivity (82.6%, 95% CI 61.2% to 95.0%) when the target condition was stage I/II epithelial ovarian cancer and women with borderline and higher-stage disease were excluded from the analysis.[82] When the target condition was borderline epithelial tumours and all women with higher-stage disease were excluded from the analysis, the sensitivity estimate was significantly lower (56.3%, 95% CI 29.9% to 80.2%).[82] These data are consistent with the observation that the proportion of women with a negative ROMA score is higher among those women with borderline disease than among those with ovarian malignancies, and may also be higher among those with lower-stage epithelial ovarian cancer than those with higher-stage epithelial ovarian cancer.

Two studies[81,96] reported accuracy data for ovarian malignancy, but without clarifying whether or not the definition of malignancy included borderline tumours (see *Appendix 5*, *Table 43*). Accuracy data for thresholds other than those recommended by the manufacturer (7.4% in premenopausal women and 25.3% in postmenopausal women) are reported in *Appendix 5*, *Table 37*; no study reported accuracy data at an alternative threshold for the inclusive target condition of all malignant tumours, including borderline, and no alternative threshold offered a clear performance advantage.

Accuracy of the Risk of Ovarian Malignancy Algorithm score using Fujirebio Diagnostics' tumour marker assays

None of the included studies used the Fujirebio Diagnostics' LUMIPULSE G automated CEIA system; hence, there were no studies of the ROMA score, using Fujirebio Diagnostics' assays, that met the inclusion criteria for this assessment. Two studies[94,98] that evaluated a ROMA score based on manual Fujirebio Diagnostics' tumour marker EIAs have been included in this report. These studies are included for information only. Both of these studies included all women in the analysis, regardless of their final histopathological diagnosis (target condition: all malignant tumours, including borderline). One study[98] reported a direct comparison of the ROMA score with the RMI 1 (threshold of 200). The results of these studies are provided in *Appendix 5*, *Tables 41* and *42*.

Accuracy of the Risk of Ovarian Malignancy Algorithm score using Roche Diagnostics' tumour marker assays

Only one[89] of the five[89,95,97,102,104] ROMA score studies, which used Roche Diagnostics' Elecsys tumour marker assays, reported a direct comparison of the ROMA score with the RMI 1 (threshold of 200). This study classified women found to have borderline ovarian tumours as disease negative and included women whose final histopathological diagnoses were epithelial ovarian cancer, non-epithelial ovarian cancer and metastases from non-ovarian primaries (target condition: all malignant tumours).[89] This study may be considered to be more applicable to clinical practice if it is considered to be preferable to manage women with borderline tumours in non-specialist settings. In these women, the sensitivity estimate for the ROMA score appeared to be slightly higher than that for the RMI 1 (83.8%, 95% CI 73.4% to 91.3%, vs. 78.4%, 95% CI 67.3% to 87.1%), and the specificity estimate for the ROMA score appeared to be slightly lower than that for the RMI 1 (68.8%, 95% CI 61.6% to 75.4%, vs. 79.6%, 95% CI, 73.1% to 85.1%), but neither difference was statistically significant.[89] A similar pattern was observed when data were stratified by menopausal status (*Table 9*). The same study also reported test performance data, whereby eight women (3%) with non-epithelial ovarian cancer and non-ovarian primaries were excluded from the analysis. This exclusion did not significantly change the test performance estimates for either the ROMA score or the RMI 1 (see *Table 9*). Although the number involved was small, it should be noted that women with malignancies other than epithelial ovarian cancer accounted for four (50%) of the FN results, using the ROMA score, and for three (37.5%) of the results, using the RMI 1.[89]

TABLE 9 Comparative accuracy of the ROMA score using Roche Diagnostics' tumour marker assays vs. the RMI 1

Study (year of publication)	Subgroup	ROMA threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)	RMI 1	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours – borderline tumours classified as disease negative																	
Yanaranop *et al.* (2016)[89]	All women	11.4%/29.9%	62	12	58	128	260	83.8 (73.4 to 91.3)	68.8 (61.6 to 75.4)	200	58	16	38	148	260	78.4 (67.3 to 87.1)	79.6 (73.1 to 85.1)
	Premenopausal women	11.4%	24	4	35	85	148	85.7 (67.3 to 96.0)	70.8 (61.8 to 78.8)	200	21	7	23	97	148	75.0 (55.1 to 89.3)	80.8 (72.6 to 87.4)
	Postmenopausal women	29.9%	38	8	23	43	112	82.6 (68.6 to 92.2)	65.2 (52.4 to 76.5)	200	37	9	15	51	112	80.4 (66.1 to 90.6)	77.3 (65.3 to 86.7)
Target condition: epithelial ovarian malignancies – borderline tumours classified as disease negative																	
Yanaranop *et al.* (2016)[89]	All women	11.4%/29.9%	58	8	58	128	252	87.9 (77.5 to 94.6)	68.8 (61.6 to 75.4)	200	53	13	38	148	252	80.3 (68.7 to 89.1)	79.6 (73.1 to 85.1)
Target condition: epithelial ovarian malignancies (stage I) – borderline tumours classified as disease negative and higher-stage tumours excluded																	
Yanaranop *et al.* (2016)[89]	All women	11.4%/29.9%	23	7	58	128	216	76.7 (57.7 to 90.1)	68.8 (61.6 to 75.4)	200	21	9	38	148	216	70.0 (50.6 to 85.3)	79.6 (73.1 to 85.1)
Target condition: epithelial ovarian malignancies (stages II–IV) – borderline tumours classified as disease negative and stage I tumours excluded																	
Yanaranop *et al.* (2016)[89]	All women	11.4%/29.9%	35	1	58	128	222	97.2 (85.5 to 99.9)	68.8 (61.6 to 75.4)	200	32	4	38	148	222	88.9 (73.9 to 96.9)	79.6 (73.1 to 85.1)

TN, true negative; TP, true positive.

The aforementioned comparative accuracy study[89] also assessed the variation in the performance of the ROMA score with different stages of epithelial ovarian cancer (see *Table 9*). The sensitivity estimate was highest for both the ROMA score (97.2%, 95% CI 95.5% to 99.9%) and the RMI 1 score (88.9%, 95% CI 73.9% to 96.9%), for which the target condition was stages II–IV epithelial ovarian cancer; women with stage I disease were excluded from the analysis.[89] As with the ROMA score using Abbott Diagnostics' ARCHITECT tumour marker assay, sensitivity estimates were lower for both the ROMA score (76.7%, 95% CI 57.7% to 90.1%) and the RMI 1 score (70.0%, 95% CI 50.6% to 85.3%), for which the target condition was stage I epithelial ovarian cancer; women with higher-stage disease were excluded from the analysis.[89] This indicates that the proportion of women with a negative ROMA score may be higher among those with lower-stage epithelial ovarian cancer than those with higher-stage epithelial ovarian cancer.

Two[97,104] of the four[95,97,102,104] additional studies that evaluated the performance of the ROMA score but did not provide a comparison with the RMI 1 score included all study participants in the analysis regardless of their final histopathological diagnoses [target condition: all malignant tumours including borderline (*Table 10*)]. The summary estimate of the sensitivity of the ROMA score (79.1%, 95% CI 74.2% to 83.5%), derived from these two studies, was lower than that reported by the comparative accuracy study[89] described earlier, in which women with borderline tumours were classified as disease negative, and the summary specificity estimate was also lower (79.1%, 95% CI 76.3% to 81.6%), but these differences were not statistically significant. Two studies[95,97] reported test performance data for the ROMA score, in which women found to have borderline tumours and those with non-ovarian primaries were excluded from the analyses. The sensitivity estimates derived from these two studies were very different (see *Table 10*) hence, no summary estimates were calculated. One of these studies[97] reported test performance estimates calculated both with and without the inclusion of women with borderline tumours and those with non-ovarian primaries. As with the ROMA score, using Abbott Diagnostics' ARCHITECT tumour marker assays, these data indicated that women with borderline tumours and those with non-ovarian primaries accounted for a high proportion (12/14; 86%) of the FN risk scores observed.[97]

One study[102] provided performance estimates for the ROMA score, using Roche Diagnostics' Elecsys tumour marker assays, for the target condition 'ovarian malignancy', when it was not clear whether or not the definition of malignancy included borderline tumours (see *Appendix 5*, *Table 44*). Accuracy data for thresholds other than those recommended by the manufacturer (11.4% in premenopausal women and 29.9% in postmenopausal women) are reported in *Appendix 5*, *Table 37*; no study reported accuracy data at an alternative threshold for the inclusive target condition of all malignant tumours including borderline, and no alternative threshold offered a clear performance advantage.

Between-assay comparisons

No study assessed variation in the performance of the ROMA score with the use of different manufacturers' tumour marker assays. However, between-study comparisons indicate that, when all study participants were included in the analyses regardless of final histopathological diagnosis (target condition: all malignant tumours including borderline), the estimates of sensitivity did not differ significantly between the two manufacturers' assays for which data were available (*Figure 3*). The sensitivity estimate for the ROMA score using Abbott Diagnostics' ARCHITECT tumour marker assay was 75.0% (95% CI 60.4% to 86.4%), derived from one study,[103] compared with 79.1% (95% CI 74.2% to 83.5%) using Roche Diagnostics' Elecsys tumour marker assay, derived from two studies.[97,104] However, the specificity estimate for Abbott Diagnostics' ARCHITECT tumour marker assay (87.9%, 95% CI 81.9% to 92.4%) was higher than that for Roche Diagnostics' Elecsys tumour marker assay (79.1%, 95% CI 76.3% to 81.6%). There were no studies of the ROMA score using Fujirebio Diagnostics' tumour marker assays that met the inclusion criteria for this assessment. There were insufficient data to compare the performance of the ROMA score with the use of different manufacturers' tumour marker assays for detecting different stages of disease.

TABLE 10 Accuracy of the ROMA score using Roche Diagnostics' tumour marker assays at the manufacturer's recommended thresholds

Study (year of publication)	Subgroup	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity %, (95% CI)
Target condition: all malignant tumours including borderline									
[a]Janas *et al.* (2015)[97]	All women	11.4%/29.9%	52	14	39	154	259	78.8 (67.0 to 87.9)	79.8 (73.4 to 85.2)
Shulman *et al.* (2016)[104]	All women	11.4%/29.9%	194	51	158	590	993	79.2 (73.7 to 83.8)	78.9 (75.8 to 81.7)
Summary estimates								**79.1 (74.2 to 83.5)**	**79.1 (76.3 to 81.6)**
[a]Janas *et al.* (2015)[97]	Premenopausal women	11.4%	9	1	22	100	132	90.0 (55.5 to 99.7)	82.0 (74.0 to 88.3)
[a]Janas *et al.* (2015)[97]	Postmenopausal women	29.9%	44	12	17	54	127	78.6 (65.6 to 88.4)	76.1 (64.5 to 88.4)
Target condition: ovarian malignancies excluding borderline									
[a]Janas *et al.* (2015)[97]	All women	11.4%/29.9%	42	2	39	154	237	95.5 (84.5 to 99.4)	79.8 (73.4 to 85.2)
Xu *et al.* (2016)[95]	All women	11.4%/29.9%	113	97	39	272	521	53.8 (46.8 to 60.7)	87.5 (83.3 to 90.9)
[a]Janas *et al.* (2015)[97]	Premenopausal women	11.4%	6	0	22	100	128	100 (54.1 to 100)	82.0 (74.0 to 88.3)
Xu *et al.* (2016)[95]	Premenopausal women	11.4%	56	51	38	226	371	54.9 (42.5 to 62.1)	85.6 (80.8 to 89.6)
[a]Janas *et al.* (2015)[97]	Postmenopausal women	29.9%	36	2	17	54	109	94.7 (82.3 to 99.4)	76.1 (64.5 to 85.4)
Xu *et al.* (2016)[95]	Postmenopausal women	29.9%	57	46	1	46	150	53.3 (45.2 to 65.1)	97.9 (88.7 to 99.9)

TN, true negative; TP, true positive.
a 2 × 2 data were calculated (other studies reported 2 × 2 data).

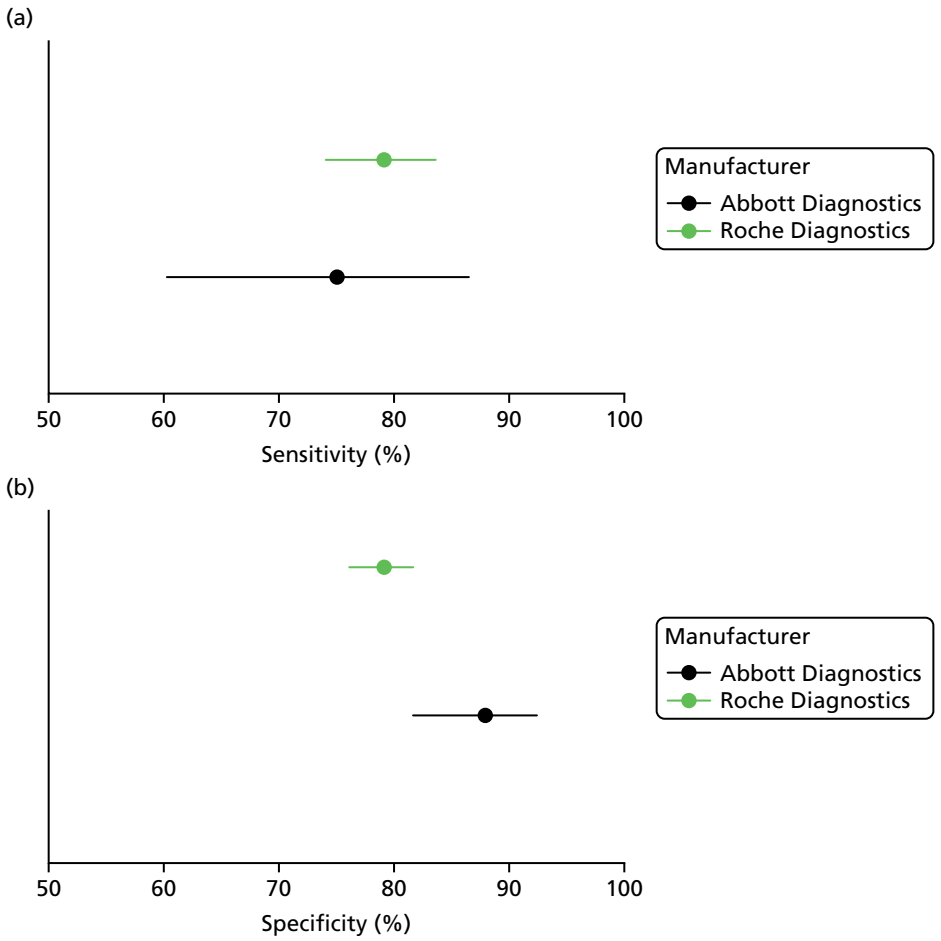

FIGURE 3 Comparison of the accuracy of the ROMA score using Abbott Diagnostics' ARCHITECT vs. the ROMA score using Roche Diagnostics' Elecsys (target condition: all malignant tumours including borderline). (a) Sensitivity estimate; and (b) specificity estimate.

Diagnostic performance of the International Ovarian Tumour Analysis group's simple ultrasound rules and the Assessment of Different NEoplasias in the adneXa model

Details of the Assessment of Different NEoplasias in the adneXa studies

Six published studies[17,42–46] and one unpublished interim report (Frances Nixon, personal communication) provided data on the diagnostic performance of the ADNEX scores at different thresholds. All studies reported accuracy data for the validated 10% decision threshold to identify women with an adnexal mass who were at a high risk of developing ovarian cancer and used a version of the ADNEX model that included a measurement of CA125 level. Four of the six published studies did not report any details of the experience of those performing the ultrasound examinations.[42–45] One study[46] reported that ultrasound examinations were performed by European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) level 2 ultrasound examiners (non-consultant gynaecology specialist, gynaecology trainee doctors and gynaecology sonographers), and the remaining study[17] used EFSUMB level 2/3 practitioners with 8–20 years' experience in gynaecological sonography. (Confidential information has been removed.)

This section reports accuracy data for only the 10% threshold. Three studies[17,43,46] provided accuracy data for additional thresholds and these are reported in *Appendix 5*, *Table 38*. All studies in this section were conducted in Europe; one study[45] was conducted solely in the UK, and two were multicentre studies[17,46] that included UK participants.

The target condition for this assessment is ovarian cancer, including conditions covered by the NICE clinical guideline CG122[1] (i.e. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma and borderline ovarian cancer). All studies in this section include women with one or more adnexal mass, and all but one study[45] included borderline tumours in their definition of malignancy; the study that did not was reported only as a conference abstract and it was not clear whether or not any borderline tumours were included (see *Appendix 5*, *Table 45*). Three published studies[17,44,46] and the unpublished interim report (Frances Nixon, personal communication) included participants with 'other malignancies', metastases from non-ovarian sites and 'non-ovarian cancers'. When the target condition was described as 'all ovarian malignancy', those participants whose postoperative histological diagnosis identified a non-ovarian primary were excluded from the estimates of test performance. Conversely, when the target condition was described as 'all malignant tumours', participants with a non-ovarian primary were not excluded and were classified as disease positive; this could potentially include participants with any tumour on the ovaries that has metastasised from another primary (e.g. CRC) and/or participants with an adnexal/pelvic mass that turns out to be non-ovarian (not clearly specified by the included studies). Full details of the final histopathological diagnoses of study participants who had a malignant mass are reported in *Appendix 4*, *Table 36*.

Accuracy of the Assessment of Different NEoplasias in the adneXa model for determining high risk of ovarian cancer

Three published studies[17,44,46] and the unpublished interim report (Frances Nixon, personal communication) included all participants in the analysis, regardless of their final histopathological diagnosis (target condition: all malignant tumours including borderline). The summary estimate of sensitivity derived from these studies was 96.3% (95% CI 95.3% to 97.1%) and the summary estimate of specificity was 69.1% (95% CI 67.4% to 70.8%; *Table 11*). These estimates did not differ significantly from those calculated from only those studies of the ADNEX model that reported a direct comparison with the RMI 1 score (at a threshold of 200 or 250; see *Table 14*). Two further studies,[42,43] reporting three data sets, excluded women with histopathological diagnoses other than primary ovarian cancer. The summary estimate of sensitivity (94%, 95% CI 88.6% to 97.4%) derived from these studies did not differ significantly from that derived from the studies that included all participants in their analyses. However, the summary estimate of specificity (77.6%, 95% CI 73.6% to 81.2%) was higher. One study,[42] which reported results from two separate cohorts (Spain and Poland), also reported accuracy data stratified by menopausal status. Menopausal status did not significantly affect sensitivity; however, the specificity estimate was significantly higher in premenopausal women than in postmenopausal women (see *Table 11*).

Accuracy data for thresholds other than the 10% validated threshold (1%, 3%, 5%, 15%, 20% and 30%) are reported in *Appendix 5*, *Table 38*. As might be expected, sensitivity estimates increase and specificity estimates decrease with decreasing threshold.

Details of the International Ovarian Tumour Analysis group's simple ultrasound rules studies

Seventeen published studies[44,47–52,55,58–65,67] and the unpublished interim report (Frances Nixon, personal communication) provided data on the diagnostic performance of the IOTA group's simple ultrasound rules, for the identification of women with an adnexal mass who are at a high risk of developing ovarian cancer. The majority (11/17) of the published studies[44,48–50,52,58,60,62–64,66] were conducted in Europe; three of the

TABLE 11 Accuracy of the ADNEX model at a threshold of 10%

Study (year of publication)	Subgroup	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline									
IOTA5 (2017)[a,b]	All women	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Meys *et al.* (2016)[44]		113	2	80	131	326	Calculated	98.0 (93.0 to 100)	62.0 (55.0 to 68.0)
Sayasneh *et al.* (2016)[46]		177	5	138	290	610	Calculated	97.3 (93.5 to 98.9)	67.7 (63.0 to 72.0)
[b]Van Calster *et al.* (2014)[17]		946	34	408	1015	2403	Calculated	96.5 (95.2 to 97.6)	71.3 (68.9 to 73.7)
Summary estimates								**96.3 (95.3 to 97.1)**	**69.1 (67.4 to70.8)**
Meys *et al.* (2016)[44]	Premenopausal women	31	0	28	69	128	Calculated	100 (86.0 to 100)	71.0 (61.0 to 80.0)
Meys *et al.* (2016)[44]	Postmenopausal women	82	2	52	62	198	Calculated	98.0 (91.0 to 100)	54.0 (44.0 to 63.0)
Target condition: ovarian malignancies including borderline									
Joyeux *et al.* (2016)[43]	All women	27	3	48	206	284	Calculated	90 (73.5 to 97.9)	81.1 (75.7 to 85.7)
Szubert *et al.* (2016)[42]	All women – Poland	66	4	37	97	204	Reported	94.3 (88.5 to 98.7)	72.4 (65.1 to 79.7)
	All women – Spain	33	1	22	67	123	Reported	97.1 (89.7 to 100)	75.3 (65.2 to 84.7)

continued

TABLE 11 Accuracy of the ADNEX model at a threshold of 10% (*continued*)

Study (year of publication)	Subgroup	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Summary estimates								**94 (88.6 to 97.4)**	**77.6 (73.6 to 81.2)**
Szubert *et al.* (2016)[42]	Premenopausal women – Poland	29	3	23	83	138	Calculated	90.6 (77.0 to 100)	78.3 (70.7 to 85.9)
	Premenopausal women – Spain	15	0	11	51	66	Calculated	100 (78.2 to 100)	82.3 (71.6 to 91.1)
Summary estimates								**93.6 (82.5 to 98.7)**	**79.8 (72.9 to 85.6)**
Szubert *et al.* (2016)[42]	Postmenopausal women – Poland	37	1	14	14	77	Calculated	97.4 (91.7 to 100)	50.0 (32.1 to 69.8)
	Postmenopausal women – Spain	18	1	11	16	46	Calculated	95.8 (85.7 to 100)	59.3 (41.5 to 77.6)
Summary estimates								**96.5 (87.9 to 99.6)**	**54.5 (40.6 to 68)**

TN, true negative; TP, true positive.
a Frances Nixon, personal communication.
b Data from the IOTA cohort.

studies were conducted in the UK.[60,62,66] Two further studies[61,65] were worldwide, with multinational studies including UK participants. Two of the remaining studies were conducted in Thailand[47,51] and one was conducted in Brazil.[55] One study did not report sufficient detail to determine the geographic location.[53] (Confidential information has been removed.)

Three published studies[50,61,65] were clearly conducted by the IOTA study core group, using data from various phases of the IOTA study; only one report was included for each phase of the IOTA study. Phase 5 of the IOTA study is ongoing and an interim report was supplied to this assessment (confidential information has been removed) (Frances Nixon, personal communication).

Ten published studies,[44,48–50,52,55,58,61,62,65] as well as the unpublished interim report (Frances Nixon, personal communication), included all participants in the analysis; participants with inconclusive IOTA group's simple ultrasound rules assessments were either assumed to have malignant tumours or classified by subjective assessment of ultrasound images. This section reports data for studies in which all participants were included in the analysis. Six further studies[47,51,53,63,64,66] excluded participants with inconclusive IOTA group's simple ultrasound rules assessments from their analyses. The results of these studies are provided in *Appendix 5*, *Table 39*. One study[60] did not report sufficient information to determine how participants with inconclusive IOTA group's simple ultrasound rules assessments were handled.

Accuracy of the International Ovarian Tumour Analysis group's simple ultrasound rules for determining a high risk of developing ovarian cancer

All studies in this section included all participants in their analyses, regardless of their final histopathological diagnosis (target condition: all malignant tumours including borderline). Eight published studies[44,48–50,52,55,62,65] and the unpublished interim report (Frances Nixon, personal communication) provided accuracy data for the IOTA group's simple ultrasound rules, whereby women with inconclusive assessments were assumed to have malignant tumours. The summary estimate of sensitivity derived from these studies was 94.2% (95% CI 93.3% to 95.1%) and the summary estimate of specificity was 76.1% (95% CI 74.9% to 77.3%). These estimates did not differ significantly from those calculated from only those studies of the IOTA group's simple ultrasound rules, in which participants with inconclusive assessments were assumed to have malignant tumours, which reported a direct comparison with the RMI 1 score (at a threshold of 200 or 250; see *Tables 14* and *15*). Four studies[44,49,50,62] of these studies reported accuracy data stratified by menopausal status. Menopausal status did not significantly affect sensitivity; however, the specificity estimate was significantly higher in premenopausal women than in postmenopausal women (*Table 12*).

Seven studies[44,49,50,52,58,62,65] provided accuracy data for the IOTA group's simple ultrasound rules, whereby participants with inconclusive assessments were classified by an expert subjective assessment. In this analysis, only studies in which the subjective assessment was done by experts or by level 2/3 examiners as per the EFSUMB classification system have been included. The summary estimates of sensitivity and specificity derived from these studies were 88.4% (95% CI 86.9% to 89.8%) and 92.5% (95% CI 91.6% to 93.4%), respectively (see *Table 12*). One of these studies[49] also assessed the effect of the training level of examiners on the diagnostic performance of the IOTA group's simple ultrasound rules and found no significant differences in test performance between EFSUMB level 2 examiners (see *Table 12*) and EFSUMB level 1 examiners (*Table 13*). However, it should be noted that all examiners received 1 half-day of practical training in the IOTA group's Simple Rules before the study.

Five of these studies[44,49,50,58,62] reported accuracy data stratified by menopausal status. Menopausal status did not significantly affect sensitivity; however, the specificity estimate was significantly higher in premenopausal women than in postmenopausal women (see *Table 12*). One study[58] (see *Appendix 5*, *Table 39*) also assessed

TABLE 12 Accuracy of the IOTA group's simple ultrasound rules, whereby inconclusive results were assumed to be malignant or classified by subjective assessment

Threshold	Study (year of publication)	Subgroup	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline										
Malignant (inconclusive results were treated as malignant)	Adballa *et al.* (2013)[48]	All women	16	1	7	63	87	Reported	94.1 (71.3 to 99.9)	90.0 (80.5 to 95.9)
	Alcazar *et al.* (2013)[52]		51	4	54	231	340	Reported	92.7 (82.4 to 98.0)	81.1 (76.0 to 85.4)
	IOTA5 2017[a,b]		Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
	Knafel *et al.* (2016)[49]		78	4	15	129	226	Reported	95.1 (88.0 to 98.7)	89.6 (83.4 to 94.1)
	Meys *et al.* (2016)[44]		107	8	67	144	326	Calculated	93.0 (86.0 to 97.0)	68.0 (61.0 to 70.0)
	Sayasneh *et al.* (2013)[62]		67	7	24	157	255	Calculated	91.0 (82.0 to 95.0)	87.0 (82.0 to 91.0)
	Silvestre *et al.* (2015)[55]		32	0	17	26	75	Reported	100 (89.1 to 100)	60.5 (44.4 to 75.0)
	[b]Testa *et al.* (2014)[50]		934	46	369	1054	2403	Calculated	95.3 (93.1 to 96.19)	74.1 (67.7 to 79.7)
	[b]Timmerman *et al.* (2010)[65]		515	27	307	1089	1938	Calculated	95.0 (92.0 to 96.0)	78.0 (75.0 to 80.0)
	Summary estimates								**94.2 (93.3 to 95.1)**	**76.1 (74.9 to 77.3)**
	Knafel *et al.* (2016)[49]	Premenopausal women	32	1	9	101	143	Calculated	96.9 (84.2 to 99.9)	91.9 (85.0 to 96.2)
	Meys *et al.* (2016)[44]		29	2	23	74	128	Calculated	94.0 (77.0 to 99.0)	76.0 (66.0 to 84.0)
	Sayasneh *et al.* (2013)[62]		24	4	16	121	165	Calculated	86.0 (69.0 to 94.0)	88.0 (83.0 to 93.0)
	[b]Testa *et al.* (2014)[50]		359	19	225	751	1354	Calculated	95.0 (91.0 to 97.0)	77.0 (70.0 to 83.0)

Threshold	Study (year of publication)	Subgroup	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
	Summary estimates								**94.5 (92.0 to 96.4)**	**79.3 (77.0 to 81.5)**
	Knafel *et al.* (2016)[49]	Postmenopausal women	46	3	6	28	83	Calculated	94 (83.1 to 98.7)	81.8 (65.5 to 93.2)
	Meys *et al.* (2016)[44]		78	6	44	70	198	Calculated	93.0 (85.0 to 97.0)	61.0 (52.0 to 70.0)
	Sayasneh *et al.* (2013)[62]		43	3	7	37	90	Calculated	93.0 (82.0 to 98.0)	84.0 (71.0 to 92.0)
	[b]Testa *et al.* (2014)[50]		578	24	152	295	1049	Calculated	96.0 (93.0 to 97.0)	66.0 (59.0 to 73.0)
	Summary estimates								**95.4 (93.7 to 96.8)**	**67.3 (63.5 to 70.9)**
Malignant (inconclusive results were classified by expert SA)	Alcázar *et al.* (2013)[52]	All women	49	6	11	274	340	Reported	89.1 (77.8 to 95.9)	96.1 (93.2 to 98.1)
Malignant (inconclusive results were classified by level 2 or level 3 by expert SA)	Knafel *et al.* (2016)[49]		78	4	9	135	226	Calculated	95.1 (88.0 to 98.7)	93.8 (88.5 to 97.1)
	Meys *et al.* (2016)[44]		102	13	21	190	326	Calculated	89.0 (81.0 to 94.0)	90.0 (85.0 to 94.0)
	Piovano *et al.* (2016)[58]		69	15	23	284	391	Calculated	82.1 (72.3 to 89.6)	92.5 (89 to 95.2)
	Sayasneh *et al.* (2013)[62]		64	10	11	170	255	Calculated	86.0 (77.0 to 92.0)	94.0 (90.0 to 97.0)
	[b]Testa *et al.* (2014)[50]		900	80	157	1266	2403	Calculated	91.8 (89.1 to 93.9)	89.0 (85.2 to 92.0)
	[b]Timmerman *et al.* (2010)[65]		494	102	48	1294	1938	Calculated	91.0 (88.0 to 93.0)	93.0 (91.0 to 94.0)

continued

TABLE 12 Accuracy of the IOTA group's simple ultrasound rules, whereby inconclusive results were assumed to be malignant or classified by subjective assessment (*continued*)

Threshold	Study (year of publication)	Subgroup	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
	Summary estimates								**88.4 (86.9 to 89.8)**	**92.5 (91.6 to 93.4)**
	Knafel *et al.* (2016)[49]	Premenopausal women	32	1	5	105	143	Calculated	96.9 (84.2 to 99.9)	95.5 (89.7 to 98.5)
	Meys *et al.* (2016)[44]		27	4	4	93	128	Calculated	87.0 (69.0 to 96.0)	96.0 (89.0 to 99.0)
	Piovano *et al.* (2016)[58]		18	3	6	194	221	Calculated	86.0 (71.0 to 100)	97.0 (94.0 to 99.0)
	Sayasneh *et al.* (2013)[62]		23	5	5	132	165	Calculated	82.0 (64.0 to 92.0)	96.0 (91.0 to 98.0)
	[b]Testa *et al.* (2014)[50]		348	24	88	888	1354	Calculated	92.0 (86.0 to 95.0)	91.0 (87.0 to 94.0)
	Summary estimates								**92.4 (89.6 to 94.6)**	**92.9 (91.5 to 94.1)**
	Knafel *et al.* (2016)[49]	Postmenopausal women	46	3	1	33	83	Calculated	94.0 (83.1 to 98.7)	97.9 (84.7 to 99.9)
	Meys *et al.* (2016)[44]		75	9	39	75	198	Calculated	89.0 (80.0 to 95.0)	85.0 (77.0 to 91.0)
	Piovanono *et al.* (2016)[58]		51	12	17	90	170	Calculated	81.0 (71.0 to 91.0)	84.0 (77.0 to 91.0)
	Sayasneh *et al.* (2013)[62]		41	5	4	40	90	Calculated	89.0 (77.0 to 95.0)	91.0 (79.0 to 96.0)
	[b]Testa *et al.* (2014)[50]		560	42	76	371	1049	Calculated	93.0 (90.0 to 95.0)	83.0 (78.0 to 87.0)
	Summary estimates								**91.6 (89.5 to 93.4)**	**81.6 (78.7 to 84.4)**

SA, subjective assessment; TN, true negative; TP, true positive.
a Frances Nixon, personal communication.
b Data from the IOTA cohort.

TABLE 13 Accuracy of the IOTA group's simple ultrasound rules using a EFSUMB level 1 examiner

Threshold	Study (year of publication)	Subgroup	Index test variations	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline											
Malignant (inconclusive were treated as malignant)	Knafel *et al.* (2016)[49]	All women	Level 1 examiner	79	3	26	118	226	Reported	96.3 (89.7 to 99.2)	81.9 (74.7 to 87.9)
	Knafel *et al.* (2016)[49]	Postmenopausal women	Level 1 examiner	46	3	12	22	83	Calculated	94 (83.1 to 98.7)	63.6 (46.5 to 80.3)
		Premenopausal women	Level 1 examiner	33	0	14	96	143	Calculated	100 (89.4 to 100)	87.4 (79.6 to 92.9)
Malignant (inconclusive were classified by expert SA)	Knafel *et al.* (2016)[49]	All women	Level 1 examiner	79	3	7	137	226	Calculated	96.3 (89.7 to 99.2)	95.1 (90.2 to 98.0)
		Postmenopausal women	Level 1 examiner	46	3	3	31	83	Calculated	93.9 (83.1 to 98.7)	90 (76.3 to 98.1)
		Premenopausal women	Level 1 examiner	33	0	4	106	143	Calculated	100 (89.4 to 100)	96.4 (91.0 to 99.0)

SA, subjective assessment; TN, true negative; TP, true positive.

whether or not the addition of biomarkers to the IOTA group's simple ultrasound rules could improve the diagnostic performance. When a positive index test was defined as a malignant classification by the IOTA group's simple ultrasound rules (with subjective assessment of inconclusives) and a ROMA score of > 11.4% out of 29.9%, the sensitivity and specificity estimates were 90.5% (95% CI 82.1% to 95.8%) and 80.1% (95% CI 75.2% to 84.4%), respectively. When a positive index test was defined as a malignant classification by the IOTA group's simple ultrasound rules (with subjective assessment of inconclusives) and a HE4 level of ≥ 70 out of 140, the sensitivity and specificity estimates were 86.9% (95% CI 77.8% to 93.3%) and 86.3% (95% CI 82% to 90%), respectively. Neither the addition of the ROMA score nor the addition of HE4 level alone significantly affected estimates of test performance. Finally, when a positive index test was defined as a malignant classification by the IOTA group's simple ultrasound rules (with subjective assessment of inconclusives) and a CA125 level of ≥ 35, the sensitivity estimate was similar to that for the IOTA group's simple ultrasound rules (90.5%, 95% CI 82.1% to 95.8%); however, the specificity estimate was significantly lower (68.1%, 95% CI 62.5% to 73.3%).

Comparison of the different methods of operationalising the IOTA group's simple ultrasound rules (i.e. 'inconclusive results treated as malignant' vs. 'inconclusive results were classified by an expert') indicates that sensitivity estimates were significantly higher when inconclusive results were treated as malignant, whereas specificity was significantly higher when patients with inconclusive results were classified by an expert. Thus, as might be expected, applying the assumption that all patients with an inconclusive result have a malignant tumour is likely to result in fewer patients with ovarian cancer being missed (FNs), whereas expert reassessment of inconclusive results is likely to result in fewer unnecessary referrals (FPs).

Comparisons between the International Ovarian Tumour Analysis group's Simple Rules, the Assessment of Different NEoplasias in the adneXa and the Risk of Malignancy Index 1

This assessment is primarily concerned with providing a comparison between the RMI 1,[78] used with a decision threshold of 250 (the current standard practice in the NHS[1]) and the specified alternative risk-scoring methods (see *Chapter 2, Intervention technologies*). Our searches did not identify any studies that reported a direct comparison (both tests were used to assess the same patient cohort) between the ADNEX model or the IOTA group's simple ultrasound rules and the RMI 1, used with a decision threshold of 250. One published study[44] and the unpublished interim report (Frances Nixon, personal communication) reported direct comparisons between the ADNEX model at the 10% threshold, the IOTA group's simple ultrasound rules (whereby patients with an inconclusive assessment were assumed to have malignant tumours) and the RMI 1, used with a decision threshold of 200 (*Table 14*). Both of these studies included all participants in the analysis, regardless of their final histopathological diagnosis (target condition: all malignant tumours including borderline). The summary estimates of sensitivity derived from these two studies were slightly higher for the ADNEX model (96%, 95% CI 94.5% to 97.1%) than for the IOTA group's simple ultrasound rules (92.8%, 95% CI 90.9% to 94.3%). The summary estimates of specificity were similar (67%, 95% CI 64.2% to 69.6%, and 71.6%, 95% CI 68.9% to 74.1%) for the ADNEX model and the ultrasound Simple Rules, respectively. The summary estimate of sensitivity for the RMI 1 at a decision threshold of 200 (66%, 95% CI 62.9% to 69%) was significantly lower than that for both the ADNEX model estimate and the IOTA group's simple ultrasound rules estimate. Conversely, the specificity estimate for the RMI 1 at a decision threshold of 200 was significantly higher (89%, 95% CI 87% to 90.7%) than that for both the ADNEX model estimate and the IOTA group's simple ultrasound rules estimate (*Figure 4*). The unpublished interim report (Frances Nixon, personal communication) also reported direct comparisons between the ADNEX model at the 10% threshold, the IOTA group's simple ultrasound rules (whereby patients with an inconclusive assessment were assumed to have

malignant tumours) and the RMI 1, used with a decision threshold of 250. The comparative accuracy estimates at this threshold did not differ from those at 200 (see *Table 14*).

Only the published study[44] reported accuracy data stratified by menopausal status. In premenopausal women, the ADNEX model at the 10% threshold and the IOTA group's simple ultrasound rules had similar sensitivities of 100% (95% CI 86% to 100%) and 94% (95% CI 77% to 99%), respectively, in comparison with the overall population. The specificities were significantly lower at 71% (95% CI 61% to 80%) and 76% (95% CI 66% to 84%), respectively. In postmenopausal women, the ADNEX model at the 10% threshold and the IOTA group's simple ultrasound rules had similar sensitivities of 98% (95% CI 91% to 100%) and 93% (95% CI 85% to 97%), respectively, in comparison with the overall population. The specificities of 54% (95% CI 44% to 63%) and 61% (95% CI 52% to 70%), respectively, were significantly lower than those of the overall population. The RMI 1, using a decision threshold of 200, had a significantly lower sensitivity of 42% (95% CI 25% to 61%) and a significantly higher specificity of 94% (95% CI 86% to 97%) in premenopausal women. Conversely, the sensitivity estimate was higher at 82% (95% CI 72% to 89%) and the specificity estimate was lower at 66% (95% CI 56% to 74%) in postmenopausal women (but this was not significant).

Four published studies[44,48,50,62] and the unpublished interim report (Frances Nixon, personal communication) reported direct comparisons between the IOTA group's simple ultrasound rules (whereby patients with an inconclusive assessment were assumed to have malignant tumours) and the RMI 1 used with a decision threshold of 200 (*Table 15*). All of these studies included all participants in the analysis, regardless of their final histopathological diagnosis (target condition: all malignant tumours including borderline). The summary estimate of sensitivity for the IOTA group's simple ultrasound rules (93.9%, 95% CI 92.8% to 94.9%) was significantly higher than that for the RMI 1 (66.9%, 95% CI 64.8% to 68.9%). Conversely, the summary estimate of specificity for the IOTA group's simple ultrasound rules (74.2%, 95% CI 72.6% to 75.8%) was significantly lower than that for the RMI 1 (90.1%, 95% CI 88.9% to 91.2%).

Three of the above studies[44,50,62] also reported comparative accuracy data for the IOTA group's simple ultrasound rules versus the RMI 1 (threshold of 200), whereby participants with inconclusive IOTA assessments were classified by an expert subjective assessment of the ultrasound images (see *Table 15*). The summary estimate of sensitivity for the IOTA group's simple ultrasound rules (91.2%, 95% CI 89.4% to 92.8%) was significantly higher than that for the RMI 1 (67.8%, 95% CI 65% to 70.4%). Conversely, the summary estimates of specificity were significantly lower for the IOTA group's simple ultrasound rules (89.6%, 95% CI 88.1% to 91%) than for the RMI 1 (98.5%, 95% CI 98.3% to 98.7%). These three studies also reported accuracy data stratified by menopausal status, and the comparative accuracy estimates for both subgroups followed the pattern observed for all participants (see *Table 15*).

In premenopausal women, using the IOTA group's simple ultrasound rules whereby participants with an inconclusive assessment were assumed to have a malignant tumour, the summary estimates of sensitivity and specificity were 94.3% (95% CI 91.7% to 96.3%) and 78.2% (95% CI 75.7% to 80.5%), respectively. These estimates were not significantly different from those for postmenopausal women (95.5%, 95% CI 93.7% to 96.9%, and 72.3%, 95% CI 68.9% to 75.5%, respectively). When participants with inconclusive IOTA group's Simple Rules assessments were classified by an expert subjective assessment, the summary estimates of sensitivity were similar (for both premenopausal women and postmenopausal women) to those obtained when inconclusive assessments were assumed to be malignant (see *Table 15*). However, in both premenopausal women and postmenopausal women, the use of a subjective assessment significantly increased the summary estimate of specificity to 92% (95% CI 90.3%

TABLE 14 Comparative accuracy of the ADNEX model, the IOTA group's simple ultrasound rules and the RMI 1

Study (year of publication)	Subgroup	Index test	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*
Target condition: all malignant tumours including borderline								
Meys *et al.* (2016)[44]	All women	ADNEX model	≥ 10%	113	2	80	131	326
IOTA5 2017[a,b]				Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Summary estimates								
IOTA5 2017[a]	All women	ADNEX model	≥ 10%	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Meys *et al.* (2016)[44]	All women	IOTA group's simple ultrasound rules	Inconclusive = malignant	107	8	67	144	326
IOTA5 2017[a,b]				Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Summary estimates								
IOTA5 2017[a]	All women	IOTA group's simple ultrasound rules	Inconclusive = malignant	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Meys *et al.* (2016)[44]	All women	IOTA group's simple ultrasound rules	Inconclusive = SA	102	13	21	190	326
	Premenopausal women	ADNEX model	≥ 10%	31	0	28	69	128
		IOTA group's simple ultrasound rules	Inconclusive by SA	27	4	4	93	128
			Inconclusive = malignant	29	2	23	74	128
	Postmenopausal women	ADNEX model	≥ 10%	82	2	52	62	198
		IOTA group's simple ultrasound rules	Inconclusive = SA	75	9	39	75	198
			Inconclusive = malignant	78	6	44	70	198

SA, subjective assessment; TN, true negative; TP, true positive.
a Frances Nixon, personal communication.
b Data from the IOTA cohort.

Sensitivity, % (95% CI)	Specificity, % (95% CI)	RMI threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
98.0 (93.0 to 100)	62.0 (55.0 to 68.0)	200	82	33	44	167	326	71.0 (62.0 to 79.0)	79.0 (72.0 to 84.0)
Confidential information has been removed	Confidential information has been removed	200	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
96.0 (94.5 to 97.1)	**67.0 (64.2 to 69.6)**	**Summary estimates**						**66.0 (62.9 to 69.0)**	**89.0 (87.0 to 90.7)**
Confidential information has been removed	Confidential information has been removed	250	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
93.0 (86.0 to 97.0)	68.0 (61.0 to 70.0)	200	82	33	44	167	326	71.0 (62.0 to 79.0)	79.0 (72.0 to 84.0)
Confidential information has been removed	Confidential information has been removed	200	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
92.8 (90.9 to 94.3)	**71.6 (68.9 to 74.1)**	**Summary estimates**						**66.0 (62.9 to 69.0)**	**89.0 (87.0 to 90.7)**
Confidential information has been removed	Confidential information has been removed	250	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
89.0 (81.0 to 94.0)	90.0 (85.0 to 94.0)	200	82	33	44	167	326	71.0 (62.0 to 79.0)	79.0 (72.0 to 84.0)
100 (86.0 to 100)	71.0 (61.0 to 80.0)	200	13	18	6	91	128	42.0 (25.0 to 61.0)	94.0 (86.0 to 97.0)
87.0 (69.0 to 96.0)	96.0 (89.0 to 99.0)								
94.0 (77.0 to 99.0)	76.0 (66.0 to 84.0)								
98.0 (91.0 to 100)	54.0 (44.0 to 63.0)	200	69	15	39	75	198	82.0 (72.0 to 89.0)	66.0 (56.0 to 74.0)
89.0 (80.0 to 95.0)	85.0 (77.0 to 91.0)								
93.0 (85.0 to 97.0)	61.0 (52.0 to 70.0)								

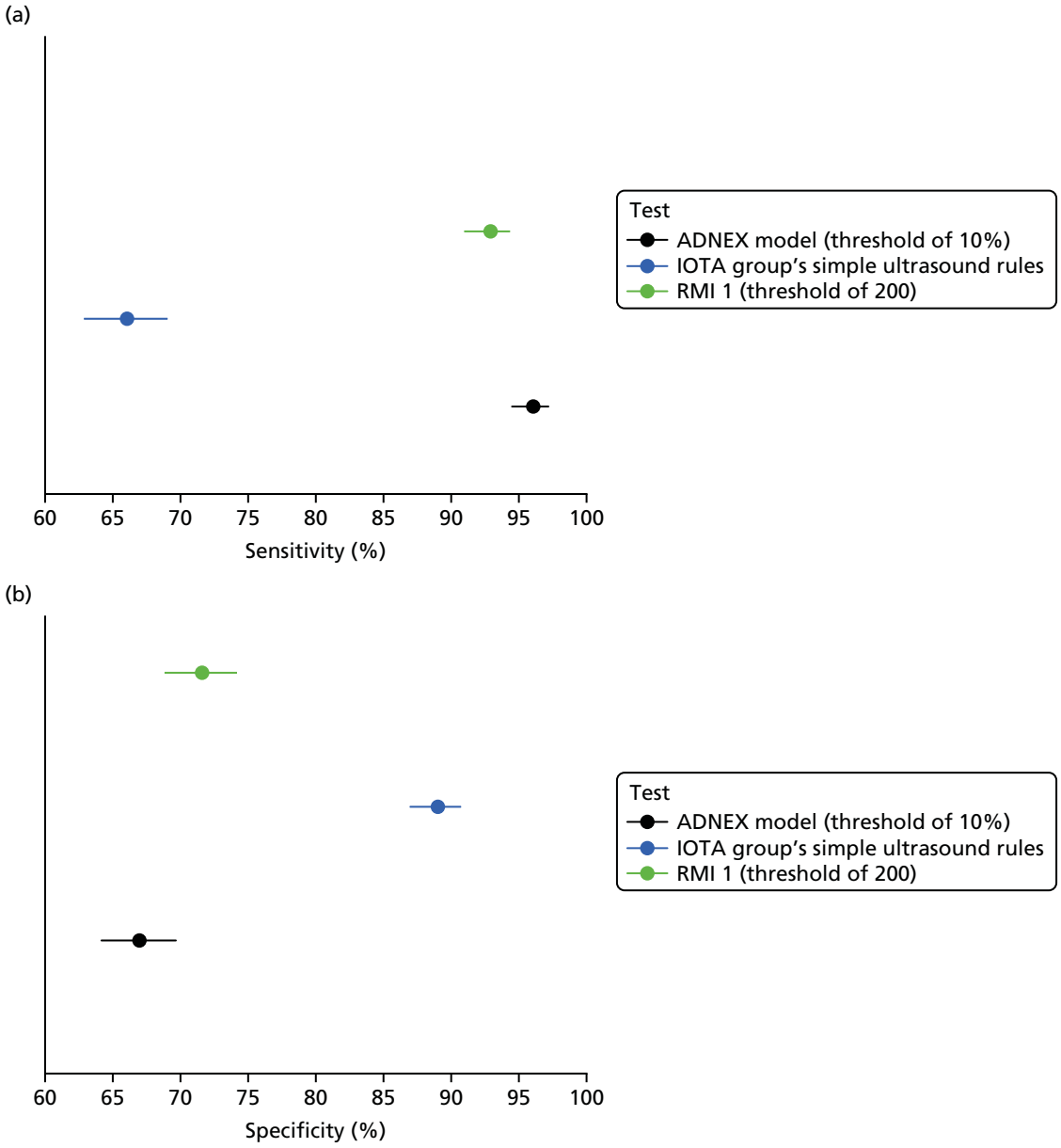

FIGURE 4 Comparison of the accuracy of the ADNEX model, the IOTA group's simple ultrasound rules and the RMI 1. (a) Sensitivity estimate and; (b) specificity estimate.

to 93.5%) and 80.3% (95% CI 76.9% to 83.4%), respectively. In premenopausal women, the summary sensitivity estimate for the RMI 1 (52.2%, 95% CI 47.4% to 56.9%) was significantly lower than that for the IOTA group's simple ultrasound rules, and the summary specificity estimate (94.2%, 95% CI 92.7% to 95.5%) was significantly higher. In postmenopausal women, the summary sensitivity estimate for the RMI 1 (78.8%, 95% CI 75.7% to 81.7%) was also significantly lower than that for the IOTA group's simple ultrasound rules; however, there were no significant differences between the specificity estimates (see *Table 15*).

One further study,[60] which was reported as only a conference abstract, did not report sufficient information to determine how participants with inconclusive IOTA group's simple ultrasound rules assessments were handled or how the target condition was defined (whether or not non-ovarian and borderline tumours were included in the definition of disease positive). This study reported sensitivity estimates of 94% for the IOTA group's simple ultrasound rules and 72% for the RMI 1, used at a decision threshold of 250. The corresponding specificity estimate was 80% for both tests.

Diagnostic performance of Overa (multivariate index assay, second generation)

Details of Overa (multivariate index assay, second-generation) studies

Three diagnostic cohort studies reported in four publications[68–70,104] provided data on the diagnostic performance of the Overa (MIA2G) score, for the identification of women with an adnexal mass who are at a high risk of developing ovarian cancer. All the studies were conducted in the USA. Only one study[70] was reported as a full paper; the remaining two studies were reported in the form of meeting slides[104] and a conference abstract.[68]

One study[70] used an Overa (MIA2G) score based on Roche Diagnostics' assays and a Roche Diagnostics analyser; the other two studies[68,104] did not report assay details.

The target condition for this assessment is ovarian cancer (i.e. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma and borderline ovarian cancer). All studies in this section included women with one or more adnexal mass and used a definition of malignancy that included borderline cancers. Histopathology indicated that all of the studies also included some women with non-ovarian malignancies and non-ovarian metastases. Full details of the final histopathological diagnoses of study participants who had a malignant mass are reported in *Appendix 4*, *Table 36*.

Accuracy of Overa (MIA2G) for determining a high risk of developing ovarian cancer

No studies were identified that directly compared Overa (MIA2G) to the RMI 1 at either decision threshold (200 or 250).

One study[104] reported comparative accuracy data for Overa (MIA2G) versus the ROMA score, using Roche Diagnostics' Elecsys tumour marker assays (*Table 16*). This study included all participants in the analysis, regardless of their final histopathological diagnosis (target condition: all malignancies including borderline). At a 29.9%) was 79.2% (95% CI 73.7%/29.9%) was 79.2% (95% CI 73.7% to 83.8%) and the specificity estimate was 78.9% (95% CI 75.8% to 81.7). These data indicate that the sensitivity of the Overa (MIA2G) score was significantly higher than that of the ROMA score, whereas the specificity of the Overa (MIA2G) score was significantly lower than that of the ROMA score (*Figure 5*).

The two remaining studies[68,70] reported data on the accuracy of Overa (MIA2G) without comparison with the RMI 1 or any other risk score and analysed data for any malignant tumour plus borderline (*Table 17*). At a threshold of 5 units, the pooled sensitivity estimate was 90.2% (95% CI 84.6% to 94.3%) and the pooled specificity estimate was 65.8% (95% CI 61.9% to 69.5%); these estimates were similar to those reported by the comparative accuracy study. One study stratified data by menopausal status and found no significant variation in test performance (see *Table 17*).[70]

TABLE 15 Comparative accuracy of the IOTA group's simple ultrasound rules and the RMI 1

Study (year of publication)	Subgroup	IOTA group's simple ultrasound rules threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)
Target condition: all malignant tumours including borderline								
Adballa *et al.* (2013)[48]	All women	Inconclusive = malignant	16	1	7	63	87	94.1 (71.3 to 99.9)
IOTA5 2017[a,b]			Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Meys *et al.* (2016)[44]			107	8	67	144	326	93.0 (86.0 to 97.0)
[c]Sayasneh *et al.* (2013)[62]			67	7	24	157	255	91.0 (82.0 to 95.0)
[b,c]Testa *et al.* (2014)[50]			934	46	369	1054	2403	95.3 (93.1 to 96.19)
Summary estimates								**93.9 (92.8 to 94.9)**
Meys *et al.* (2016)[44]	All women	Inconclusive = SA	102	13	21	190	326	89.0 (81.0 to 94.0)
[c]Sayasneh *et al.* (2013)[62]			64	10	11	170	255	86.0 (77.0 to 92.0)
[b,c]Testa *et al.* (2014)[50]			900	80	157	1266	2403	91.8 (89.1 to 93.9)
Summary estimates								**91.2 (89.4 to 92.8)**
Meys *et al.* (2016)[44]	Premenopausal women	Inconclusive = malignant	29	2	23	74	128	94.0 (77.0 to 99.0)
[c]Sayasneh *et al.* (2013)[62]			24	4	16	121	165	86.0 (69.0 to 94.0)
[b,c]Testa *et al.* (2014)[50]			359	19	225	751	1354	95.0 (91.0 to 97.0)
Summary estimates								**94.3 (91.7 to 96.3)**
Meys *et al.* (2016)[44]		Inconclusive = SA	27	4	4	93	128	87.0 (69.0 to 96.0)
[c]Sayasneh *et al.* (2013)[62]			23	5	5	132	165	82.0 (64.0 to 92.0)
[b,c]Testa *et al.* (2014)[50]			348	24	88	888	1354	92.0 (86.0 to 95.0)
Summary estimates								**92.3 (89.4 to 94.7)**
Meys *et al.* (2016)[44]	Postmenopausal women	Inconclusive = malignant	78	6	44	70	198	93.0 (85.0 to 97.0)
[c]Sayasneh *et al.* (2013)[62]			43	3	7	37	90	93.0 (82.0 to 98.0)
[b,c]Testa *et al.* (2014)[50]			578	24	152	295	1049	96.0 (93.0 to 97.0)
Summary estimates								**95.5 (93.7 to 96.9)**
Meys *et al.* (2016)[44]	Postmenopausal women	Inconclusive = SA	75	9	39	75	198	89.0 (80.0 to 95.0)
[c]Sayasneh *et al.* (2013)[62]			41	5	4	40	90	89.0 (77.0 to 95.0)
[b,c]Testa *et al.* (2014)[50]			560	42	76	371	1049	93.0 (90.0 to 95.0)

Specificity, % (95% CI)	RMI threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
90.0 (80.5 to 95.9)	200	15	2	8	62	87	88.2 (63.6 to 98.5)	88.6 (78.7 to 94.9)
Confidential information has been removed	200	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
68.0 (61.0 to 70.0)	200	82	33	44	167	326	71.0 (62.0 to 79.0)	79.0 (72.0 to 84.0)
87.0 (82.0 to 91.0)	200	53	21	11	170	255	72.0 (60.0 to 81.0)	94.0 (90.0 to 97.0)
74.1 (67.7 to 79.7)	200	657	323	134	1289	2403	67.1 (61.4 to 72.4)	90.6 (87.3 to 93.1)
74.2 (72.6 to 75.8)	**Summary estimates**						**66.9 (64.8 to 68.9)**	**90.1 (88.9 to 91.2)**
90.0 (85.0 to 94.0)	200	82	33	44	167	326	71.0 (62.0 to 79.0)	79.0 (72.0 to 84.0)
94.0 (90.0 to 97.0)	200	53	21	11	170	255	72.0 (60.0 to 81.0)	94.0 (90.0 to 97.0)
89.0 (85.2 to 92)	200	657	323	134	1289	2403	67.1 (61.4 to 72.4)	90.6 (87.3 to 93.1)
89.6 (88.1 to 91)	**Summary estimates**						**67.8 (65.0 to 70.4)**	**98.5 (98.3 to 98.7)**
76.0 (66.0 to 84.0)	200	13	18	6	91	128	42.0 (25.0 to 61.0)	94.0 (86.0 to 97.0)
88.0 (83.0 to 93.0)	200	15	13	5	132	165	54.0 (36.0 to 70.0)	96.0 (92.0 to 98.0)
77.0 (70.0 to 83.0)	200	200	178	59	917	1354	53.0 (45.0 to 61.0)	94.0 (92.0 to 96.0)
78.2 (75.7 to 80.5)	**Summary estimates**						**52.2 (47.4 to 56.9)**	**94.2 (92.7 to 95.5)**
96.0 (89.0 to 99.0)	200	13	18	6	91	128	42.0 (25.0 to 61.0)	94.0 (86.0 to 97.0)
96.0 (91.0 to 98.0)	200	15	13	5	132	165	54.0 (36.0 to 70.0)	96.0 (92.0 to 98.0)
91.0 (87.0 to 94.0)	200	200	178	59	917	1354	53.0 (45.0 to 61.0)	94.0 (92.0 to 96.0)
92 (90.3 to 93.5)	**Summary estimates**						**52.2 (47.4 to 56.9)**	**94.2 (92.7 to 95.5)**
61.0 (52.0 to 70.0)	200	69	15	39	75	198	82.0 (72.0 to 89.0)	66.0 (56.0 to 74.0)
84.0 (71.0 to 92.0)	200	38	8	5	39	90	83.0 (69.0 to 91.0)	89.0 (76.0 to 95.0)
66.0 (59.0 to 73.0)	200	470	132	85	362	1049	78.0 (72.0 to 83.0)	81.0 (76.0 to 85.0)
72.3 (68.9 to 75.5)	**Summary estimates**						**78.8 (75.7 to 81.7)**	**78.7 (75.2 to 81.9)**
85.0 (77.0 to 91.0)	200	69	15	39	75	198	82.0 (72.0 to 89.0)	66.0 (56.0 to 74.0)
91.0 (79.0 to 96.0)	200	38	8	5	39	90	83.0 (69.0 to 91.0)	89.0 (76.0 to 95.0)
83.0 (78.0 to 87.0)	200	470	132	85	362	1049	78.0 (72.0 to 83.0)	81.0 (76.0 to 85.0)

TABLE 15 Comparative accuracy of the IOTA group's simple ultrasound rules and the RMI 1 (*continued*)

Study (year of publication)	Subgroup	IOTA group's simple ultrasound rules threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)
Summary estimates								**92.3 (90.2 to 94.2)**
[b]Di Legge *et al.* (2012)[57]	Tumour size of < 4 cm	Inconclusive = malignant	42	9	13	332	396	82.0 (69.0 to 92.0)
	Tumour size of ≥ 10 cm		281	23	66	222	592	92.0 (89.0 to 95.0)
	Tumour size of 4–9.9 cm		303	27	60	1067	1457	92.0 (88.0 to 95.0)
Target condition: ovarian borderline tumours – higher-stage malignancies excluded								
[b,c]Testa *et al.* (2014)[50]	All women	Inconclusive = malignant	133	20	367	1056	1576	87.5 (79.3 to 92.8)
		Inconclusive = SA	121	32	152	1271	1576	79.5 (70.8 to 86.1)

SA, subjective assessment; TN, true negative; TP, true positive.
a Frances Nixon, personal communication.
b Data from the IOTA study cohort.
c 2 × 2 data were calculated; studies that are not denoted had reported 2 × 2 data.

Specificity, % (95% CI)	RMI threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
80.3 (76.9 to 83.4)	**Summary estimates**						**78.8 (75.7 to 81.7)**	**78.7 (75.2 to 81.9)**
96.0 (94.0 to 98.0)	200	29	22	16	329	396	56.0 (43.0 to 70.0)	95.0 (93.0 to 98.0)
77.0 (72.0 to 82.0)	200	224	80	38	250	592	74.0 (69.0 to 79.0)	87.0 (83.0 to 91.0)
95.0 (93.0 to 96.0)	200	220	110	68	1059	1457	67.0 (62.0 to 72.0)	94.0 (92.0 to 95.0)
74.2 (66.5 to 80.7)	200	45	108	134	1289	1576	29.6 (21.2 to 39.7)	90.6 (87.1 to 93.2)
89.3 (84.7 to 92.7)	200	45	108	134	1289	1576	29.6 (21.2 to 39.7)	90.6 (87.1 to 93.2)

TABLE 16 Comparative accuracy of Overa (MIA2G) vs. the ROMA score

Study (year of publication)	Index test	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline									
Shulman *et al.* (2016)[104]	Overa (MIA2G)	5 units	223	22	258	490	993	91.0 (86.8 to 94.0)	65.5 (62.0 to 68.8)
	ROMA score using Roche Diagnostics' tumour marker assay	11.4%/29.9%	194	51	158	590	993	79.2 (73.7 to 83.8)	78.9 (75.8 to 81.7)

TN, true negative; TP, true positive.

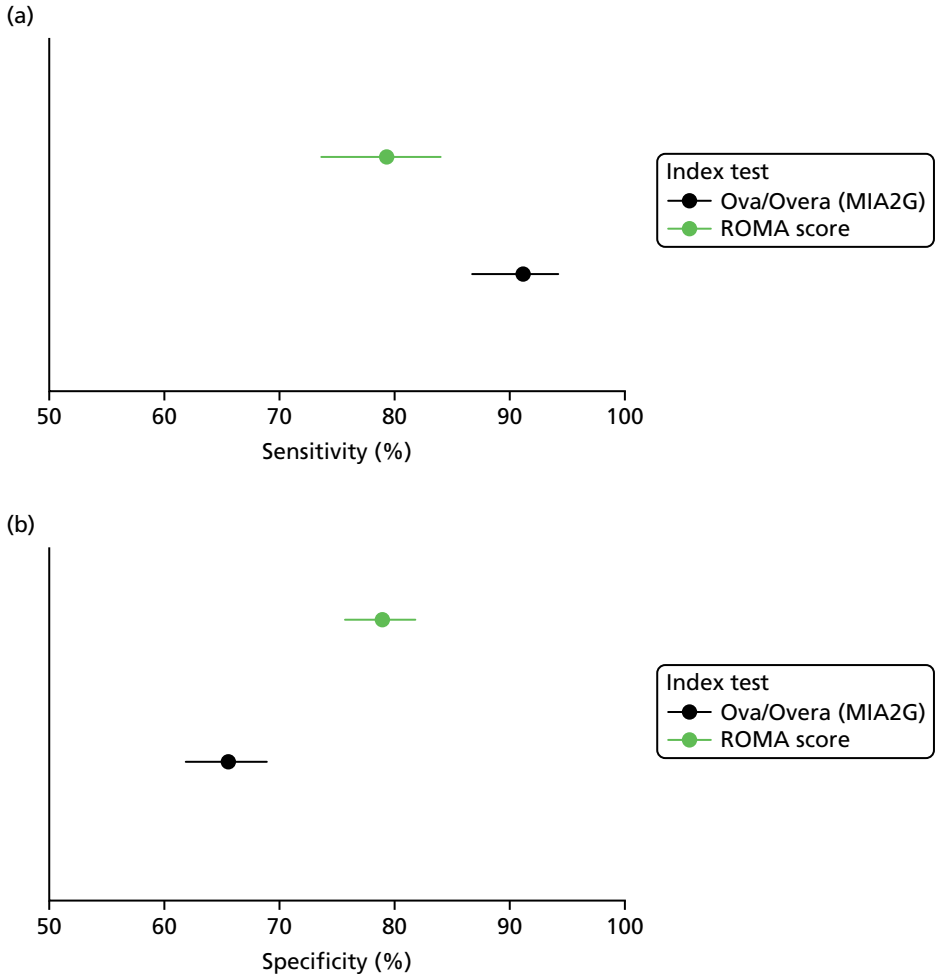

FIGURE 5 Comparison of the summary estimates for Overa (MIA2G) and ROMA score using Roche Diagnostics' tumour marker assay (all malignant tumours plus borderline). (a) Sensitivity estimate; and (b) specificity estimate.

TABLE 17 Accuracy of the Overa (MIA2G) score at a threshold of 5 units

Study (year of publication)	Subgroup	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline									
Coleman *et al.* (2016)[70]	All women	84	8	124	277	493	Reported	91.3 (83.8 to 95.5)	69.1 (64.4 to 73.4)
Zhang *et al.* (2015)[68]	All women	64	8	93	140	305	Reported	88.9 (79.3 to 95.1)	60.1 (53.5 to 66.4)
Summary estimates								**90.2 (84.6 to 94.3)**	**65.8 (61.9 to 69.5)**
Coleman *et al.* (2016)[70]	Premenopausal women	28	3	70	175	276	Reported	90.3 (75.1 to 96.7)	71.4 (65.5 to 76.7)
	Postmenopausal women	56	5	54	102	217	Reported	91.8 (82.2 to 96.4)	65.4 (57.6 to 72.4)

TN, true negative; TP, true positive.

Diagnostic performance of the Risk of Malignancy Index 1 using decision thresholds other than 250

Details of Risk of Malignancy Index 1 studies

Ten diagnostic cohort studies,[71–80] reported in 10 full-paper publications, provided data comparing the diagnostic performance of the RMI 1 at multiple decision thresholds, including a decision threshold of 250, for the identification of women with an adnexal mass who were at a high risk of developing ovarian cancer.

Two studies[78,79] specifically included women from the UK, two studies were European (from Italy and Norway)[76,80] and six studies were from five non-European countries (Turkey, Pakistan, China, India and Japan).[71–75,77]

Three studies[75,76,78] used a RMI 1 score based on an Abbott Diagnostics' CA125 assay, three studies[71,72,74] used a Roche Diagnostics assay, one study[77] used an IMMULITE® assay, one study[79] used CIS Bioindustries, one study[80] used a commercial kit by Centocor (Malvern, PA, USA) and one study[73] did not report the CA125 assay used.

This assessment is primarily concerned with providing a comparison between the RMI 1,[78] used with a decision threshold of 250 (current standard practice in the NHS[1]) and the specified alternative risk-scoring methods (see *Chapter 2, Intervention technologies*). The identified studies for the RMI 1 reported test performance data for multiple thresholds, and full data are reported in *Appendix 5*, *Table 40*. All of the identified studies that provided comparative accuracy data for alternative risk-scoring methods versus the RMI 1 used a decision threshold of 200. In order to assess the applicability of these data to the stated objective of this assessment, this section therefore focuses on the comparative accuracy of the RMI 1, using decision thresholds of 200 and 250.

The target condition for this assessment is ovarian cancer (i.e. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma and borderline ovarian cancer), defined as those conditions covered by the NICE clinical guideline CG122.[1] All studies in this section included women with one or more adnexal mass. Seven studies[72–74,76,78–80] used a definition of malignancy that included borderline tumours, two studies[71,75] excluded women found to have borderline tumours from the analyses and, in the remaining study,[77] it was unclear whether or not women with borderline tumours were included in the analysis (no histopathology was reported with which to confirm the tumour type). Six studies[73,74,76,78–80] included all study participants in the analyses and included some women with 'other malignancies', metastases from non-ovarian sites and 'non-ovarian cancers.' Full details of the final histopathological diagnoses of study participants who had a malignant mass are reported in *Appendix 4*, *Table 26*.

Accuracy of Risk of Malignancy Index 1 for determining a high risk of developing ovarian cancer using different decision thresholds

Six studies[73,74,76,78–80] included all study participants in the analyses, regardless of final histopathological diagnosis [target condition: all malignant tumours including borderline (*Table 18*)]. At the decision threshold of 200, the summary estimate of sensitivity derived from these studies was 70.8% (95% CI 65.6% to 75.6%) and the summary estimate of specificity was 91.2% (95% CI 88.9% to 93.1%). At the decision threshold of 250, the summary estimate of sensitivity was 69% (95% CI 63.7% to 73.9%) and the summary estimate of specificity was 91.6% (95% CI 89.3% to 93.5%). The sensitivity and specificity estimates did not differ significantly between the two decision thresholds [200 and 250 (*Figure 6*)]. Studies compared multiple thresholds (between 25 and 500); as would be expected, the sensitivity estimate for the RMI 1 increased and the specificity estimate decreased with a decreasing threshold (see *Appendix 5*, *Table 40*).

One study[72] reported a direct comparison of the RMI 1 at decision thresholds of 250 and 200 and excluded women with a final histopathological diagnosis other than primary ovarian cancer from the analysis [target condition: ovarian malignancies including borderline (see *Table 18*)]. At the decision threshold of 200, the sensitivity estimate was 80% (95% CI 65.2% to 89.5%) and the specificity estimate was 86.4% (95% CI 81.8% to 89.9%). At the decision threshold of 250, the sensitivity estimate was 72.5% (95% CI 57.2% to 83.9%) and the specificity estimate was 88.7% (95% CI 84.4% to 92.0%). Although the sensitivity estimate was higher for the 200 threshold and the specificity estimate was higher for the 250 threshold, these differences were not significantly different. In addition, the sensitivity and specificity estimates from this study did not differ significantly from the summary estimates described earlier.

Two further studies[71,75] excluded participants found to have borderline tumours from the analysis (target condition: all malignant tumours including borderline). At the decision threshold of 200, the summary estimate of sensitivity was 73.5% (95% CI 64.3% to 81.3%) and the summary estimate of specificity was 89.6% (95% CI 83.2% to 94.2%). At the decision threshold of 250, the summary estimate of sensitivity was 66.4% (95% CI 56.9% to 75.0%) and the summary estimate of specificity was 93.3% (95% CI 87.7% to 96.9%). The sensitivity and specificity estimates did not differ significantly between the two decision thresholds (200 and 250). In addition, these summary sensitivity and specificity estimates did not differ significantly from those derived from the six studies that included all participants in their analyses.

One study[77] included participants with malignant tumours, but it was unclear whether or not borderline tumours were included (see *Appendix 5*, *Table 43*).

TABLE 18 Comparative accuracy of the RMI 1 at decision thresholds of ≥ 200 and ≥ 250

Study (year of publication)	Threshold: 200 TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)	250 TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline														
Davies *et al.* (1993)[79]	33	4	11	76	124	89.2 (74.6 to 97.0)	87.4 (78.5 to 93.5)	34	3	21	66	124	91.9 (78.1 to 98.3)	75.9 (65.5 to 84.4)
Jacobs *et al.* (1990)[78]	35	6	3	95	139	85.4 (70.8 to 94.4)	96.9 (91.3 to 99.4)	32	9	1	97	139	78.0 (62.4 to 89.4)	99.0 (94.5 to 100)
Lou *et al.* (2010)[73]	34	27	5	157	223	55.7 (42.4 to 68.5)	96.9 (92.9 to 99.0)	35	26	3	159	223	57.4 (44.1 to 70.0)	98.1 (94.7 to 99.6)
Morgante *et al.* (1999)[80]	18	13	5	88	124	58.1 (39.1 to 75.5)	94.6 (87.9 to 98.2)	17	14	4	89	124	54.8 (36.0 to 72.7)	95.7 (89.4 to 98.8)
Tingulstad *et al.* (1996)[76]	40	16	5	112	173	71.4 (57.8 to 82.7)	95.7 (90.3 to 98.6)	38	18	5	112	173	67.9 (54.0 to 79.7)	95.7 (90.3 to 98.6)
Ulusoy *et al.* (2007)[74]	75	31	37	153	296	71.1 (62.1 to 80)	80.5 (74.2 to 85.9)	73	33	29	161	296	68.9 (59.1 to 77.5)	84.7 (78.8 to 89.5)
Summary estimates						**70.8 (65.6 to 75.6)**	**91.2 (88.9 to 93.1)**	**Summary estimates**					**69.0 (63.7 to 73.9)**	**91.6 (89.3 to 93.5)**
Target condition: ovarian malignancies including borderline														
Yamamoto *et al.* (2009)[72]	32	8	29	184	253	80.0 (65.2 to 89.5)[a]	86.4 (81.8 to 89.9)[a]	29	11	24	189	253	72.5 (57.2 to 83.9)[a]	88.7 (84.4 to 92)[a]
All malignant tumours excluding borderline														
Aktürk *et al.* (2011)[71]	15	5	9	71	100	75.0 (50.9 to 91.3)	88.8 (79.7 to 94.7)	13	7	4	76	100	65.0 (40.8 to 84.6)	95.0 (87.7 to 98.6)
Manjunath *et al.* (2001)[75]	68	25	5	50	148	73.1 (62.9 to 81.8)	90.9 (80.0 to 97.0)	62	31	5	50	148	66.7 (56.1 to 76.1)	90.9 (80.0 to 97.0)
Summary estimates						**73.5 (64.3 to 81.3)**	**89.6 (83.2 to 94.2)**	**Summary estimates**					**66.4 (56.9 to 75.0)**	**93.3 (87.7 to 96.9)**

TN, true negative; TP, true positive.
a Calculated values.

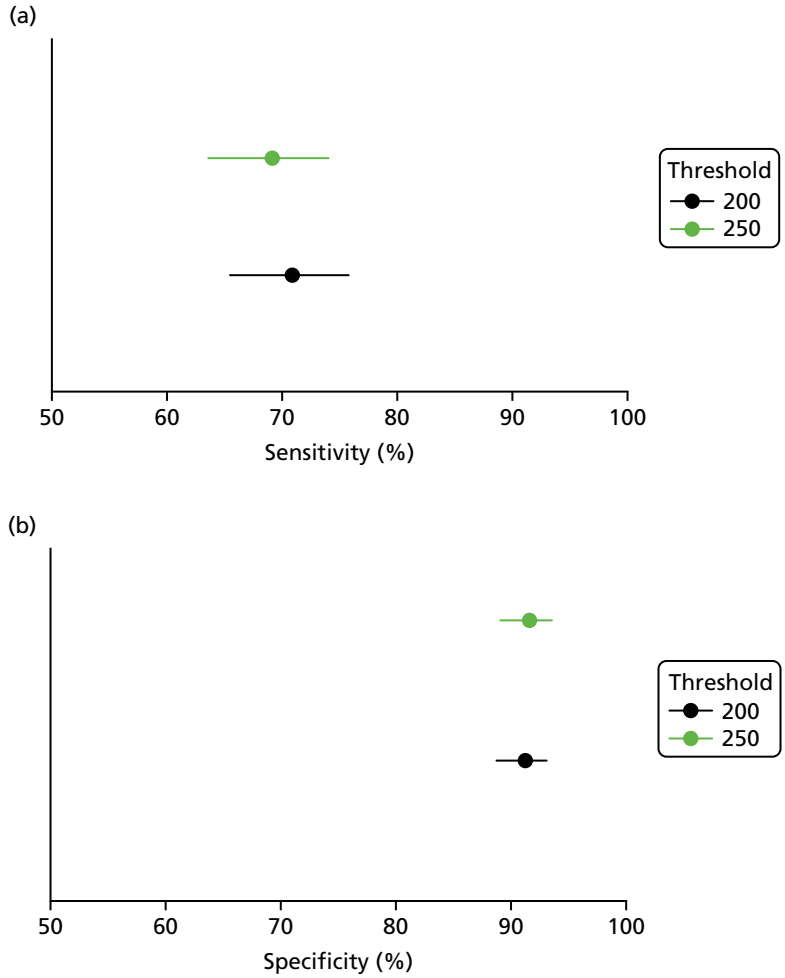

FIGURE 6 Comparison of the summary estimates for the RMI 1 at thresholds of 200 and 250 (all malignant tumours plus borderline). (a) Sensitivity estimate; and (b) specificity estimate.

Selection of diagnostic performance estimates for inclusion in cost-effectiveness modelling

Data for the target condition 'all malignant tumours including borderline' were prioritised. This is because the scope and protocol for this assessment specified that the definition of ovarian cancer should include borderline tumours. In addition, the population in which risk-scoring would be applied in practice is likely to include some women who will ultimately be found to have a non-ovarian primary and some who will have cancers that fall outside the scope of conditions covered in CG122[1] (e.g. germ cell tumours and sex cord–stromal tumours of the ovary); therefore, it was considered that studies that include all participants in their analysis, irrespective of final histological diagnosis, are more likely to produce estimates of risk score performance that are representative of what might be expected in clinical practice.

Comparative accuracy data were available for the risk scores ROMA, IOTA group's simple ultrasound rules and the ADNEX model versus the RMI 1 (i.e. studies evaluated the diagnostic performance both of the risk score and the RMI 1 in the same patient cohort). No studies were identified that provided a direct comparison of Overa (MIA2G) with the RMI 1. Summary estimates of the diagnostic performance of risk scores, calculated using all available data sets for a given target condition, did not differ significantly from those calculated from only those studies that reported a direct comparison with the RMI 1. Cost-effectiveness modelling therefore used the summary estimates of diagnostic performance of these larger data sets, making maximum use of the available data.

Estimates of the diagnostic performance of the comparator, the RMI 1 with a decision threshold of 250, were derived from a meta-analysis of all available RMI 1 data sets with the corresponding target condition (e.g. all malignant tumours including borderline or all ovarian tumours including borderline) and population (e.g. all participants, premenopausal women or postmenopausal women). When no data were available for the RMI 1 with a decision threshold of 250, data for a decision threshold of 200 were used; the analysis reported in *Diagnostic performance of the Risk of Malignancy Index 1 using decision thresholds other than 250* indicated no significant difference in the performance of the RMI 1 at these two thresholds.

Chapter 4 Assessment of cost-effectiveness

This chapter examines the cost-effectiveness of alternative risk scores, which include HE4 levels, CA125 levels or ultrasound, compared with the RMI 1 score as used in current practice for women with suspected ovarian cancer in secondary care, to guide decisions about referral to a SMDT. More specifically, the following research question is addressed:

- What is the cost-effectiveness of alternative risk scores (including alternative RMI 1 score thresholds), which include HE4 levels, CA125 levels or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current practice), when routinely used in secondary care to guide decisions about referral to a SMDT, for people with suspected ovarian cancer?

Review of economic analyses of ovarian cancer risk scores

Search strategy

Searches were undertaken to locate relevant economic evaluations of the target condition (ovarian cancer) and diagnosis with ultrasound, CA125 levels, HE4 levels or biomarkers.

Methodological study design filters were included in the search strategy when relevant. No restrictions on language or publication status were applied. The main EMBASE strategy was independently peer reviewed by a second information specialist using the Canadian Agency for Drugs and Technologies in Health Peer Review checklist.[31] Identified references were downloaded in EndNote X6 software for further assessment and handling. References in retrieved articles were checked for additional studies.

The following databases were searched for relevant studies:

- MEDLINE (via Ovid) – 1946 to week 2 November 2016
- MEDLINE In-Process Citations (via Ovid) – to 22 November 2016
- MEDLINE Daily Update (via Ovid) – to 22 November 2016
- MEDLINE Epub Ahead of Print (via Ovid) – to 23 November 2016
- EMBASE (via Ovid) – 1974 to 22 November 2016
- NHS Economic Evaluation Database (via Wiley Online Library) – to Issue 2 of 4, April 2015
- EconLit (via EBSCO*host*) – 1966 to 25 November 2016
- Cost-Effectiveness Analysis Registry (via the internet: http://www.cearegistry.org) – to 25 November 2016
- Research Papers in Economics (via the internet: http://repec.org/) – to 25 November 2016.

The full search strategies are presented in *Appendix 1*.

Inclusion criteria

Studies reporting outcomes of a full cost-effectiveness analysis, examining QALYs, with (at least) one of the comparators, were eligible for inclusion. Studies conducted in primary care settings and screening studies were included to ensure that no potentially relevant information on costs or health-related quality of life was missed.

Quality assessment

Included studies are appraised using a quality checklist based on Drummond *et al.*[120]

Results

The literature search identified 749 records from bibliographic database searches and supplementary searching (e.g. reference/citation checking and additional database searches, including the database search for the assessment of clinical effectiveness). After title and abstract screening, 10 records were considered

to be potentially relevant; after full-text screening, five studies[121–125] (five publications, including one abstract) were considered to be eligible for inclusion (*Figure 7*). These studies are described in more detail below and summarised in *Table 19*. The results of the quality assessment are shown in *Table 20*.

Havrilesky *et al.* (2015)

Havrilesky *et al.*[123] constructed a Markov model [in TreeAge Pro 2013 (TreeAge Software, Inc., Williamstown, MA, USA)] using alive and death health states. From a societal US perspective, the authors estimated the costs and outcomes of five strategies to help clinicians to decide which women with an adnexal mass requiring surgery would most benefit from subspecialist referral:

1. American Congress of Obstetricians and Gynecologists (ACOG)'s guidelines
2. multivariate index assay (MIA) algorithm
3. ROMA
4. CA125 level alone with lowered cut-off values to prioritise test sensitivity over specificity (15 U/ml for postmenopausal and 22 U/ml for premenopausal women)
5. referral of all women.

The analyses indicated that CA125 level is a cost-effective test for willingness-to-pay thresholds below US$9423 and US$10,644 per LYs gained for postmenopausal women and premenopausal women, respectively. The refer-all strategy was cost-effective above these thresholds. The other strategies are dominated. Therefore, it was concluded that referral of all women to a subspecialist is a cost-effective strategy for managing women with adnexal masses requiring surgery. However, if a test-based triage strategy is needed (e.g. because of capacity constraints), CA125 level with lowered cut-off values should be considered.

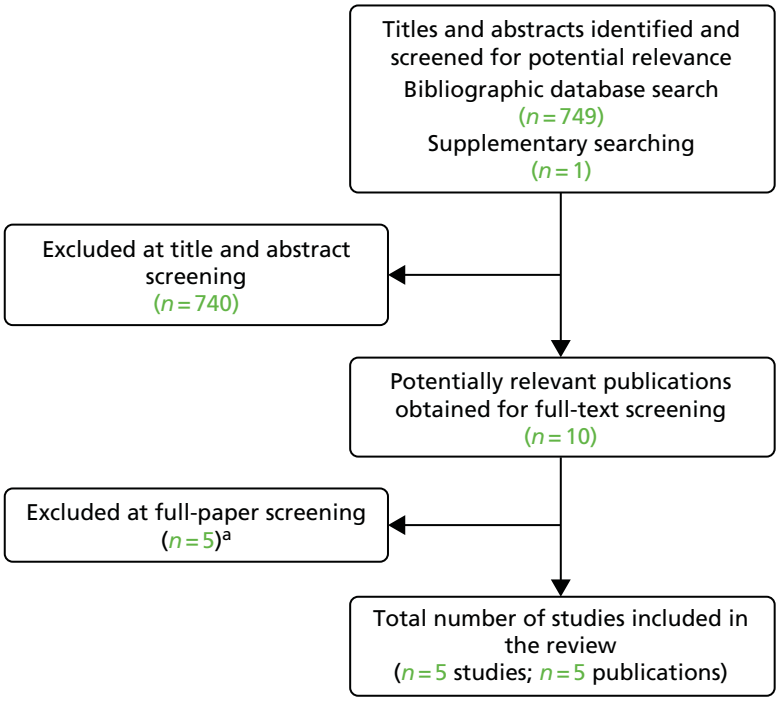

FIGURE 7 Flow chart (review of economic analyses). a, Reasons for exclusion: did not report outcomes of a full cost-effectiveness analysis ($n = 2$) and did not report quality-adjusted LYs or LYs as an outcome ($n = 3$).

TABLE 19 Summary of included economic evaluations (all abstracts)

Study characteristics	Study (year of publication)				
	Havrilesky *et al.* (2015)[123]	Drescher *et al.* (2012)[124]	Kearns *et al.* (2016)[125]	Forde *et al.* (2016)[122]	Ding *et al.* (2010)[121]
Population	Women with an adnexal mass	Women aged 45–85 years	Postmenopausal women aged 50–74 years in the UK	Women with adnexal masses	Postmenopausal females aged 65–69 years
Setting	At generalist obstetrician–gynaecologist (decision to refer to a subspecialist)	First-line screening	Secondary care	Secondary care	Screening
Time horizon	NR	NR	Lifetime	Lifetime	Lifetime
Objective	To compare the estimated costs and outcomes of five strategies to help clinicians decide which women with an adnexal mass requiring surgery would most benefit from subspecialist referral	To estimate the mortality reduction, years of life saved, and cost-effectiveness of epithelial ovarian cancer screening protocols in a hypothetical cohort of women aged 45–85 years	To evaluate the potential cost-effectiveness of screening for ovarian cancer in the UK and to estimate the value of further research into ovarian cancer screening	To evaluate the cost-effectiveness of the MIA for use in triaging women with an adnexal mass	To assess the cost-effectiveness of annual MMS vs. no screening for postmenopausal females aged 65–69 years
Source of effectiveness information	Literature	Literature	UKTOCS study and extrapolation of mortality data	Published data on survival, prognostic factors, effectiveness of surgical cytoreduction	NCT00058032 clinical trial
Comparators	1. ACOG guidelines 2. MIA algorithm 3. ROMA 4. CA125 levels alone with lowered cut-off values to prioritise test sensitivity over specificity 5. Referral of all women	1. No screening 2. CA125 levels and TVS 3. CA125 levels and hypothetical imaging 4. Hypothetical biomarker and TVS 5. Hypothetical biomarker and hypothetical imaging	1. USS (TVS by sonographer in first line and by a more experienced member of staff in second line) 2. MMS (CA125 levels interpreted using ROCA in first line and ultrasound by a more experienced member of staff in second line)	1. MIA: based on five biomarkers, transthyretin, apolipoprotein, alpha-1 microglobulin, beta-2 microglobulin, TRF and CA125 levels 2. Modified ACOG referral guidelines (including patient's history, physical, pelvic ultrasound and CA125 levels) 3. CA125 levels alone	1. Annual MMS (with CA125 marker, followed by TVS for those at an increased risk according to CA125 level) 2. No screening
Costs items	• Test costs • Surgery costs • Subspecialist costs • End-of-life costs	• Test costs • Surgery costs • Treatment costs • End-of-life costs	Multimodal and USS dropouts and complete screening, screening invitation, diagnosis and treatment of borderline or stages I–IV ovarian cancer, end-of-life costs	Chemotherapy with different cycle lengths and for CRC; diagnosis-related group costs and professional fees for surgery for malignancy, non-malignancy, staging surgery; CT scan; CA125 level; modified ACOG guidelines; MIA	Not stated
Main measure of benefit	LY	LY	QALY	QALY	QALYs

continued

TABLE 19 Summary of included economic evaluations (all abstracts) (*continued*)

Study characteristics	Study (year of publication)				
	Havrilesky *et al.* (2015)[123]	Drescher *et al.* (2012)[124]	Kearns *et al.* (2016)[125]	Forde *et al.* (2016)[122]	Ding *et al.* (2010)[121]
Main assumptions	• 75% of all ovarian cancers are postmenopausal • Survival advantage for women who undergo surgery by a subspecialist (recurrence rates after 80 months are independent of the specialty of the original surgeon) • The 'average' postsurgical treatment, including chemotherapy, is similar for women with ovarian cancer, no matter who performed the initial surgery. Because the costs and impact on quality of life of this postsurgical cancer treatment are not expected to be different on average, these were not included in the analysis • All FN tests (initial surgery performed by a generalist) result in postoperative subspecialist referral, with a CT scan performed, followed by restaging/debulking surgery (immediately or following initiation of chemotherapy) in 50% of patients • Women referred to a subspecialist after oophorectomy by a generalist for unsuspected ovarian cancer would also undergo a CT scan prior to a decision regarding a second surgical procedure; 50% of these women would undergo additional staging/debulking surgery • Various cost assumptions (see methods section of the paper for more details)	• Women alive 15 years after an epithelial ovarian cancer diagnosis are assumed to be cured • TVS is equally sensitive throughout the disease duration once CA125 values have elevated above the positivity threshold • Risk of developing epithelial ovarian cancer following bilateral salpingo-oophorectomy is assumed to be zero • A stage shift is assumed to occur whenever a tumour destined to be diagnosed clinically in late-stage (III or IV) ovarian cancer is detected in early-stage (I or II) ovarian cancer by screening	Log-normal for modelling survival in screening arms and Weibull in no-screening arm; disutility associated with diagnosis relates to treatment and lasts for only 1 year; no disutility associated with screening, use of ROCA does not increase costs	Major treatment-related costs occur during the first year of treatment; quality-of-life utility weights change with disease progression and differ by stage and type of cancer	Not stated

Study characteristics	Study (year of publication)				
	Havrilesky *et al.* (2015)[123]	Drescher *et al.* (2012)[124]	Kearns *et al.* (2016)[125]	Forde *et al.* (2016)[122]	Ding *et al.* (2010)[121]
Perspective	Societal perspective	NR	NHS and Personal Social Services	Public payer	US societal perspective
Discount rate	3%	3%	3.5% for costs and QALYs	3% for costs and QALYs	3% for costs and QALYs
Uncertainty around cost-effectiveness ratio expressed	Yes	Yes	Yes, EVPI and EVPPI was also performed	Yes, ICERs with one-way sensitivity analysis are given	No, but stated that cancer incidence rates and time required for screening exhibited substantial impact in sensitivity analyses
Sensitivity analysis	PSA	Threshold and scenario analyses	Yes, one-way sensitivity analysis and PSA	Yes, one-way sensitivity analysis	Yes
Monetary outcomes	2013 US$	2010 US$	£	2014 US$	2009 US$
Outcomes per comparator	Postmenopausal (costs; LYs): • CA125 – US$17,428; 16.93 • ACOG – US$17,469; 16.92 • ROMA – US$17,485; 16.91 • Refer all – US$17,510; 16.94 • MIA – US$18,004; 16.92 Premenopausal (costs; LYs): • CA125 – US$9876; 28.58 • ACOG – US$9892; 28.55 • ROMA – US$9897; 28.57 • Refer all – US$9999; 28.60 • MIA – US$10,354; 28.58	• No screening: US$865 • CA125 level and TVS: US$1741 • Absolute LYs are not provided (see results section of the paper for the results of hypothetical strategies)	MMS vs. USS vs. no screening: • QALYs – 14.357 vs. 14.297 vs. 14.29 • Costs – £598 vs. £824 vs. £179	• MIA vs. modified ACOG (only direct costs; direct and indirect costs): US$35,094; dominating • MIA vs. CA125 level (only direct costs; direct and indirect costs): US$12,189; dominating	NR
Summary of incremental analysis	CA125 level is cost-effective for willingness-to-pay thresholds below US$9423 and US$10,644 per LY gained for postmenopausal women and premenopausal women, respectively. Refer all is cost-effective above these thresholds. The other strategies are dominated	CA125 level and TVS led to 1.68 more LYs than no screening, resulting in an ICER of US$88,993 per LY gained (see results section of the paper for the results of hypothetical strategies)	MMS and USS are likely to be associated with benefits for patients, but also with additional costs. The ICER of MMS vs. no screening was £8864 and USS was dominated by MMS	MIA and referral to a gynaecologic oncologist (instead of surgery by a gynaecologist) for all patients are the most cost-effective triage strategies for women with adnexal masses	MMS resulted in additional costs and QALYs of US$820 and 0.0037, respectively vs. no screening. This resulted in an ICER of US$226,622 per QALY gained

EVPI, expected value of perfect information; EVPPI, expected value of partial perfect information; ICER, incremental cost-effectiveness ratio; MMS, multimodal screening; NA, not applicable; NR, not reported; PSA, probabilistic sensitivity analysis; ROCA, Risk of Ovarian Cancer Algorithm; UKTOCS, UK Collaborative Trial of Ovarian Cancer Screening; USS, ultrasound screening.

TABLE 20 Study quality checklist for included studies

Study details	Study (year of publication)				
	Havrilesky *et al.* (2015)[123]	Drescher *et al.* (2012)[124]	Kearns *et al.* (2016)[125]	Forde *et al.* (2016)[122]	Ding *et al.* (2010)[121]
Study design					
The research question is stated	✓	✓	✓	✓	✓
The economic importance of the research question is stated	✓	✓	✓	✓	✗
The viewpoint(s) of the analysis are clearly stated and justified	✓	✗	✓	✓	✓
The rationale for choosing alternative programmes or interventions compared is stated	✓	✓	✓	✓	✗
The alternatives being compared are clearly described	✓	✓	✓	✓	✗
The form of economic evaluation used is stated	✓	✓	✓	✓	✓
The choice of form of economic evaluation is justified in relation to the questions addressed	✓	✗	✓	✓	✓
Data collection					
The source(s) of effectiveness estimates used are stated	✓	✓	✓	✓	NA
Details of the design and results of the effectiveness study are given (if based on a single study)	✗	✗	✓	✗	✗
Details of the methods of synthesis or meta-analysis of estimates are given (if based on a synthesis of a number of effectiveness studies)	NA	NA	NA	NA	✗
The primary outcome measure(s) for the economic evaluation are clearly stated	✓	✓	✓	✓	✓
Methods to value benefits are stated	NA	NA	✓	✗	✗
Details of the subjects from whom valuations were obtained were given	NA	NA	✗	✗	✗
Productivity changes (if included) are reported separately	NA	NA	NA	✓	✗
The relevance of productivity changes to the study question is discussed	NA	NA	NA	✓	✗
Quantities of resource use are reported separately from their unit costs	✗	✗	✓	✗	✗
Methods for the estimation of quantities and unit costs are described	✗	✗	✓	✗	✗

TABLE 20 Study quality checklist for included studies (*continued*)

Study details	Study (year of publication)				
	Havrilesky *et al.* (2015)[123]	Drescher *et al.* (2012)[124]	Kearns *et al.* (2016)[125]	Forde *et al.* (2016)[122]	Ding *et al.* (2010)[121]
Currency and price data are recorded	✓	✓	✓	✓	✓
Details of currency of price adjustments for inflation or currency conversion are given	✓	✓	✓	✓	✗
Details of any model used are given	✓	✗	✓	✓	✗
The choice of model used and the key parameters on which it is based are justified	✗	✗	✓	✓	✗
Analysis and interpretation of results					
Time horizon of costs and benefits is stated	✗	✗	✓	✓	✓
The discount rate(s) is stated	✓	✓	✓	✓	✓
The choice of discount rate(s) is justified	✗	✗	✓	✓	✗
An explanation is given if costs and benefits are not discounted	NA	NA	NA	NA	NA
Details of statistical tests and CIs are given for stochastic data	✓	✗	✓	✗	✗
The approach to sensitivity analysis is given	✓	✓	✓	✓	✓
The choice of variables for sensitivity analysis is justified	✗	✓	✓	✓	✗
The ranges over which the variables are varied are justified	✗	✓	✓	✓	✗
Relevant alternatives are compared	✓	✓	✓	✓	✓
Incremental analysis is reported	✓	✓	✓	✓	✓
Major outcomes are presented in a disaggregated, as well as aggregated form	NA	NA	✓	✓	✗
The answer to the study question is given	✓	✓	✓	✓	✓
Conclusions follow from the data reported	✓	✓	✓	✓	✓
Conclusions are accompanied by the appropriate caveats	✓	✗	✓	✓	✓

✗, no; ✓, yes; NA, not applicable.

Drescher *et al.* (2012)

Drescher *et al.*[124] used an unspecified model type and structure to estimate the cost-effectiveness of first-line testing of women with an adnexal mass, TVS and CA125 level, in women aged 45–85 years (US setting, perspective not stated). The following multimodal testing strategies were considered:

- no primary care testing
- CA125 level followed by TVS
- CA125 level followed by hypothetical imaging with 50% improvement in sensitivity compared with TVS
- hypothetical biomarker with twofold greater sensitivity followed by TVS
- hypothetical biomarker and hypothetical imaging (as above).

The analysis indicated that CA125 level and TVS led to 1.68 more LYs than no primary care testing, resulting in an incremental cost-effectiveness ratio (ICER) of US$88,993 per LY gained. Moreover, it was concluded that testing outcomes are relatively insensitive to second-line test performance and costs. Identification of a first-line test that does substantially better than CA125 level and has similar costs is required for primary care testing to reduce ovarian mortality by at least 25% and be reasonably cost-effective.

Kearns *et al.* (2016)

Using the NHS Personal Social Services perspective, Kearns *et al.*[125] developed a Markov model to estimate the cost-effectiveness of different screening strategies in postmenopausal women and to estimate the value of further research. The following screening strategies were considered:

- multimodal screening (MMS): first-line screening with CA125 level interpreted with risk of ovarian cancer algorithm, followed by TVS performed by senior staff
- ultrasound screening (USS): first-line screening with TVS performed by less-experienced staff, followed by TVS performed by more-experienced staff
- no screening.

Results indicated that USS was dominated by MMS, being both more costly and less effective. Compared with no screening, MMS cost £419 more and generated 0.047 additional QALYs, resulting in an ICER of £8864 per QALY gained, but alternative mortality extrapolation methods increased the ICER. The conclusion was that MMS for ovarian cancer is both more effective and more expensive than no screening, but that substantial uncertainty remains regarding the extrapolated long-term effectiveness.

Forde *et al.* (2016)

From the perspective of the public payer, Forde *et al.*[122] developed a Markov model to evaluate the cost-effectiveness of different strategies for use in triaging women with an adnexal mass. The following triage strategies were considered:

- the MIA (Ova1; Vermillion, Inc., Austin, TX, USA) based on five biomarkers, including CA125 level
- the modified ACOG (mACOG) referral guidelines
- CA125 level testing alone.

The MIA resulted in fewer reoperations and pretreatment CT scans, and was cost-effective compared with the ACOG referral guidelines, with an ICER of US$35,094 per QALY gained. The MIA dominated CA125 level alone, by being cost-saving and QALY-increasing. The MIA is expected to increase the percentage of women with ovarian cancer referred to gynaecological oncologists, thereby improving clinical outcomes.

Ding and Hay (2010)

Ding and Hay[121] assess the cost-effectiveness of annual MMS (with CA125 marker, followed by TVS for those women at an increased risk of developing ovarian cancer according to their CA125 level) versus no screening for postmenopausal females aged 65–69 years from a US societal perspective. It should be noted

that the available information for this assessment is restricted to one abstract (despite efforts to contact the authors). The incremental analysis indicated that, over a lifetime, MMS was both more costly (incremental costs of US$820) and more effective (incremental QALY of 0.004), resulting in an ICER of US$221,622 per QALY gained compared with no screening.

Quality assessment and summary of studies in the cost-effectiveness review

In total, three[121,122,125] out of the five included studies reported QALYs as the outcome. Of these studies, two[121,125] considered population screening, whereas the remaining study[122] considered the assessment of women referred to secondary care from the US perspective. The last study was of reasonable quality (see *Table 20*). The UK screening study[125] indicated that multimodal triage consisting of CA125 level followed by TVS could be cost-effective compared with ultrasound only and no triage. The two studies considering MIA, both from the US perspective, provided conflicting results; one[122] indicated that MIA might be cost-effective, whereas the other indicated that it was dominated by other strategies (when considering LYs).[123] This latter study[123] was the only one to consider the ROMA score, and also indicated that this strategy would be dominated by other strategies (when considering LYs). Moreover, this study indicated that a refer-all strategy is cost-effective for thresholds above US$10,644 per LY gained.[123] In conclusion, there is limited and conflicting evidence regarding the cost-effectiveness of alternative risk scores, which include HE4 level, CA125 level or morphological features seen on ultrasound, compared with the RMI 1 score with a referral threshold of ≥ 250 (current UK practice[1]) for women with suspected ovarian cancer in secondary care. The population screening studies were included as a potential source of information for our cost-effectiveness analysis in case all data gaps could not be filled with the more relevant second-line studies. However, because all data gaps could be addressed with the more relevant studies, the population screening studies were not used.

Model structure and methodology

Interventions and comparators

The health economic analysis estimates the cost-effectiveness of different risk scores to estimate an individual's risk of malignancy. This risk score can inform decisions about SMDT referral. The following risk scores are considered in the model:

1. RMI 1 score (at a threshold of 200)
2. RMI 1 score (at a threshold of 250)
3. ROMA score using Abbott Diagnostics' ARCHITECT
4. ROMA score using Roche Diagnostics' Elecsys
5. Overa (MIA2G) from Vermillion (at a threshold of 5 units)
6. IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)
7. IOTA group's ADNEX model (threshold most commonly used in studies: 10%).

An optimised risk assessment that reduces the number of women with ovarian cancer who are not referred for further specialist care (i.e. those with a FN risk assessment) has the potential to improve prognosis, be cost-saving in terms of avoiding unnecessary further investigations and optimising staging and surgical treatment, and to reduce associated anxiety. It is likely that women who are believed to have a benign explanation for any pelvic mass will be operated on in secondary care. If they actually have ovarian cancer, then the prognosis might be worse than if they had been operated on by a specialist gynaecological oncology surgeon.

The current standard assessment (RMI 1 score at a decision thresholds of ≥ 250) has been reported as having poor sensitivity (69%) for the prediction of malignancy (see *Table 18*). If referral decisions are based on the RMI 1 score at this threshold, there remains the potential for significant numbers of women with ovarian cancer to remain unreferred and experience consequential delays in diagnosis and detrimental effects on prognosis. This risk score was used as reference strategy. Alternative risk scores evaluated in the

model are the ROMA score [ARCHITECT tumour marker assays (CA125 and HE4 levels) from Abbott Diagnostics; Elecsys tumour marker assays (CA125 and HE4) from Roche Diagnostics], the simple ultrasound rules classification system from the IOTA group, the ADNEX model from the IOTA group and the Overa [(MIA2G) from Vermillion], and alternative decision thresholds for the RMI 1. The model does not include LUMIPULSE G HE4 (Fujirebio Diagnostics), as no studies of this technology were identified, or LUMIPULSE HE4 EIA (Fujirebio Diagnostics), as this test was outside the scope of our assessment (see *Accuracy of the Risk of Ovarian Malignancy Algorithm score using Fujirebio Diagnostics' tumour marker assays*). Alternative threshold values for the IOTA group's ADNEX model were not considered, as the 10% threshold is the most commonly studied and has been used in model validation studies.[42,46]

For the IOTA group's simple ultrasound rules risk score, it was assumed that inconclusive assessments would be classified as malignant, as this was assumed to be most representative of what would be available in secondary care (no additional input from a specialist ultrasonographer needed). Concerning the alternative decision thresholds for the RMI 1, a threshold of ≥ 200 (used in the original publication[78]) was used in the base-case analysis and other RMI 1 thresholds were considered in scenario analyses.

Model structure

This assessment uses the economic model from CG122[1] as a starting point. CG122[1] reviewed clinical and economic questions that involve the detection in primary care, diagnosis in secondary care and initial management of early- and advanced-stage ovarian cancer (AOC). The CG122[1] model consisted of a decision tree outlining the assessment strategies, and a Markov process to model the progression and survival of women with ovarian cancer based on the results of the diagnostic tests and the subsequent treatment of women presenting with symptom(s) of ovarian cancer. The CG122[1] model was constructed using TreeAge Pro [(2009) TreeAge Software Inc., Williamstown, MA, USA] software. The assessment group used the description of this model as a starting point to develop a de novo model [in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA)], adapted to better fit the scope of the current assessment. Consistent with the CG122[1] model, the population age in the base case was assumed to be 40 years. In the subgroup analysis, different ages were used to reflect the premenopausal (mean age of 38 years) and postmenopausal (mean age of 68 years) groups of women. The mean age for both groups was estimated based on information on the distribution of ovarian cancer patients pre- and post-age 50 years from Cancer Research UK, assuming that menopause occurs approximately at the age of 50 years.[126]

In the de novo health economic model, the mean expected costs and QALYs were calculated for each alternative risk score. These long-term consequences were estimated based on the accuracy of the different risk scores to predict ovarian cancer, followed by referral to and treatment by a SMDT, or no tertiary care referral. It was also taken into account that a small proportion of the women with pelvic masses, who tested positive based on the risk score, were ultimately diagnosed with CRC (consistent with CG122). These women were therefore included in the model and the prognosis for women with CRC was included in the Markov model.

A decision tree and a Markov model were developed. The decision tree was used to model the short-term (up to 30 days after surgery) outcomes. It was assumed that women who receive a high-risk test result (either true or false) are referred to a SMDT, and women who receive a low-risk test result (either true or false) are not referred to a SMDT. After the risk assessment and referral decision, women in the decision tree are allocated to 'early ovarian cancer', 'AOC', 'benign mass', 'colorectal cancer' and 'death'. Death was included as an outcome in the decision tree to account for 30 days' post-surgery mortality. Women referred to a SMDT receive surgery by gynaecological oncology specialists that has shown to achieve better patient outcomes than those for patients not referred to a SMDT and, for a proportion of patients, this surgery is extensive. Women not referred to a SMDT receive surgery by secondary care gynaecologists. In the case of a FN diagnosis (i.e. when a woman has a malignancy and is incorrectly classified as having a low risk score), there is an increased risk of progressing to AOC and/or death. This increased risk is likely to be the result of a combination of factors, such as a delay in appropriate treatment, because the woman would be operated on and then referred to a SMDT for another surgery (based on clinical experts' feedback). Women with a benign mass who are incorrectly classified as being at a high risk of developing ovarian cancer

and referred to a SMDT receive surgery and have their benign mass removed. This incorrect referral has only cost implications, as women are identified as having a benign mass at surgery. Alternatives in the women's care pathway are explored in scenario analyses. The decision tree is shown in *Figure 8*.

The long-term consequences in terms of costs and QALYs were estimated using a Markov cohort model (*Figure 9*) with a lifetime time horizon. The cycle time was 1 year. The following health states were included:

- benign mass
- early ovarian cancer, not referred to a tertiary care SMDT
- early ovarian cancer, referred to a tertiary care SMDT
- AOC, not referred to a tertiary care SMDT
- AOC, referred to a tertiary care SMDT
- colorectal cancer Dukes' stage A
- colorectal cancer Dukes' stage B
- colorectal cancer Dukes' stage C
- colorectal cancer Dukes' stage D
- death.

A distinction between the decision to refer to a SMDT or not was made only for 'early-stage ovarian cancer' and 'AOC'. This was done as it was assumed that a referral to the SMDT would have an impact on the long-term outcomes in terms of LYs and QALYs for women with ovarian cancer only.

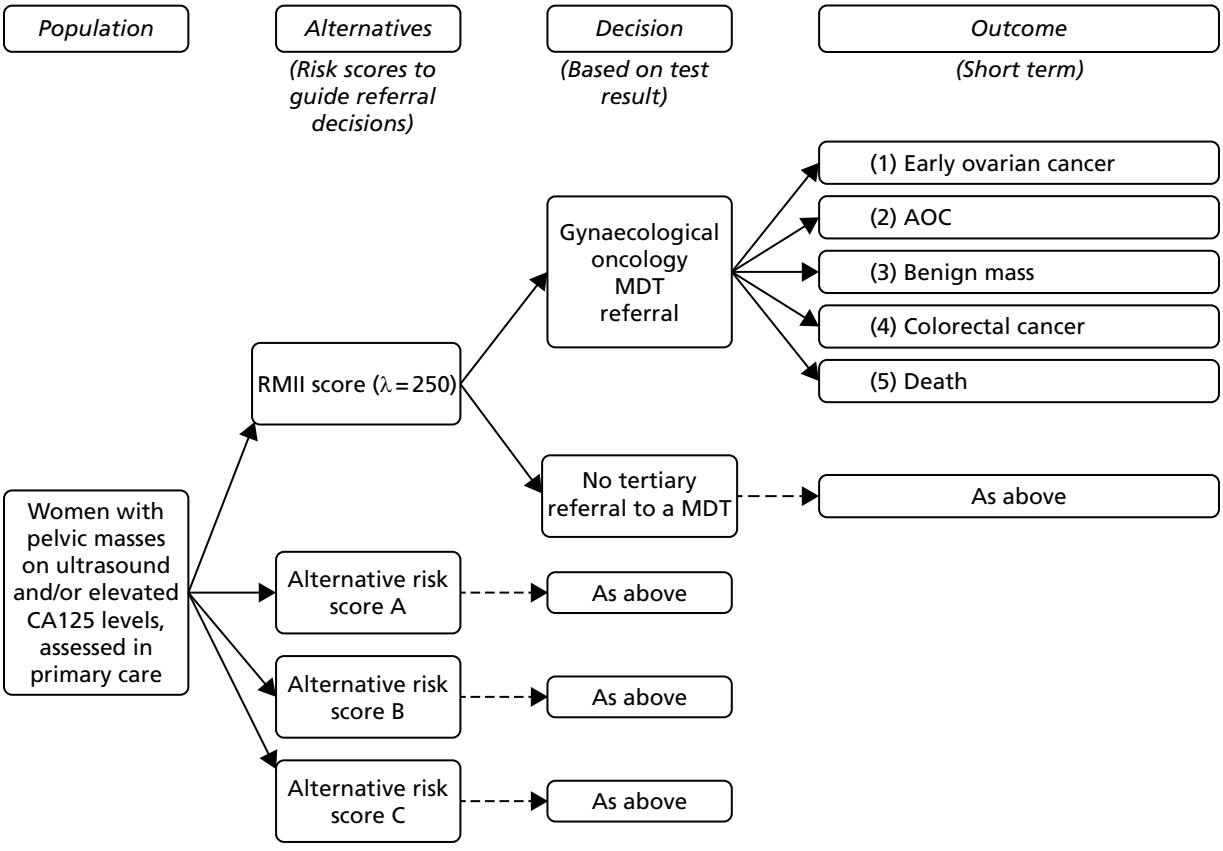

FIGURE 8 Decision tree structure.

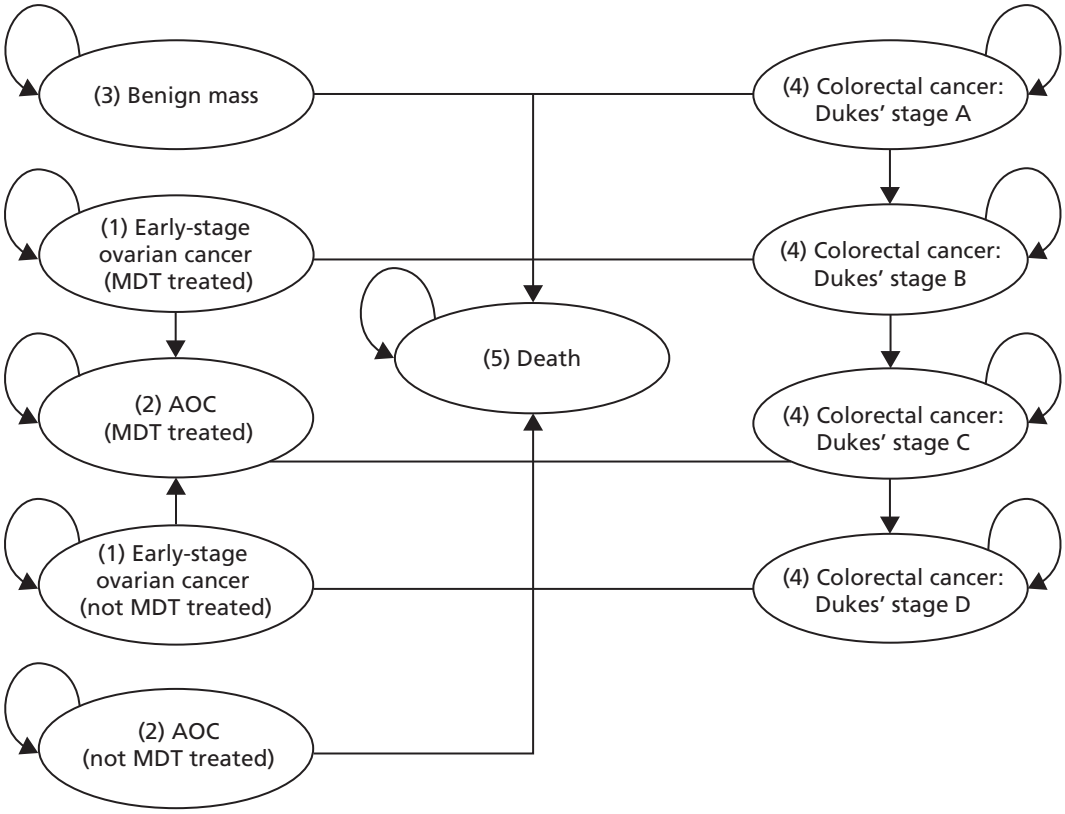

FIGURE 9 Markov process structure.

Model parameters

Estimates for the model input parameters were retrieved from the literature and by consulting experts for unpublished data. For consistency, and when the same parameters were required, the same sources were used as those used in CG122.[1] Accuracy estimates were derived from the systematic review component of this assessment (see *Results of the assessment of clinical effectiveness*). In case empirical estimates of standard errors were unavailable, it was assumed that the standard error would be equal to 20% of the expected values.

Probabilities not related to the risk scores

An overview of the disease-related (ovarian cancer, CRC and benign mass) probabilities for both the decision tree and the Markov model is provided in *Table 21*. It was assumed that all patients are female (used for utility estimation).

The prevalence of malignancies (all, including borderline and non-ovarian malignancies) as well as the proportion of women diagnosed with other malignancies (assumed to be CRC) were obtained using a random-effects meta-analysis (with log-transformation) of diagnostic cohort studies, included in our systematic review, which reported data for the relevant target condition and subgroup. The following parameters were estimated as in CG122:[1]

- percentage of early-stage ovarian cancer versus AOC
- 30 days post-surgery ovarian cancer mortality
- 10-year overall survival and progression-free survival of ovarian cancer [using updates of the same trials: the International Collaborative Ovarian Neoplasm (ICON) Group's study 1 was used to model progression-free survival and overall survival for early ovarian cancer;[133] and the ICON Group's study 3 was used to model these outcomes for AOC[134]].

TABLE 21 Ovarian cancer and CRC probabilities

	Estimate	SE	Distribution	Source (year of publication)
Decision tree (short term)				
Prevalence (all malignancies)	21.3%	1.0%	Beta	Aktürk *et al.* (2011),[71] Colombo *et al.* (2009),[127] Coleman *et al.* (2016),[70] Di Legge *et al.* (2012),[61] Jacobs *et al.* (1990),[78] Janas *et al.* (2015),[97] Knafel *et al.* (2016),[49] Lou *et al.* (2010),[73] Meys *et al.* (2016),[44] Moore *et al.* (2011),[101] Morgante *et al.* (1999),[80] Sayasneh *et al.* (2016),[46] Shulman *et al.* (2016),[104] Testa *et al.* (2014),[50] Timmerman *et al.* (2010),[65] van Gorp *et al.* (2012),[98] Xu *et al.* (2016)[95] and Yanaranop *et al.* (2016)[89]
Prevalence (all malignancies) – premenopausal women	16.2%	2.0%	Beta	Al Musalhi *et al.* (2016),[103] Coleman *et al.* (2016),[70] Janas *et al.* (2015),[97] Knafel *et al.* (2016),[49] Meys *et al.* (2016),[44] Piovano *et al.* (2016),[58] Sayasneh *et al.* (2013),[62] Testa *et al.* (2014),[50] van Gorp *et al.* (2012)[98] and Yanaranop *et al.* (2016)[89]
Prevalence (all malignancies) – postmenopausal women	45.9%	3.3%	Beta	Al Musalhi *et al.* (2016),[103] Coleman *et al.* (2016),[70] Janas *et al.* (2015),[97] Knafel *et al.* (2016),[49] Meys *et al.* (2016),[44] Piovano *et al.* (2016),[58] Sayasneh *et al.* (2013),[62] Testa *et al.* (2014),[50] van Gorp *et al.* (2012)[98] and Yanaranop *et al.* (2016)[89]
Prevalence non-ovarian malignancies (colorectal) within malignancies	2.9%	0.3%	Beta	Aktürk *et al.* (2011),[71] Colombo *et al.* (2009),[127] Coleman *et al.* (2016),[70] Di Legge *et al.* (2012),[61] Jacobs *et al.* (1990),[78] Janas *et al.* (2015),[97] Knafel *et al.* (2016),[49] Lou *et al.* (2010),[73] Meys *et al.* (2016),[44] Moore *et al.* (2011),[101] Morgante *et al.* (1999),[80] Sayasneh *et al.* (2016),[46] Shulman *et al.* (2016),[104] Testa *et al.* (2014),[50] van Gorp *et al.* (2012),[98] Xu *et al.* (2016)[95] and Yanaranop *et al.* (2016)[89]
Advanced stage if ovarian malignancy[a]	75%		Fixed	Bell *et al.* (1998)[128]
If CRC, proportion of Dukes' stage A	13.2%	0.1%	Dirichlet	National Cancer Registration and Analysis Service (2010)[129]
If CRC, proportion of Dukes' stage B	36.9%	0.1%	Dirichlet	
If CRC, proportion of Dukes' stage C	35.9%	0.1%	Dirichlet	
If CRC, proportion of Dukes' stage D	14.0%		Dirichlet	
If FN result, proportion of ovarian cancer	100.0%		Fixed	Assumption
If FP result, proportion having benign mass	100.0%		Fixed	Assumption
If TN result, proportion having benign mass	100.0%		Fixed	Assumption
30-day post-surgery mortality, early-stage ovarian cancer	1.1%	0.5%	Beta	National Collaborating Centre for Cancer (2011),[1] and Venesmaa and Ylikorkala (1992)[130]

continued

TABLE 21 Ovarian cancer and CRC probabilities (*continued*)

	Estimate	SE	Distribution	Source (year of publication)
30-day post-surgery mortality, AOC	2.9%	0.3%	Beta	National Collaborating Centre for Cancer (2011)[1] and Gerestein *et al.* (2009)[131]
30-day post-surgery mortality related to benign surgery	0.2%	0.0%	Beta	National Collaborating Centre for Cancer (2011),[1] Loft *et al.* (1991)[132]
Markov model (long-term)				
10-year progression-free survival for early-stage ovarian cancer	70.0%	4.7%	Beta	ICON Group study 1, Collinson *et al.* (2014)[133]
10-year overall survival for early-stage ovarian cancer	73.0%	4.0%	Beta	ICON Group study 1, Collinson *et al.* (2014)[133]
2-year overall survival for AOC	62.6%	1.8%	Beta	ICON Group study 3 (2002)[134]
HR overall survival with SMDT treatment vs. no SMDT treatment	0.900	0.048 = SE ln(HR)	Log-normal	Woo *et al.* (2012)[135]
HR progression-free survival with SMDT treatment vs. no SMDT treatment	Assumed equal to HR for overall survival given that overall survival and progression-free survival HRs for teaching vs. general hospitals are very similar[135]			
Annual progression for Dukes' stage A to B	58.3%	0.5%	Beta	Westwood *et al.* (2016)[136] and Tappenden *et al.* (2007)[137]
Annual progression for Dukes' stage B to C	65.6%	0.8%	Beta	
Annual progression for Dukes' stage C to D	86.7%	0.8%	Beta	
Mortality CRC	Time dependent, see *Appendix 7* in Westwood *et al.*[136] for more details			

HR, hazard ratio; ICON, International Collaborative Ovarian Neoplasm; SE, standard error; TN, true negative; TPs, true positives.
a Only for TPs (see text).

The following parameters were estimated as in the most recent diagnostic appraisal review (DAR) in CRC:[136]

- percentage in each of the Dukes' stages
- annual progression between Dukes' stages
- mortality by Dukes' stage.

The effect of SMDT treatment (i.e. with gynaecological oncologists on site) versus women treated in secondary care was estimated from Woo *et al.*,[135] who reported a hazard ratio (HR) of 0.90 (95% CI 0.820 to 0.990) for the overall survival of women with ovarian cancer treated in teaching hospitals versus general hospitals. This HR was also assumed for progression-free survival, as the analyses by Woo *et al.*[135] indicated that the HR for overall and progression-free survival for teaching hospitals versus general hospitals is very similar. This study was obtained from a focused literature search, which was pragmatic in design. For this, the following resources were searched:

- MEDLINE (via Ovid): 1946 to week 3 January 2017
- MEDLINE In-Process Citations (via Ovid): to 30 January 2017
- MEDLINE Daily Update (via Ovid): to 30 January 2017
- MEDLINE Epub Ahead of Print (via Ovid): to 30 January 2017
- EMBASE (via Ovid): 1974 to 30 January 2017
- Cochrane Database of Systematic Reviews (via Wiley Online Library): to Issue 1 of 12, January 2017

- Database of Abstracts of Reviews of Effects (via Wiley Online Library): to Issue 2 of 4, April 2015
- Cochrane Central Register of Controlled Trials (via Wiley Online Library): to Issue 11 of 12, November 2016
- NHS Economic Evaluation Database (via Wiley Online Library): to Issue 2 of 4, April 2015.

Full search strategies are presented in *Appendix 1*.

Finally, age-dependent mortality from the general population was used for women with a benign mass, after the 30-day post-surgery period. All input parameters for the Markov model are reported in *Table 21*.

Risk score accuracy parameters

The proportions of women testing positive (and thus referred to a SMDT) or negative were based on the estimated accuracy of the risk scores considered (see *Chapter 3, Selection of diagnostic performance estimates for inclusion in cost-effectiveness modelling* and *Table 22*) and the estimated prevalence of all malignancies detected in this population (21.3% with a standard error of 1.0%). The proportions of true positives (TPs), FPs, FNs and true negatives (TNs) were calculated as follows:

$$TP = \text{prevalence} \times \text{sensitivity}. \quad (5)$$

$$FP = (1 - \text{prevalence}) \times (1 - \text{specificity}). \quad (6)$$

$$FN = \text{prevalence} \times (1 - \text{sensitivity}). \quad (7)$$

$$TN = (1 - \text{prevalence}) \times \text{specificity}. \quad (8)$$

TABLE 22 Test accuracy

Risk score	Sensitivity (SE)	Specificity (SE)	Source (systematic review; see *Chapter 3*)
RMI 1 (threshold of 250)	64.4% (1.4%)	91.8% (0.7%)	Summary estimate derived from all studies, six published studies[73,74,76,78–80] and one unpublished study[a] that reported data for RMI 1 (at a threshold of 250) and the target condition 'all malignant tumours'
ROMA score using Abbott Diagnostics' ARCHITECT	75.0% (6.6%)	87.9% (2.7%)	Al Musalhi *et al.* (2016)[103] (see *Table 7*)
ROMA score using Roche Diagnostics' Elecsys	79.1% (2.4%)	79.1% (1.4%)	Summary estimate derived from two studies[97,104] (see *Table 10*)
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	90.2% (2.5%)	65.8% (1.9%)	Summary estimate derived from two studies[68,70] (see *Table 17*)
IOTA group's Simple Rules (inconclusive, assumed to be malignant)	94.2% (0.5%)	76.1% (0.6%)	Summary estimate derived from eight published studies[44,48–50,52,55,62,65] and one unpublished study (see *Table 12*)
IOTA group's ADNEX model (threshold of 10%)	96.3% (0.5%)	69.1% (0.9%)	Summary estimate derived from three published studies[17,44,46] and one unpublished study (see *Table 11*)
RMI 1 (threshold of 200)	68.1% (0.9%)	90.1% (0.5%)	Summary estimate derived from all studies, 12 published studies[44,48,50,62,73,74,76,78–80,98,103] and one unpublished study that reported data for the RMI 1 (threshold of 200) and the target condition 'all malignant tumours'

SE, standard error.
a Frances Nixon, personal communication.

Subsequently, the proportions of women who are referred to a SMDT (TPs and FPs), and who are not referred to a SMDT (TNs and FNs) were calculated. The results are listed in *Table 23*.

After the risk assessment and referral decision, women in the decision tree were allocated to 'early-stage ovarian cancer', 'AOC', 'benign mass', 'CRC' and 'death'. One of the main assumptions in the decision tree was that the women categorised as testing FN all had early-stage disease based on expert opinion (i.e. the value of 25% for early-stage ovarian cancer refers only to the TPs).

Health state utilities

For women with a benign mass, age-dependent general population utility estimates were used.[138] The utilities for early-stage and AOC were taken from Havrilesky *et al.*[139] and Grann *et al.*,[140] respectively. These estimates were also used in the economic model in CG122.[1] As in the latest CRC DAR, the study by Ness *et al.*[136,141] was used to inform utilities for the four stages of CRC. Utility estimates and sources are summarised in *Table 24*.

TABLE 23 Test outcomes

Risk score	TP (%)	FP (%)	FN (%)	TN (%)	PPV	NPV	LR+	LR–
RMI 1 (at a threshold of 250)	13.7	6.5	7.6	72.2	0.68	0.90	7.85	0.39
ROMA score using Abbott Diagnostics' ARCHITECT	16.0	9.5	5.3	69.2	0.63	0.93	6.20	0.28
ROMA score using Roche Diagnostics' Elecsys	16.9	16.4	4.5	62.2	0.51	0.93	3.78	0.26
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	19.2	26.9	2.1	51.8	0.42	0.96	2.64	0.15
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	20.1	18.8	1.2	59.9	0.52	0.98	3.94	0.08
IOTA group's ADNEX model (at a threshold of 10%)	20.5	24.3	0.8	54.4	0.46	0.99	3.12	0.05
RMI 1 (at a threshold of 200)	14.5	7.8	6.8	70.9	0.65	0.91	6.88	0.35

LR+, positive likelihood ratio; LR–, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

TABLE 24 Utility scores

Population	Estimate (SE)	Distribution	Source (year of publication)
Benign mass (assumed to be equal to the general population)	Age dependent		Ara *et al.* (2010)[138]
Early-stage ovarian cancer treated by a SMDT	0.830 (0.063)	Beta	Havrilesky *et al.* (2009)[139]
Early-stage ovarian cancer not treated by a SMDT	Assumed to be equal to treatment by a SMDT		
AOC treated by a SMDT	0.630 (0.247)	Beta	Grann *et al.* (1998)[140]
AOC not treated by a SMDT	Assumed to be equal to treatment by a SMDT		
CRC Dukes' stage A	0.740 (0.023)	Beta	Ness *et al.* (1999)[141]
CRC Dukes' stage B	0.670 (0.026)	Beta	
CRC Dukes' stage C	0.500 (0.031)	Beta	
CRC Dukes' stage D	0.250 (0.028)	Beta	

SE, standard error.

Resource use and costs related to the risk scores

Risk score costs are listed in *Table 25*.

To derive the costs associated with these risk scores, several assumptions were made. These pertained to the different components of the RMI 1 score, the ROMA score, Overa (MIA2G), the IOTA group's simple ultrasound rules and the ADNEX model, and are summarised in the following sections.

Ultrasound costs

All risk scores entail, or are intended to be derived partly from, TVS scans. The costs for these were informed by the costs for TVS scans used in CG122,[1] and inflated to 2015–16 Personal Social Services Research Unit (PSSRU) values, resulting in a cost of £77.

Cancer antigen 125 test costs

The RMI 1, the ROMA score, Overa (MIA2G) from Vermillion and the IOTA group's ADNEX model risk scores all require an estimated cost for CA125 tests. These can differ in practice depending on which company's test is used. Only Roche Diagnostics made CA125 costs available. However, these costs referred to only the cost of the kit, not to the overall CA125 test cost. The cost used here was therefore taken from CG122[1] and estimated to be £26 (adjusted for inflation).

Risk of Malignancy Index 1 costs

The RMI 1 entails ultrasound scans and CA125 testing. RMI 1 costs were therefore the sum of the costs of ultrasound scans and CA125 testing (£102).

International Ovarian Tumour Analysis group's simple ultrasound rules

The IOTA group's simple ultrasound rules entail the costs of ultrasound scans only (£77). The IOTA group stated that using the simple rules algorithm would not add to the time needed to conduct the scan or interpret the results (Thomas Walker, Maastricht University, 2016, personal communication).

The International Ovarian Tumour Analysis group's Assessment of Different NEoplasias in the adneXa model

The IOTA group's ADNEX model consists of the costs of ultrasound scans and CA125 testing, and was therefore estimated to be £102.

TABLE 25 Risk score costs

Risk score	Estimate (£)	SE (£)	Distribution	Source
RMI 1	102	20	Gamma	More detail on the calculation of these costs is provided in the following sections and in *Appendix 6*
ROMA score using Abbott Diagnostics' ARCHITECT	130	26	Gamma	
ROMA score using Roche Diagnostics' Elecsys	126	25	Gamma	
Overa (MIA2G) from Vermillion	176	35	Gamma	
IOTA group's simple ultrasound rules	77	15	Gamma	
IOTA group's ADNEX model	102	20	Gamma	

SE, standard error.

Vermillion's Overa (MIA2G)

Vermillion provided a cost estimate of £99 chargeable for the provision of its test. It was unclear whether or not this cost included all materials. The cost of ultrasound was added to it, resulting in £176. This was added because the manufacturer recommends the use of Overa (MIA2G) alongside clinical and radiological evaluation and states that the product is not recommended as a stand-alone screening or diagnostic test.

Risk of Ovarian Malignancy Algorithm (Abbott Diagnostics' ARCHITECT and Roche Diagnostics' Elecsys)

The estimation of costs related to HE4 testing relied on the information provided by the different manufacturers. The cost of ultrasound was added to both of the ROMA risk scores as determined in the final scope based on clinical expert opinion. Manufacturers of both tests stated that final costs may be subject to volume-based discounts. Not all companies provided the same cost items and assumptions were made in order to fill in data gaps. The following is a list of these assumptions:

- Cost per HE4 test kit:
  - The Abbott Diagnostics' ARCHITECT HE4 test kit cost provided by the manufacturer was £21.33 per single test (e-mail from Abbott Diagnostics to Thomas Walker, NICE, 2017, personal communication).
  - The Roche Diagnostics' Elecsys HE4 test kit cost provided by the manufacturer was £1594.65, with each kit containing 100 tests, resulting in a cost per test of £15.95.
- Capital costs:
  - Analyser equipment costs were assumed to be the same for Fujirebio Diagnostics' LUMIPULSE G HE4, Abbott Diagnostics' ARCHITECT HE4 and Roche Diagnostics' Elecsys HE4, and these were based on the average analyser equipment cost of Fujirebio Diagnostics' LUMIPULSE G1200 and G600II.
  - These analyser equipment costs were annuitised with an assumed lifetime of 10 years and using a discount rate of 3.5%, resulting in an annuity factor of 8.32.
  - To calculate the analyser equipment capital cost per each test, an average of 253 working days per year with 7 work hours per day was assumed, and it was also assumed that two tests on average would be run per hour. This resulted in 3542 HE4 tests run on one analyser per year. However, many tests were run at the same time and some laboratories would only run these tests weekly, whereas some others would run these daily, resulting in a relatively crude estimate of numbers of tests per year. The resulting capital cost per test might, therefore, be an overestimate. However, it amounted to only £1.92 per test and, therefore, did not significantly affect the model outcomes.
- Maintenance costs:
  - Only Fujirebio Diagnostics provided cost estimates for maintenance. The average of the maintenance costs for LUMIPULSE G1200 and G600 II were assumed to be representative of the maintenance costs for all of the analysers of the different manufacturers. The two options of maintenance cover (fully comprehensive and preventative) were assumed to be adopted in equal proportions.
- Quality control:
  - Abbott Diagnostics and Roche Diagnostics provided cost estimates for their quality control. These were applied for each.

- Calibration:
  - Calibration costs were provided by Abbott Diagnostics and Roche Diagnostics, but it was not clear how often the calibration costs provided by Abbott Diagnostics were going to be incurred. The Roche Diagnostics calibration costs were therefore applied to both tests.
- Shipping:
  - Only Fujirebio Diagnostics provided costs for shipping each month. These were assumed to apply to Roche Diagnostics. Abbott Diagnostics stated that one shipment per month was free of charge and that further shipments were unlikely, so no further shipment costs were added to the Abbott Diagnostics' test costs.
- Personnel costs:
  - An estimate of 0.05 hours to prepare and perform one test was used. Given that many tests can be performed at the same time, this is likely to be an overestimate. The personnel cost was assumed to be that of a health-care scientist derived from Curtis (*Unit Costs of Health and Social Care 2014*)[142] at £2.76 per test. Given that these costs are still relatively low, the potential overestimation of personnel costs was not likely to affect the model outcomes. These costs were therefore added to both Roche Diagnostics and Abbott Diagnostics' tests.

The Roche Diagnostics' Elecsys HE4 test costs amounted to £23.75 and the Abbott Diagnostics' ARCHITECT HE4 test costs amounted to £27.97. The main difference in costs between Roche Diagnostics' Elecsys HE4 and Abbott Diagnostics' ARCHITECT HE4 thus stemmed from the cost per kit; other differences were caused by shipping costs and quality control costs.

Resource use and costs not related to the risk scores

All women with a high-risk test result are assumed to be referred to the SMDT. Based on expert opinion, the additional resource use for this is assumed to be the cost of a SMDT meeting, assuming no additional cost of the surgery or other investigations. The cost of this meeting (£116) is estimated to be that of the SMDT meetings, from the 2015–16 NHS reference costs.[143]

The treatment of ovarian cancer may consist of surgery and/or chemotherapy or supportive care, depending on the disease stage. As assumed in CG122,[1] chemotherapy consists of six cycles of carboplatin for women with early-stage ovarian cancer and six cycles of carboplatin/paclitaxel for women with advanced-stage disease.[1] Surgery costs were also based on CG122,[1] and calculated as a weighted average of the relevant NHS reference costs, taking into account the probability of complications and the underlying disease (early-stage or AOC or benign mass).[143] The proportions of women receiving each type of care were based on CG122,[1] and the unit costs of care from the latest PSSRU publication.[144] The frequency of follow-up costs for women with ovarian cancer was based on CG122,[1] and the unit costs were based on the PSSRU publication.[144] Annual costs for the CRC states were estimated from the lifetime costs of CRC and mean survival as was done in the most recent DAR in CRC[136,137] (*Table 26*). Women with a high risk score but a benign mass would be identified at the SMDT meeting to be FP cases, and would undergo SMDT surgery without any further costs incurred in the tertiary care setting.

TABLE 26 Health state costs, event costs and unit prices

Health state or event	Estimate	SE	Distribution	Source
Health state costs of CRC				
CRC Dukes' stage A (lifetime costs)	£10,683	£3959	Gamma	Westwood *et al.* (2016)[136] and Tappenden *et al.* (2007)[137]
CRC Dukes' stage B (lifetime costs)	£18,015	£6677	Gamma	
CRC Dukes' stage C (lifetime costs)	£29,141	£10,800	Gamma	
CRC Dukes' stage D (lifetime costs)	£19,392	£7187	Gamma	
CRC Dukes' stage A (annual costs)	£264			Calculated (using mean survival as in the previous DAR); Westwood *et al.* (2016)[136]
CRC Dukes' stage B (annual costs)	£609			
CRC Dukes' stage C (annual costs)	£2039			
CRC Dukes' stage D (annual costs)	£8391			
SMDT visit				
SMDT visit (necessary or unnecessary for benign mass)	£116		Fixed	Calculated
Treatment costs of ovarian cancer/benign mass				
Chemotherapy for early-stage ovarian cancer (six cycles of carboplatin)	£1898	£380	Gamma	Calculated
Unit costs of one cycle of carboplatin	£316.29		Fixed	National Collaborating Centre for Cancer (2011)[1] and the *British National Formulary* (2016)[145]
Chemotherapy administration for early-stage ovarian cancer	£1417	£283		National Collaborating Centre for Cancer (2011)[1]
Simple parenteral chemotherapy administration (per cycle)	£236		Fixed	*NHS Reference Costs 2015-2016*– SB12Z[143]
Chemotherapy for AOC (six cycles of carboplatin and paclitaxel)	£5905	£1118	Gamma	National Collaborating Centre for Cancer (2011)[1]
Unit costs of paclitaxel	£667.88			National Collaborating Centre for Cancer (2011)[1] and the *British National Formulary* (2016)[145]
Chemotherapy administration for AOC	£1918	£384		Calculated
More complex parenteral chemotherapy administration (per cycle)	£320		Fixed	*NHS Reference Costs 2015-2016* – SB13Z[143]
Laparotomy malignancy without complication	£3615	£723	Gamma	*NHS Reference Costs 2015-2016* – M06C[143]
Laparotomy malignancy with complication	£4566	£913	Gamma	*NHS Reference Costs 2015-2016* – MA06A and MA06B[143]
Laparotomy benign mass without complication	£3.301	£660	Gamma	*NHS Reference Costs 2015-2016* – MA07G and MA08B[143]
Laparotomy benign mass with complication	£4112	£822	Gamma	*NHS Reference Costs 2015-2016* – MA07E, MA07F and MA08A[143]
Proportion complication laparotomy benign mass	5.0%	1.0%		National Collaborating Centre for Cancer (2011)[1]
Proportion complication laparotomy early ovarian cancer	5.0%	1.0%		
Proportion complication laparotomy AOC	12.5%	2.5%		

TABLE 26 Health state costs, event costs and unit prices (*continued*)

Health state or event	Estimate	SE	Distribution	Source
Number of hospital specialist care support visits	14.0	2.8	Gamma	
Unit costs of hospital specialist care support	£100		Fixed	*NHS Reference Costs 2015-2016* – SD03A[143]
Number of hospital specialist care visits	14.0	2.8	Gamma	National Collaborating Centre for Cancer (2011)[1]
Unit costs of hospital specialist care visit	£396		Fixed	*NHS Reference Costs 2015-2016* – SD01A[143]
Number of GP visits	1.0	0.2	Gamma	National Collaborating Centre for Cancer (2011)[1]
Unit costs of GP visits	£76		Fixed	PSSRU[144]
Number of district nurse visits	4.0	0.8	Gamma	National Collaborating Centre for Cancer (2011)[1]
Unit costs of district nurse visit	£42		Fixed	PSSRU[144]
Number of nurse specialist visits	2.0	0.4	Gamma	National Collaborating Centre for Cancer (2011)[1]
Unit costs of nurse specialist visits	£50		Fixed	PSSRU[144]
Total costs of supportive care	£7290			Calculated
Proportion of chemotherapy administered for early-stage ovarian cancer	50%	10%	Beta	National Collaborating Centre for Cancer (2011)[1]
Proportion chemotherapy AOC	95%	(100% minus the proportion of supportive care)		
Proportion surgery early ovarian cancer	100%		fixed	
Proportion surgery AOC	85%	17%	Beta	
Proportion supportive care early ovarian cancer	0%		Fixed	
Proportion supportive care AOC	5%	1%	Beta	
Total treatment costs for benign mass	£3342			Calculated
Total treatment costs for early-stage ovarian cancer	£5320			
Total treatment costs for AOC	£10,606			
Follow up costs				
Annual number of follow-up visits (years 1–3)	4	0.8	Gamma	National Collaborating Centre for Cancer (2011)[1]
Annual number of follow-up visits (> year 3)	1	0.2	Gamma	
Unit costs of follow-up visits	£92		Fixed	PSSRU[144]
Total annual follow-up costs (years 1–3)	£398			Calculated
Total annual follow-up costs (> year 3)	£92			

SE, standard error.

Overview of main model assumptions

The main assumptions in the health economic analyses were:

- All non-ovarian malignancies are CRC malignancies.
- False-negative tests are more likely to be early-stage ovarian cancer than AOC.
- For the IOTA group's simple ultrasound rules, inconclusive assessments would be assumed to be malignant.
- Carboplatin costs (six cycles) for early ovarian cancer and carboplatin plus paclitaxel costs (six cycles) for AOC [without bevacizumab (Avastin®; Roche Diagnostics, Hertford, UK)].
- List prices are used for carboplatin and paclitaxel.
- The HR of 0.900 retrieved from the Cochrane review by Woo *et al.*,[135] which focused on the comparison of institutions with gynaecologic oncologists on site versus community or general hospitals, is representative of the relative progression-free survival and overall survival for SMDT treatment versus no SMDT treatment.
- All FP and FN patients will be operated on for a benign mass.
- No disutility is incorporated for FP women (i.e. women who are incorrectly told that they have ovarian cancer).

The impact of all of the assumptions listed above on the model outcomes is explored in the scenario analyses.

Model analyses

Expected costs, LYs and QALYs were estimated for all risk scores from the perspective of the NHS. Discount rates of 3.5% and a half-cycle correction were applied for both costs and effects. Incremental costs and QALYs for each strategy versus the RMI 1 at a decision threshold of 250 and versus the next best alternative were calculated (full incremental analysis). The ICER was then calculated by dividing the incremental costs by the incremental QALYs. Probabilistic sensitivity analyses (PSAs; 15,000 simulations) were performed and cost-effectiveness acceptability curves were constructed.

Sensitivity analyses

Deterministic one-way sensitivity analyses were performed, using all input parameters incorporated as stochastic parameters in the PSAs as well as the discount rates, to assess the impact input parameters on the estimated outcomes. The results of these analyses are presented using tornado diagrams.

Scenario analyses

Various scenario analyses were performed to assess the impact of the assumptions on the estimated outcomes:

- Assuming a prevalence of 20% for all malignancies.
- Assuming a prevalence of 30% for all malignancies.
- Assuming a 0% prevalence of non-ovarian (CRC) malignancies.
- Assuming an equal proportion of early-stage versus AOC in the FN and TP groups (in the base case, it was assumed that FN women would all have early-stage ovarian cancer).
- Assuming that, for the IOTA group's simple ultrasound rules, subjective assessment would be used for inconclusive assessments (instead of being assumed to be malignant). A subjective assessment of the ultrasound images was done by experts or by level 2/3 examiners as per the EFSUMB classification system.
- Assuming equal test costs for all risk scores.
- Assuming that no ultrasound is performed in conjunction with the ROMA and Overa (MIA2G) risk scores, thus reducing the costs of these risk scores.
- Assuming additional costs for the FP group (surgery costs with malignancy instead of without) and additional costs for the FN group (additional costs of benign surgery).

- Assuming additional costs for the FP group (surgery costs with malignancy instead of without) and additional costs for the FN group (additional costs of benign surgery and SMDT costs).
- Assuming a discount of 92% for carboplatin (CG122:[1] a discount in England of 91.8%; and a discount in Wales of 92.1%).
- Assuming a discount of 95% for paclitaxel (CG122:[1] a discount in England of 91.0%; and a discount in Wales of 95.4%).
- Assuming an alternative HR for progression-free and overall survival for SMDT treatment versus no SMDT treatment (of 0.808).[146] This study was selected because it was not included in the Cochrane review that was used in the base case, and it provided an alternative HR ($n = 275$, $n = 238$ for this comparison).
- Assuming an alternative HR for progression-free and overall survival for SMDT treatment versus no SMDT treatment (of 0.990; the upper bound of the CI used in the base case).
- Assuming that the proportion of women receiving supportive care (for advanced-stage cancer) is 10% (instead of 5%).
- Assuming an alternative TVS cost of £142.46 [(MA36Z) instead of £76.75 based on CG122].[1]
- Assuming an alternative TVS cost of £142.46 [(MA36Z) instead of £76.75 based on CG122][1] and increasing the TVS cost for the IOTA groups' risk scores by 20% (to reflect the potential training costs).
- Assuming an additional SMDT cost of £2500 to reflect higher surgery costs, given that, according to expert opinion, 1 in 3 or 4 patients referred to a SMDT may receive extensive surgery for ovarian cancer (IPG 470), for which the price is unknown.
- Assuming 90% of the non-malignancy surgery and complication costs for TN, which reflects a scenario wherein 90% of the TN group are operated on (instead of all).
- Assuming Avastin for advanced-stage cancer. Assuming an additional treatment cost of £17,760 per treated woman[147] and assuming a median survival rate of 39.7 months (95% CI 36.0 to 44.2 months), derived from a clinically predefined high-risk subgroup of the ICON7 trial.[148] This subgroup was used because an overall survival benefit was recorded in poor-prognosis patients, in contrast with the study population as a whole, providing further evidence towards the optimum use of bevacizumab in the treatment of ovarian cancer.[148]
- Assuming a disutility for the FP group during the first year in a state-transition model of 0.100.
- Assuming a disutility for the FP group during the first year in a state-transition model of 0.010.
- Using the optimal RMI 1 threshold (i.e. the RMI 1 threshold is cost-effective at £20,000 and/or £30,000 per QALY gained in the former scenario), based on a comparison of only different RMI 1 thresholds (see *Appendix 7* for accuracy estimates).

Subgroup analyses

Various subgroup analyses were performed (if applicable, see *Appendix 7* for accuracy estimates):

- premenopausal women (mean age of 38 years, subgroup-specific accuracy data)
- postmenopausal women (mean age of 68 years, subgroup-specific accuracy data)
- using a baseline age of 50 years for the base case (instead of 40 years, no other changes)
- early-stage disease only
- advanced-stage disease only.

Results of the cost-effectiveness analyses

This section describes the results using probabilistic analyses for the base-case analysis. In addition, the sensitivity (deterministic), scenario (deterministic) and subgroup (probabilistic) analyses are described.

Base-case analysis

The base-case analysis included seven risk scores. *Tables 27* and *28*, as well as *Figure 10*, show the probabilistic results of this analysis. The RMI 1, with a threshold of 250, was the least effective (16.926 LYs, 13.820 QALYs) and the second cheapest (£5669). The IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant), was the cheapest (£5667) and the second most effective (16.954 LYs, 13.841 QALYs), thereby dominating the RMI 1 (at both the 200 and 250 thresholds). The IOTA group's ADNEX

TABLE 27 Probabilistic results for the base-case analysis: LYs

Risk score	LYs (95% CI)	Compared with the RMI 1 (at a threshold of 250)
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	16.954 (16.651 to 17.247)	0.029
RMI 1 (at a threshold of 250)	16.926 (16.619 to 17.223)	
RMI 1 (at a threshold of 200)	16.928 (16.621 to 17.225)	0.002
IOTA group's ADNEX model (at a threshold of 10%)	16.957 (16.653 to 17.250)	0.031
ROMA score using Abbott Diagnostics' ARCHITECT (at a threshold of 7.4%/25.3%)	16.934 (16.627 to 17.229)	0.008
ROMA score using Roche Diagnostics' Elecsys (at a threshold of 11.4%/29.9%)	16.936 (16.631 to 17.231)	0.011
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	16.950 (16.646 to 17.243)	0.024

TABLE 28 Probabilistic results for the base-case analysis: costs, QALYs and incremental analysis

			Compared with the RMI 1 (threshold of 250)			Full incremental analysis
Risk score	Costs, £ (95% CI)	QALYs (95% CI)	ΔCosts (£)	ΔQALYs	ΔCosts/ ΔQALYs	ΔCosts/ ΔQALYs
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5667 (4551 to 6941)	13.841 (13.477 to 14.154)	–2	0.021	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5669 (4553 to 6934)	13.820 (13.456 to 14.134)				Dominated
RMI 1 (at a threshold of 200)	5673 (4557 to 6939)	13.821 (13.456 to 14.135)	4	0.002	£2483	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5699 (4585 to 6973)	13.843 (13.480 to 14.155)	30	0.023	£1274	£15,304
ROMA score using Abbott Diagnostics' ARCHITECT	5707 (4593 to 6976)	13.825 (13.458 to 14.138)	38	0.005	£7506	Dominated
ROMA score using Roche Diagnostics' Elecsys	5713 (4597 to 6985)	13.826 (13.461 to 14.141)	44	0.007	£6409	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5775 (4655 to 7049)	13.837 (13.472 to 14.151)	105	0.017	£6038	Dominated

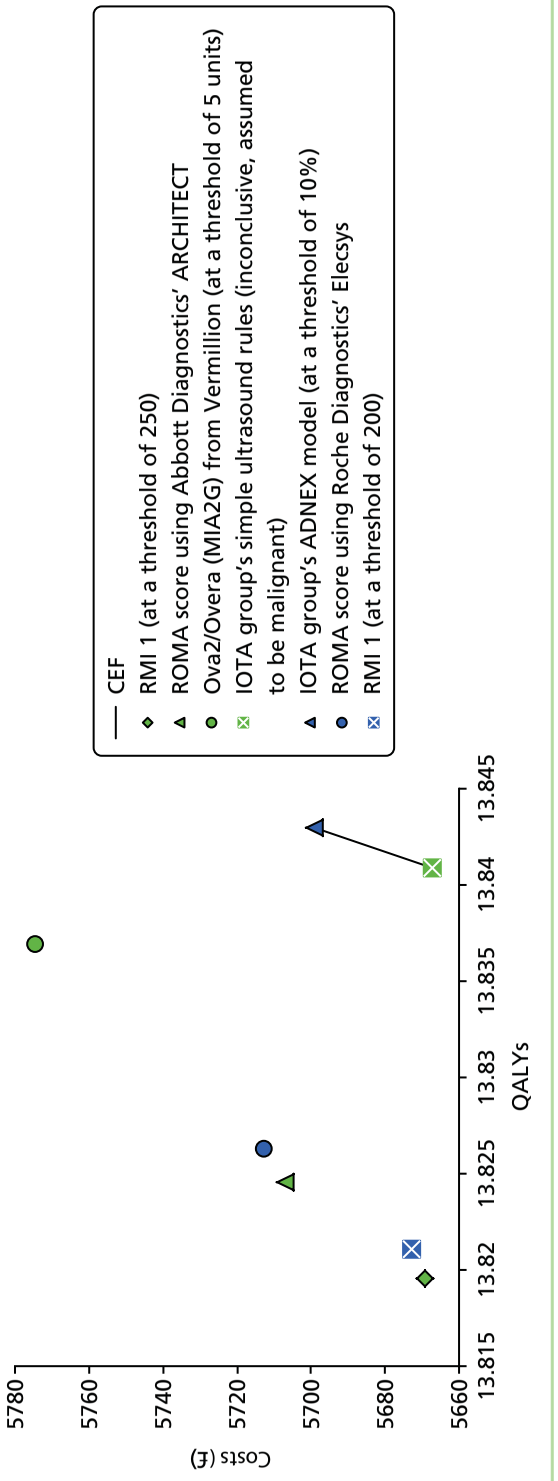

FIGURE 10 Probabilistic results. CEF, cost-effectiveness frontier.

model (at a threshold of 10%), costing £5699, was the most effective (16.957 LYs, 13.843 QALYs), and compared with the IOTA group's simple ultrasound rules, resulted in an ICER of £15,304 per QALY gained. The remaining risk scores [ROMA score using Abbott Diagnostics' ARCHITECT (at a threshold of 7.4%/25.3%); ROMA score using Roche Diagnostics' Elecsys (at a threshold of 11.4%/29.9%); and Overa (MIA2G) from Vermillion (at a threshold of 5 units)] were dominated. As a result, the incremental analysis indicated that, up to thresholds of £15,304 per QALY gained, the IOTA group's simple ultrasound rules are cost-effective, whereas the IOTA group's ADNEX model (threshold of 10%) is cost-effective for higher thresholds.

At willingness-to-pay thresholds of both £20,000 and £30,000 per QALY, the RMI 1 at a decision threshold of 250 had a probability of being cost-effective of 1%. For the IOTA group's simple ultrasound rules and the IOTA group's ADNEX model (threshold of 10%), this was 39% and 60%, respectively, at the £20,000 threshold, and 23% and 75%, respectively, at the £30,000 threshold. The probabilities for the other risk scores were < 1% for these thresholds (*Figure 11*).

Sensitivity analyses

The deterministic one-way sensitivity analyses (conditional upon the base-case analysis) indicated that the following parameters were the most influential in regard to the impact on the ICER versus the RMI 1 at the 250 threshold:

1. progression-free and overall survival HRs for SMDT referral versus no SMDT referral
2. test costs
3. utility of AOC
4. specialist multidisciplinary team costs
5. test sensitivity
6. discount rate
7. test specificity.

When considering a threshold of £20,000 per QALY gained, the IOTA group's ADNEX model (threshold of 10%) remained cost-effective, except in five sensitivity analyses wherein the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) became cost-effective:

- upper bound (0.990) for the overall survival HR for SMDT referral versus no SMDT referral
- upper bound (95.1%) for sensitivity for the IOTA group's simple ultrasound rules
- lower bound (95.3%) for sensitivity for the IOTA group's ADNEX model
- lower bound (£47) for the IOTA group's simple ultrasound rules costs
- upper bound (£142) for the IOTA group's ADNEX model costs.

When considering a threshold of £30,000 per QALY gained, the IOTA group's ADNEX model (threshold of 10%) remained cost-effective, except in two sensitivity analyses wherein the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) became cost-effective:

- upper bound (0.990) for the overall survival HR for SMDT referral versus no SMDT referral
- upper bound (£142) for the IOTA group's ADNEX model costs.

These results are shown in the tornado diagrams in *Appendix 8*.

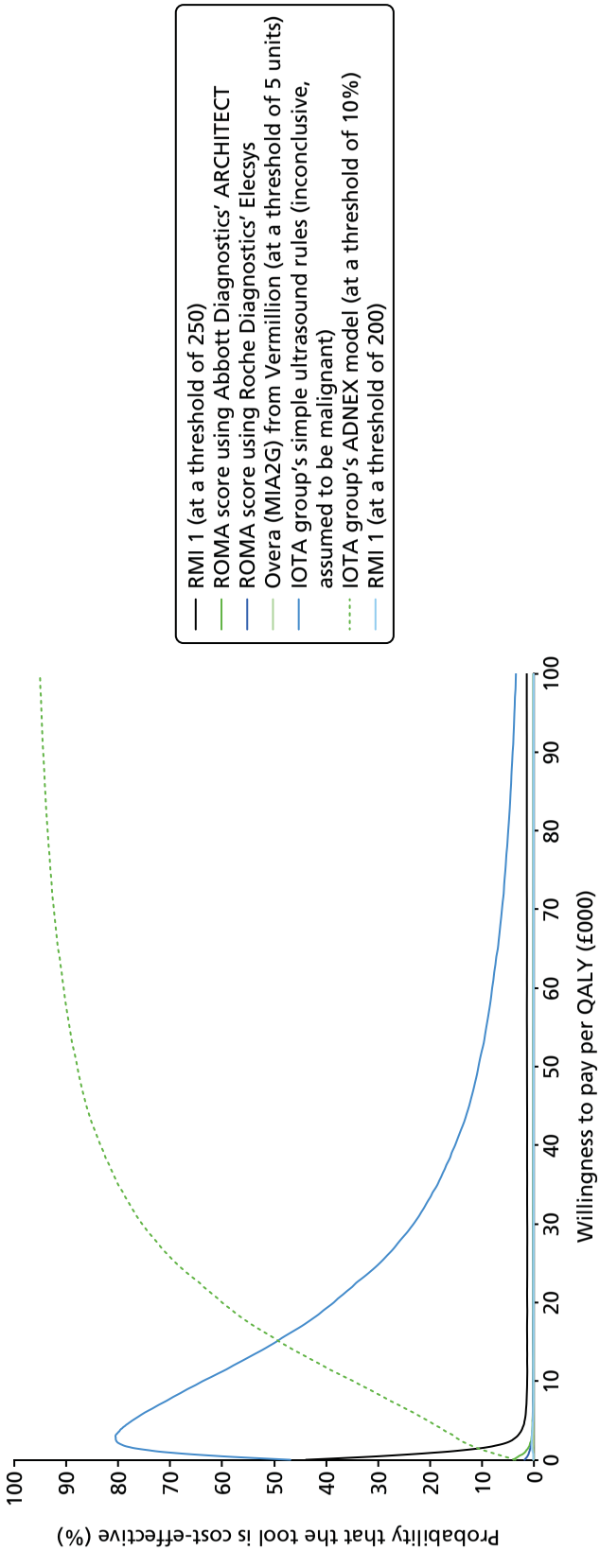

FIGURE 11 Cost-effectiveness acceptability curve for the base-case analysis.

Scenario analyses

The scenario analyses included the same seven risk scores. The tabulated results are provided in *Appendix 9*. The scenario analyses indicated that at thresholds of £20,000 and £30,000 per QALY gained, the IOTA group's ADNEX model (threshold of 10%) remained the most cost-effective strategy. This excludes the following scenarios:

- assuming an equal proportion of early-stage ovarian cancer versus AOC in the FN and TP groups (in the base case it was assumed, based on expert opinion, that the FN group would have only early-stage ovarian cancer)
- assuming 90% of the non-malignancy surgery and complications costs for the TN group, reflecting a scenario wherein 90% of the TN group are operated on (instead of all)
- assuming a disutility for the FP group during the first year in the state-transition model of 0.010.

In these scenario analyses, the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) was cost-effective at a threshold of £20,000 per QALY gained, whereas again the IOTA group's ADNEX model (threshold of 10%) was cost-effective at a threshold of £30,000 per QALY gained. Moreover, in the following scenario analyses, the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) became cost-effective at thresholds of £20,000 and £30,000 per QALY gained:

- assuming an alternative HR for progression-free and overall survival for SMDT referral versus no SMDT referral (of 0.990; the upper bound of the CI used in the base-case analysis)
- assuming a disutility for the FP group during the first year in the state-transition model of 0.100.

Finally, in the scenario with increased SMDT surgery costs, given that, according to expert opinion, 1 in 3 or 4 patients referred to a SMDT may receive extensive surgery for ovarian cancer[149] for which the price is unknown, the RMI 1 (threshold of 250) was cost-effective at a threshold of £20,000 per QALY gained, whereas the IOTA group's simple ultrasound rules were cost-effective at a threshold of £30,000 per QALY gained.

When comparing different RMI 1 thresholds only, it was found that the RMI 1 with a threshold of 25 would be cost-effective at all thresholds of £2890 per QALY gained or higher. However, when including this RMI 1 threshold with optimal sensitivity (instead of the RMI 1 with a threshold of 200) in the base-case analysis, the RMI 1 was still dominated.

Subgroup analysis

Premenopausal subgroup

The premenopausal subgroup analysis used a different mean starting age (38 years) and drew on subgroup-specific accuracy data (see *Appendix 7*). *Tables 29* and *30*, as well as *Figure 12*, show the probabilistic results of the subgroup analysis in premenopausal women. The ROMA score using Abbott Diagnostics' ARCHITECT was the least effective (18.108 LYs, 14.927 QALYs) and the RMI 1 with a threshold of 200 was the cheapest (£5188), followed by the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) at £5189. The most effective options were the IOTA group's ADNEX model (18.137 LYs, 14.948 QALYs) followed by the IOTA group's simple ultrasound rules (18.135 LYs, 14.946 QALYs). Consequently, the incremental analysis indicated that between thresholds of £15 and £18,304 per QALY gained, the IOTA group's simple ultrasound rules are cost-effective, whereas the IOTA group's ADNEX model (threshold of 10%) is cost-effective for higher thresholds.

TABLE 29 Probabilistic results for the premenopausal subgroup analysis

Risk score	LYs (95% CI)	Compared with the RMI 1 (at a threshold of 250)
RMI 1 (at a threshold of 200)	18.108 (17.435 to 18.720)	–0.006
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	18.135 (17.470 to 18.740)	0.021
RMI 1 (at a threshold of 250)		
ROMA score using Abbott Diagnostics' ARCHITECT	18.108 (17.434 to 18.720)	–0.006
IOTA group's ADNEX model (at a threshold of 10%)	18.137 (17.472 to 18.741)	0.024
ROMA score using Roche Diagnostics' Elecsys	18.132 (17.461 to 18.737)	0.018
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	18.131 (17.464 to 18.736)	0.018

TABLE 30 Probabilistic results for the premenopausal subgroup analysis: costs, QALYs and incremental analysis

			Compared with the RMI 1 (at a threshold of 250)			Full incremental analysis
Risk score	Costs, £ (95% CI)	QALYs (95% CI)	ΔCosts (£)	ΔQALYs	ΔCosts/ ΔQALYs	ΔCosts/ ΔQALYs
RMI 1 (at a threshold of 200)	5188 (4045 to 6510)	14.927 (14.309 to 15.471)	–7	–0.003	£1954	Cheapest
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5189 (4046 to 6515)	14.946 (14.331 to 15.486)	–6	0.016	Dominant	£15
RMI 1 (at a threshold of 250)	5195 (4051 to 6516)	14.93 (14.311 to 15.473)	0	0.000		Dominated
ROMA score using Abbott Diagnostics' ARCHITECT	5219 (4076 to 6542)	14.927 (14.308 to 15.471)	24	–0.004	Dominated	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5223 (4081 to 6549)	14.948 (14.335 to 15.487)	28	0.018	£1564	£18,466
ROMA score using Roche Diagnostics' Elecsys	5235 (4090 to 6571)	14.944 (14.329 to 15.484)	40	0.013	£2993	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5295 (4150 to 6617)	14.943 (14.327 to 15.484)	100	0.013	£7748	Dominated

At willingness-to-pay thresholds of both £20,000 and £30,000 per QALY gained, the RMI 1 (at a threshold of 250) had a probability of being cost-effective of < 1%. For the IOTA group's ADNEX model (at a threshold of 10%), the IOTA group's simple ultrasound rules and the ROMA score using Roche Diagnostics' Elecsys, the probability of being cost-effective was 46%, 37% and 16%, respectively, (at the £20,000 threshold) and 52%, 27% and 19%, respectively (at the £30,000 threshold). The probabilities for the other risk scores were < 2% for these thresholds (*Figure 13*).

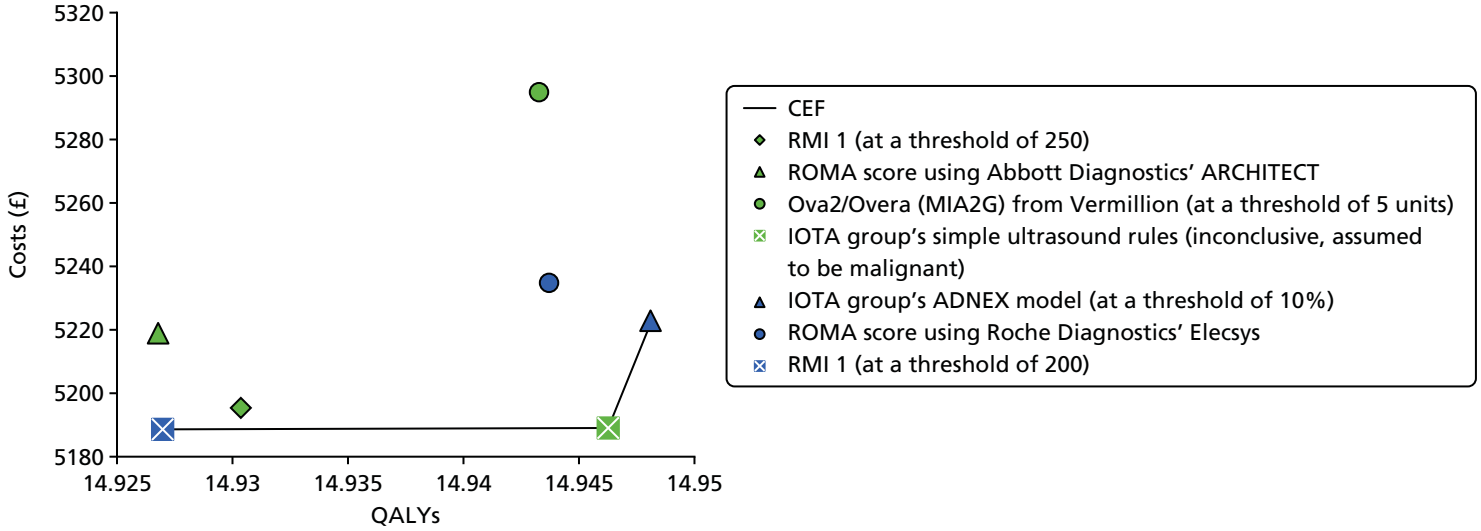

FIGURE 12 Probabilistic results for the premenopausal subgroup. CEF, cost-effectiveness frontier.

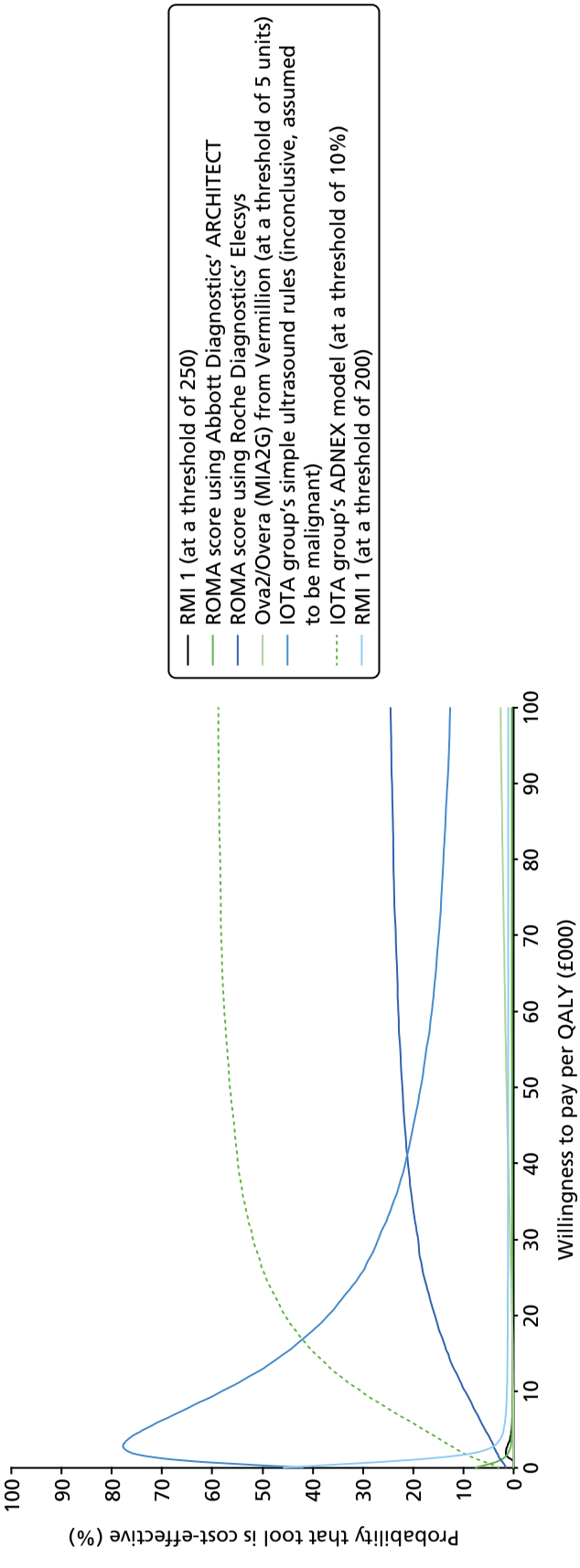

FIGURE 13 Cost-effectiveness acceptability curve for the premenopausal subgroup analysis.

Postmenopausal subgroup

The postmenopausal subgroup analysis used a different mean starting age (68 years) and drew on subgroup-specific accuracy data (see *Appendix 7*). *Tables 31* and *32*, as well as *Figure 14*, show the probabilistic results of the subgroup analysis in postmenopausal women (mean age 68 years). The RMI 1, with a threshold of 250, was the least effective (8.031 LYs, 5.690 QALYs). The cheapest risk score was the IOTA group's simple ultrasound rules (£7742), which was £1 cheaper than the RMI 1 with a threshold of 250. The most effective option was the IOTA group's ADNEX model (8.076 LYs, 5.721 QALYs). The IOTA group's simple ultrasound rules was the second most effective (8.072 LYs, 5.718 QALYs) and cost-effective up to a threshold of £12,876 per QALY gained; thereafter, the IOTA group's ADNEX model was cost-effective. The other risk scores [the RMI 1 at a threshold of 200, ROMA score using Abbott Diagnostics' ARCHITECT, ROMA score using Roche Diagnostics' Elecsys and Overa (MIA2G) from Vermillion] were dominated.

At willingness-to-pay thresholds of both £20,000 and £30,000 per QALY gained, the RMI 1 at the threshold of 250 had a probability of being cost-effective of < 2%. For the IOTA group's simple ultrasound rules and the IOTA group's ADNEX model (at a threshold of 10%), the probability of being cost-effective was 40% and 59%, respectively, at the £20,000 threshold, and 24% and 74%, respectively, at the £30,000 threshold. The probabilities for the other risk scores were < 1% for these thresholds (*Figure 15*).

Additional subgroup analyses

The results for the subgroup analyses, for which only the mean age was changed, to a mean age of 50 years (instead of 40 years as in the base case), and consisting of early-stage ovarian cancer only, were similar to the base-case results; at the thresholds of £20,000 and £30,000 per QALY gained, the IOTA group's ADNEX model (threshold of 10%) remained the most cost-effective strategy. In contrast, for the subgroup consisting of AOC only, the IOTA group's simple ultrasound rules were cost-effective at the thresholds of £20,000 and £30,000 per QALY gained. The tabulated results are provided in *Appendix 10*.

TABLE 31 Probabilistic results for the postmenopausal subgroup analysis (mean age 68 years): LYs

Risk score	LYs (95% CI)	Compared with the RMI 1 (at a threshold of 250)
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	8.072 (7.623 to 8.505)	0.042
RMI 1 (at a threshold of 250)	8.031 (7.57 to 8.472)	
RMI 1 (at a threshold of 200)	8.052 (7.597 to 8.487)	0.021
IOTA group's ADNEX model (at a threshold of 10%)	8.076 (7.626 to 8.508)	0.045
ROMA score using Abbott Diagnostics' ARCHITECT (at a threshold of 7.4%/25.3%)	8.069 (7.618 to 8.502)	0.038
ROMA score using Roche Diagnostics' Elecsys (at a threshold of 11.4%/29.9%)	8.051 (7.597 to 8.487)	0.020
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	8.068 (7.618 to 8.5)	0.037

TABLE 32 Probabilistic results for the postmenopausal subgroup analysis (mean age 68 years): costs, QALYs and incremental analysis

Risk score	Costs, £ (95% CI)	QALYs (95% CI)	Compared with the RMI 1 (at a threshold of 250)			Full incremental analysis
			ΔCosts (£)	ΔQALYs	ΔCosts/ ΔQALYs	ΔCosts/ ΔQALYs
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	7742 (6338 to 9281)	5.718 (5.061 to 6.178)	–1	0.028	Dominance	Cheapest
RMI 1 (at a threshold of 250)	7743 (6334 to 9289)	5.69 (5.035 to 6.153)	0	0.000		Dominated
RMI 1 (at a threshold of 200)	7765 (6356 to 9309)	5.703 (5.043 to 6.168)	22	0.013	£1746	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	7774 (6370 to 9318)	5.721 (5.063 to 6.181)	31	0.031	£1013	£12,876
ROMA score using Abbott Diagnostics' ARCHITECT (at a threshold 7.4%/ 25.3%)	7788 (6381 to 9329)	5.716 (5.059 to 6.177)	45	0.026	£1759	Dominated
ROMA score using Roche Diagnostics' Elecsys (at a threshold 11.4%/29.9%)	7789 (6377 to 9334)	5.702 (5.044 to 6.168)	46	0.012	£3738	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	7842 (6429 to 9396)	5.715 (5.058 to 6.176)	99	0.025	£3992	Dominated

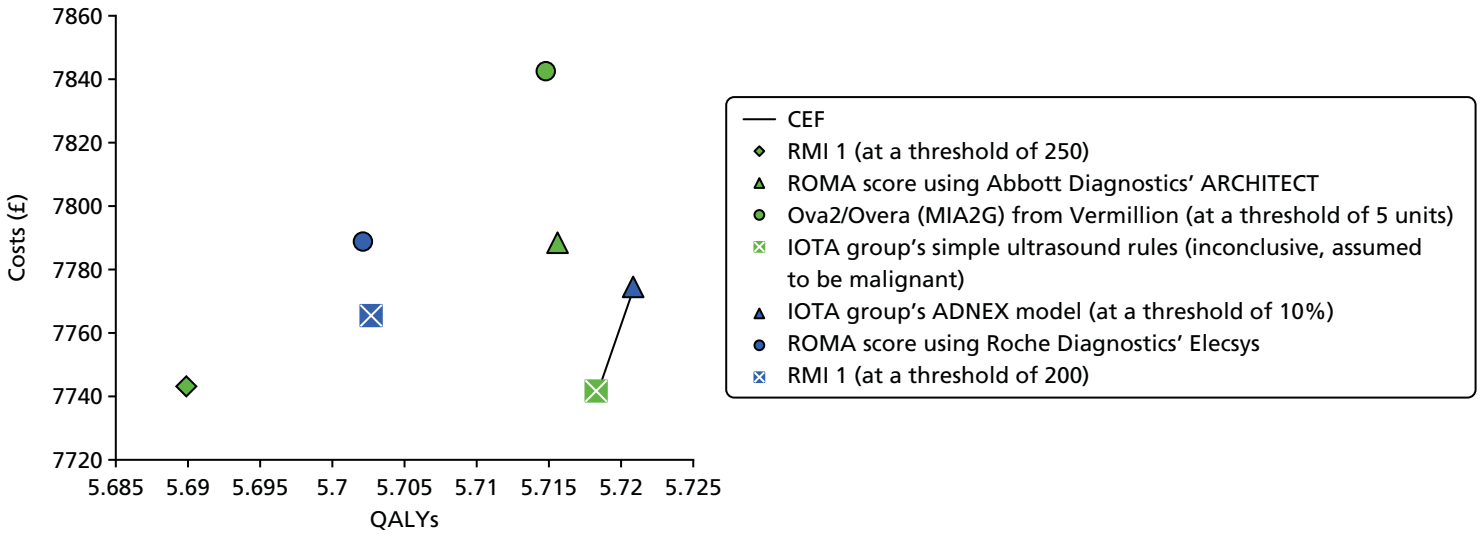

FIGURE 14 Probabilistic results for the postmenopausal subgroup (mean age 68 years). CEF, cost-effectiveness frontier.

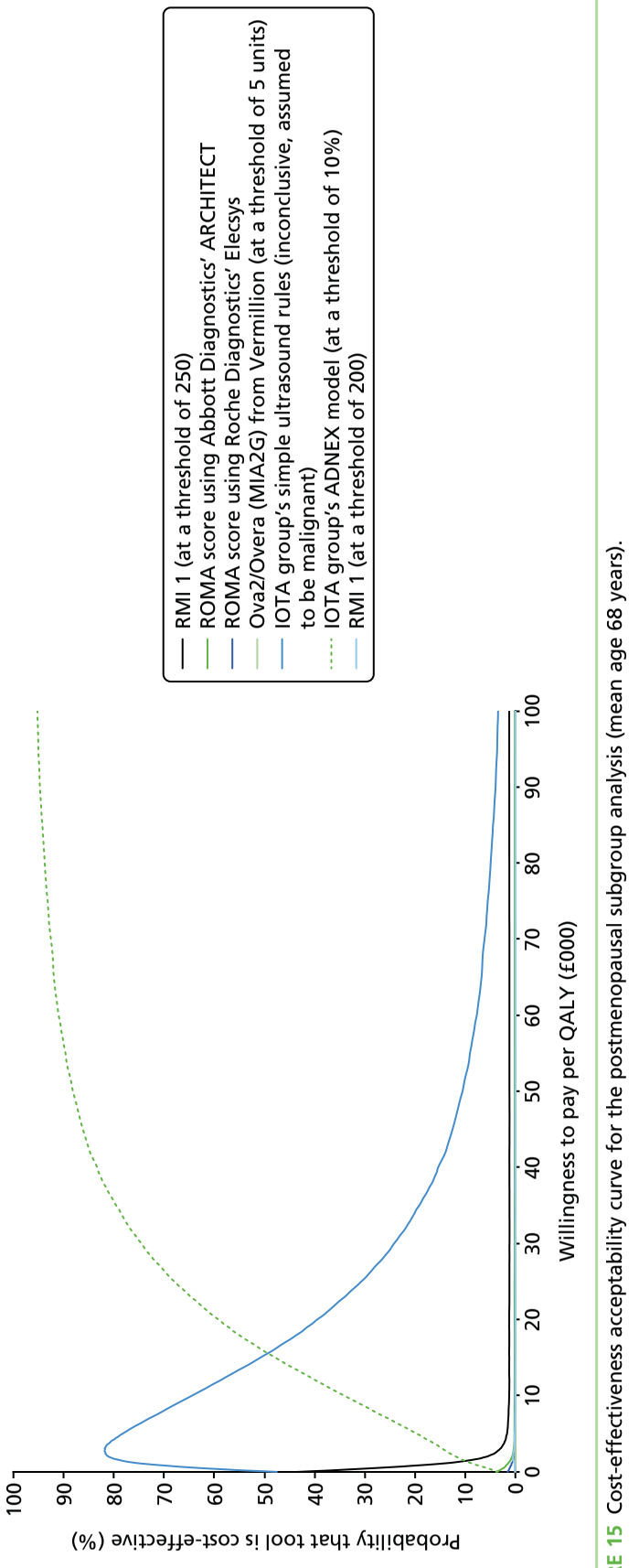

FIGURE 15 Cost-effectiveness acceptability curve for the postmenopausal subgroup analysis (mean age 68 years).

Chapter 5 Discussion

Statement of principal findings

Clinical effectiveness

All of the studies included in our systematic review were diagnostic cohort studies that reported data on the diagnostic accuracy of one or more ovarian cancer risk scores [the ROMA score, the IOTA group's simple ultrasound rules, the ADNEX model or Overa (MIA2G)] or provided data on the accuracy of the RMI 1 at different decision thresholds (including a threshold of 250, as specified in the current NICE guidelines[1]). With the exception of Overa (MIA2G), studies were identified that provided direct comparisons of the performance of each included risk score versus the RMI 1 (at a threshold of 200), that is, the performance of the intervention risk score and the performance of the RMI 1 (at a threshold of 200) were assessed in the same patient cohort. No study reported data to allow a direct comparison of all included index tests (risk scores) with each other and the RMI 1 in the same patient cohort. No RCTs or CCTs were identified, and no studies provided data on patient-relevant outcomes following different risk assessment strategies.

Studies evaluating the ROMA score used either the Roche Diagnostics Elecsys or the Abbott Diagnostics ARCHITECT tumour marker assays. None of the included studies used the Fujirebio Diagnostics LUMIPULSE G automated CEIA system. Two studies[94,98] that used a ROMA score based on the manual Fujirebio Diagnostics tumour marker EIAs (see *Appendix 5*, *Tables 41* and *42*) were included. These studies are included for information only and it should be noted that the manual assays are not specified interventions for this assessment.

The target condition for this assessment is ovarian cancer, defined as those conditions covered by NICE clinical guideline CG122,[1] namely epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma and borderline ovarian cancer; excluded conditions are pseudomyxoma peritonei, relapsed cancers, germ cell tumours of the ovary and sex cord–stromal tumours of the ovary. All studies in our systematic review included women with one or more adnexal or pelvic masses. The definition of ovarian cancer varied between studies and did not always include borderline tumours. In addition, the definition of disease/reference standard positive could include all malignancies or only ovarian malignancies. Although studies that report the performance of risk scores for the specific target condition of ovarian cancer (as described in CG122)[1] could be considered the most applicable to the scope of this assessment (and, accordingly, have been rated as giving rise to 'low concerns regarding applicability' in our QUADAS-2 assessments), it should be noted that the calculation of accuracy estimates for ovarian cancer or epithelial ovarian cancer requires the post hoc exclusion of women with other histological diagnoses from the analysis. In practice, such patients form part of the population in whom risk-scoring would be applied and, hence, their exclusion from the analyses may result in estimates of test performance that cannot be achieved in real-world clinical settings.

The majority of the studies that assessed the performance of the ROMA score used the Abbott Diagnostics ARCHITECT tumour marker assays. The summary sensitivity estimate for the ROMA score (using the manufacturer's recommended cut-off values of 7.4% in premenopausal women and 25.3% in postmenopausal women) was highest (96.4%, 95% CI 93.6% to 98.2%) when analyses excluded women with borderline tumours and those with malignancies other than epithelial ovarian cancer. It was lowest (75.0%, 95% CI 60.4% to 86.4%) when all women were included in the analysis, regardless of their final histopathological diagnosis and if different cut-off values (13.1% and 27.7%) were used; non-exclusion is more likely to reflect the performance of the score in a clinical setting. The study that included all women in the analysis reported similar sensitivity and specificity estimates for the ROMA score and the RMI 1 [(at a threshold of 200) 75%, 95% CI 60.4% to 86.4% vs. 77.1%, 95% CI 62.7% to 88.0%; and 87.9%, 95% CI 81.9% to 92.4% vs. 81.8%, 95% CI 75.1% to 87.4%, respectively].[103] By contrast, when women with borderline tumours and/or those with malignancies other than epithelial ovarian cancer were excluded

from the analyses, the summary specificity estimate for the ROMA score (53.3%, 95% CI 50.0% to 56.7%) was significantly lower than that for the RMI 1 [(at a threshold of 200) 80.3%, 95% CI 77.5% to 82.9%]. The only study to report a direct comparison of the ROMA score using Roche Diagnostics' Elecsys tumour marker assays (with the manufacturer's recommended thresholds of 11.4% in premenopausal women and 29.9% in postmenopausal women) with the RMI 1 included all study participants in the analysis irrespective of final histological diagnosis, but classified women with borderline tumours as being disease negative. In this study, the sensitivity estimate for the ROMA score appeared to be slightly higher than that for the RMI 1 (83.8%, 95% CI 73.4% to 91.3% vs. 78.4%, 95% CI 67.3% to 87.1%), and the specificity estimate for the ROMA score appeared to be slightly lower than that for the RMI 1 (68.8%, 95% CI 61.6% to 75.4% vs. 79.6%, 95% CI 73.1% to 85.1%), but neither difference was statistically significant.[89] This study may be considered to be more applicable to clinical practice if it is considered preferable to manage women with borderline tumours in non-specialist settings. The summary estimates of sensitivity and specificity for the ROMA score, using Roche Diagnostics' Elecsys tumour marker assays at the manufacturer's recommended thresholds, derived from non-comparative accuracy studies in which all women were included in the analysis (target condition: all malignancy) were 79.1% (95% CI 74.2% to 83.5%) and 79.1% (95% CI 76.3% to 81.6%), respectively. In studies in which the manufacturers' recommended cut-off values were used, the performance of the ROMA score did not differ significantly between premenopausal women and postmenopausal women.

When considering the risk-scoring methods produced by the IOTA group, our report focuses on data for the ADNEX model for which the validated 10% threshold is used and on data for the IOTA group's simple ultrasound rules, for which all study participants have an index test-based classification (either by assuming that inconclusive classifications are malignant or by applying subjective judgement to inconclusive assessments). Accuracy data for studies in which women with an inconclusive IOTA group's simple ultrasound rules assessment were not classified (excluded from the analyses) are reported in *Appendix 5*, *Table 39*. However, these results are considered to be of limited clinical value, as it is unclear which alternative methods might be used to select the most appropriate care pathway for these women. The majority of these studies included all participants in the analyses, irrespective of final histological diagnosis (i.e. the target condition was all malignant tumours including borderline). The summary estimates of sensitivity were high for both the ADNEX model (96.3%, 95% CI 95.3% to 97.1%) and the IOTA group's simple ultrasound rules for which inconclusive results were assumed to be malignant (94.2%, 95% CI 93.3% to 95.1%); when subjective assessment was applied to inconclusive and IOTA group's simple ultrasound rules results, the summary sensitivity estimate was significantly lower (88.4%, 95% CI 86.9% to 89.8%). Conversely, the summary estimates of specificity were low for both the ADNEX model (69.1%, 95% CI 67.4% to 70.8%) and the IOTA group's simple ultrasound rules, for which inconclusive results were assumed to be malignant (76.1%, 95% CI 74.9% to 77.3%), and significantly higher (92.5%, 95% CI 91.6% to 93.4%) when subjective assessment was applied to inconclusive results and the IOTA group's simple ultrasound rules results. Menopausal status did not significantly affect the performance of either the ADNEX model or the IOTA group's simple ultrasound rules, but the specificity estimate was significantly higher in premenopausal women than in postmenopausal women for both instruments. One published study[44] and one unpublished interim report (Frances Nixon, personal communication) provided comparative accuracy data for the ADNEX model, the IOTA group's simple ultrasound rules, for which inconclusive results were assumed to be malignant, and the RMI 1, using a decision threshold of 200. The summary estimates of sensitivity derived from these two studies were slightly higher for the ADNEX model (96%, 95% CI 94.5% to 97.1%) than for the IOTA group's simple ultrasound rules (92.8%, 95% CI 90.9% to 94.3%). Likewise, the summary estimates of specificity were similar (67%, 95% CI 64.2% to 69.6% and 71.6%, 95% CI 68.9% to 74.1%) for the ADNEX model and the IOTA group's simple ultrasound rules, respectively. The summary estimate of sensitivity for the RMI 1 at a decision threshold of 200 (66%, 95% CI 62.9% to 69%) was significantly lower than both the ADNEX model and IOTA group's simple ultrasound rules estimates. Conversely, the specificity estimate for the RMI 1 at a decision threshold of 200 was significantly higher (89%, 95% CI 87% to 90.7%) than both the ADNEX model and the IOTA group's simple ultrasound rules estimates.

No studies were identified that directly compared Overa (MIA2G) with the RMI 1 at either decision threshold (200 or 250). One study[104] reported comparative accuracy data for Overa (MIA2G) versus the ROMA score, using the Roche Diagnostics Elecsys tumour marker assays. This study included all participants in the analysis, regardless of their final histopathological diagnosis (target condition: all malignancies including borderline). At a threshold of 5 units, the sensitivity estimate for Overa (MIA2G) was 91% (95% CI 86.8% to 94%) and the specificity estimate was 65.5% (95% CI 62.0% to 68.8%). The sensitivity of the Overa (MIA2G) score was significantly higher than that of the ROMA score (79.2%, 95% CI 73.7% to 83.8%), whereas the specificity of the Overa (MIA2G) score was significantly lower than that of the ROMA score (78.9%, 95% CI 75.8% to 81.7%).

Summary estimates derived from studies that compared the diagnostic performance of different RMI 1 decision thresholds (between 25 and 500) and included all study participants in the analyses, regardless of final histopathological diagnosis (target condition: all malignant tumours including borderline), indicated that sensitivity and specificity estimates did not differ significantly between the two decision thresholds (200 and 250). At the decision threshold of 200, the sensitivity estimate was 70.8% (95% CI 65.2% to 75.6%) and the specificity estimate was 91.2% (95% CI 88.9% to 93.1%). At the decision threshold of 250, the sensitivity estimate was 69.0% (95% CI 63.7% to 73.9%) and the specificity estimate was 91.6% (95% CI 89.3% to 93.5%). The summary estimates of sensitivity for the RMI 1, derived from studies included in our systematic review, were lower than those reported in a recent systematic review[150] [75%, 95% CI 72% to 74% (based on 14 studies)]; however, the difference was not statistically significant and the specificity estimate was similar (92%, 95% CI 88% to 94%). It should be noted that this systematic review[150] included studies of women undergoing surgery for an adnexal mass and excluded any studies that selectively excluded some histopathological subtypes of ovarian cancer or classified borderline tumours as benign. As would be expected, the sensitivity estimate for the RMI 1 increased and the specificity decreased with decreasing threshold.

For both the IOTA group's simple ultrasound rules and the ADNEX model, there was evidence that specificity can be significantly decreased in postmenopausal women in comparison with overall populations or premenopausal women. Neither of these risk scores incorporates menopausal status; preliminary evidence suggests that menopausal status should be taken into account when applying these tools in practice.

The base case for the cost-effectiveness analysis considers the target condition 'all malignant tumours including borderline'. This is because the scope and protocol for this assessment specified that the definition of ovarian cancer should include borderline tumours. In addition, as previously outlined, the population in which risk-scoring would be applied in practice is likely to include some women who will ultimately be found to have a non-ovarian primary and some who will have cancers that fall outside the definition of ovarian cancer as used in CG122[1] (e.g. germ cell tumours and sex cord–stromal tumours of the ovary); therefore, it was considered that studies that include all participants in their analysis, irrespective of the final histological diagnosis, are more likely to produce estimates of risk-score performance that are representative of what might be expected in clinical practice. For all index tests (risk scores), there were no significant differences between the summary performance estimates calculated from all available data and those that included only those studies reporting a direct comparison with the RMI 1 (see *Chapter 3, Diagnostic performance of the Risk of Ovarian Malignancy Algorithm score*, *Diagnostic performance of International Ovarian Tumour Analysis group's simple ultrasound rules and the Assessment of Different NEoplasias in the adneXa model* and *Diagnostic performance of Overa (multivariate index assay, second generation)*. Therefore, the cost-effectiveness modelling used summary estimates of the diagnostic performance of risk scores, calculated using all available data sets for a given target condition. The ROMA score is considered to be a separate intervention for each tumour marker manufacturer (Roche Diagnostics' Elecsys and Abbott Diagnostics' ARCHITECT; none of the included studies used the Fujirebio Diagnostics LUMIPULSE G automated CEIA system and, therefore, the ROMA score using this assay option is not included in the cost-effectiveness analysis). Estimates of the diagnostic performance of the comparator, the RMI 1 with a decision threshold of 250, were derived from a meta-analysis of all available RMI 1 data sets with the corresponding target condition (e.g. all malignant tumours including borderline or all ovarian tumours including borderline) and population

(e.g. all participants, premenopausal women or postmenopausal women). When no data were available for the RMI 1 with a decision threshold of 250, data for a decision threshold of 200 were used; the analysis reported in *Chapter 3, Diagnostic performance of the Risk of Malignancy Index 1 using decision thresholds other than 250* indicated no significant difference in the performance of the RMI 1 at these two thresholds.

Cost-effectiveness

The review of economic analyses examined studies reporting outcomes of a full cost-effectiveness analysis, examining QALYs, with at least one of the comparators. In total, five studies were included, of which three studies reported QALYs as an outcome. Of these studies, one considered screening, whereas the remaining two considered secondary care from the UK and US perspectives. The UK study indicated that MMS consisting of CA125 testing followed by TVS could be cost-effective compared with USS and no screening. The two studies considering MIA, both from the US perspective, provided conflicting results: one study indicated that MIA might be cost-effective, whereas the other indicated that it was dominated by other strategies (when considering LYs). This latter study was the only one considering the ROMA score and also indicated that this score would be dominated by other strategies (when considering LYs). Moreover, this study indicated that a 'refer all' approach is cost-effective for thresholds above US$10,644 per LY gained. In conclusion, there is limited and conflicting evidence regarding the cost-effectiveness of alternative risk scores, which include HE4 testing, CA125 testing and ultrasound, compared with the RMI 1 score with a referral threshold of $\geq$ 250 (current UK practice) for people with suspected ovarian cancer in secondary care.

In our health economic analysis, the cost-effectiveness of different risk scores, which include HE4 testing, CA125 testing and ultrasound, compared with the RMI 1 score, as used in current practice for patients with suspected ovarian cancer in secondary care, was assessed to guide decisions about referral to a SMDT. The base-case analysis included seven risk scores:

1. RMI 1 (at a threshold of 250)
2. ROMA score using Abbott Diagnostics' ARCHITECT
3. ROMA using Roche Diagnostics' Elecsys
4. Overa (MIA2G) from Vermillion (at a threshold of 5 units)
5. IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)
6. IOTA group's ADNEX model (at a threshold of 10%)
7. RMI 1 (at a threshold of 200).

In the base-case analysis, the RMI 1 with a threshold of 250 was the least effective (16.926 LYs, 13.820 QALYs) and the second cheapest (£5669). The IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) was the cheapest (£5667) and the second most effective (16.954 LYs, 13.841 QALYs), and thereby dominated the RMI 1 (at both the 200 and 250 thresholds). The IOTA group's ADNEX model (threshold of 10%, cost of £5699) was the most effective (16.957 LYs, 13.843 QALYs), and compared with the IOTA group's simple ultrasound rules, resulted in an ICER of £15,304 per QALY gained. The remaining risk scores [ROMA score using Abbott Diagnostics' ARCHITECT, ROMA score using Roche Diagnostics' Elecsys and Overa (MIA2G) from Vermillion] were dominated. As a result, the incremental analysis indicated that, up to thresholds of £15,304 per QALY gained, the IOTA group's simple ultrasound rules are cost-effective, whereas the IOTA group's ADNEX model (threshold of 10%) is cost-effective for higher thresholds. Consequently, at willingness-to-pay thresholds of both £20,000 and £30,000 per QALY, the RMI 1 at a threshold of 250 had a probability of being cost-effective of 1%. For the IOTA group's simple ultrasound rules and IOTA group's ADNEX model (threshold of 10%), this was 39% and 60%, respectively, at the £20,000 threshold, and 23% and 75%, respectively, at the £30,000 threshold. The probabilities for the other risk scores were < 1% for these thresholds.

The sensitivity and scenario analyses indicated that the HR for SMDT referral versus no SMDT referral (for women with ovarian cancer) was the most influential parameter in the model, and the results are reasonably robust. Most scenario analyses indicated that, at thresholds of £20,000 and £30,000 per QALY gained, the IOTA group's ADNEX model (threshold of 10%) remained the cost-effective strategy. In two

scenario analyses, the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant) were considered to be cost-effective at a threshold of £20,000 and/or £30,000 per QALY gained. For the scenario comparing the optimal sensitivity RMI 1 threshold, which was found to be 25 (at all thresholds of £2890 per QALY gained or higher), the RMI 1 was still dominated.

For the premenopausal and postmenopausal subgroups, the IOTA group's ADNEX model (threshold of 10%) was cost-effective at thresholds of £20,000 and £30,000 per QALY gained.

Strengths and limitations of the assessment

Clinical effectiveness

We are not aware of any previous systematic review that has considered the performance of both ultrasound-based risk scores, such as the IOTA group's Simple Rules, and biomarker-based scores, such as the ROMA score and Overa (MIA2G). The most recent systematic review[151] of the ROMA score completed searching in November 2014. In addition, previous systematic reviews[9,151,152] of the ROMA score have focused on predicting ovarian cancer (no definition reported) or epithelial ovarian cancer and have combined data from studies using different manufacturers' tumour marker assays and thresholds, and have not clearly described how study participants with borderline tumours and those with non-ovarian primaries were classified. A more recent systematic review[150] (searches completed in July 2015) is available for ultrasound-based risk scores, but previous systematic reviews[150,153] have tended to focus on comparing these scores with subjective ultrasound evaluation rather than with other types of risk-scoring. Risk-scoring for ovarian malignancy is a rapidly evolving field and it is believed that the full update, comparing all options currently available to the NHS, provided by this assessment, will be of value to clinicians and decision-makers. In addition, there is currently a large, ongoing Cochrane review, entitled 'Symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer'[154] that will provide data on testing options that lie outside the scope of this assessment.

Extensive literature searches were conducted in an attempt to maximise the retrieval of relevant studies. These included electronic searches of a variety of bibliographic databases, as well as screening of clinical trials registers and conference abstracts to identify unpublished studies. Because of the known difficulties in identifying test accuracy studies using study design-related search terms,[30] search strategies were developed to maximise sensitivity at the expense of reduced specificity. Thus, large numbers of citations were identified and screened, relatively few of which met the inclusion criteria of the review.

The possibility of publication bias remains a potential problem for all systematic reviews. Considerations may differ for systematic reviews of test accuracy studies. It is relatively simple to define a positive result for studies of treatment (e.g. a significant difference between the treatment and control groups that favours treatment). This is not the case for test accuracy studies, which measure agreement between the index test and the reference standard. It would seem likely that studies finding greater agreement (high estimates of sensitivity and specificity) will be published more often, but the relative priorities given to sensitivity and specificity estimates may vary depending upon the intended application of the test. In addition, test accuracy data are often collected as part of routine clinical practice or by retrospective review of records; test accuracy studies are not subject to the formal registration procedures applied to RCTs and are therefore more easily discarded when results appear to be unfavourable. The extent to which publication bias occurs in studies of test accuracy remains unclear, but simulation studies have indicated that the effect of publication bias on meta-analytic estimates of test accuracy is minimal.[155] Formal assessment of publication bias in systematic reviews of test accuracy studies remains problematic, and reliability is limited.[29] A statistical assessment of publication bias in this review was not undertaken. However, our search strategy included a variety of routes to identify unpublished studies and resulted in the inclusion of a number of conference abstracts and an unpublished interim report (Frances Nixon, personal communication).

Despite our extensive searches, no studies were identified that assessed the diagnostic performance of the ROMA score using the Fujirebio Diagnostics' LUMIPULSE G automated CEIA system. It is not considered appropriate to treat ROMA scores calculated using different manufacturers' tumour marker assays as equivalent technologies, as each uses different thresholds and is CE marked for use with the specified tumour marker assays. Furthermore, no studies that reported a direct comparison of the diagnostic performance of the ROMA score using different manufacturers' tumour marker assays in the same patient cohort were identified.

No studies were identified that directly compared the performance of the Overa (MIA2G) score with that of the RMI 1 score; the data included in the systematic review component of this assessment refer only to the performance of the Overa (MIA2G) score compared with that of the ROMA score (using Roche Diagnostics' Elecsys tumour marker assays) and not to its performance in relation to the specified comparator, the RMI 1 score.

Clear inclusion criteria were specified in the protocol for this review, a copy of which is available online (www.nice.org.uk/guidance/GID-DG10012/documents/final-protocol, accessed 5 July 2018). The eligibility of studies for inclusion is therefore transparent. In addition, specific reasons for exclusion have been provided for all of the studies that were considered to be potentially relevant at the initial citation screening and were subsequently excluded on assessment of the full publication (see *Appendix 2*). The review process followed recommended methods to minimise the potential for error and/or bias;[27] studies were independently screened for inclusion by two reviewers and data extraction and quality assessment were done by one reviewer and checked by a second reviewer (MW, SD or SL). Any disagreements were resolved by consensus.

All studies included in this review were assessed for risk of bias and applicability using the QUADAS-2 tool,[36] which is recommended by the Cochrane Collaboration.[29] The QUADAS-2 tool is structured into four key domains, covering participant selection, index test, reference standard and the flow of patients through the study (including the timing of tests). Each domain is rated for risk of bias (low, high or unclear); the participant selection, index test and reference standard domain are also separately rated for concerns regarding the applicability of the study to the review question (low, high or unclear). The results of the QUADAS-2 assessment are reported, in full, for all included studies in *Appendix 3* and are summarised in *Chapter 3, Study quality*. Those studies that reported the development of risk scores, in addition to test accuracy data, were also assessed using the PROBAST.[37] The PROBAST has been designed to assess both the risk of bias and concerns regarding applicability of a study that evaluates (develops and/or validates) a multivariable diagnostic or prognostic prediction model. It has a domain-based structure, similar to that of QUADAS-2, and is intended to be used for the assessment of primary studies included in a systematic review. The PROBAST is not yet published but has been used with the consent of the steering group, of which the lead author of this assessment report is a member.

The studies included in our systematic review used a variety of definitions of disease/reference standard positive. In order to facilitate clinically relevant comparisons, it was decided to group studies according to whether or not they included borderline tumours in their definition of malignancy and whether women found to have non-ovarian malignancies were included in the analyses or excluded post hoc. However, a detailed breakdown of histopathological diagnoses was not always reported (see *Appendix 4*, *Table 36*), and, hence, the within-group variation in the distribution of diagnoses cannot be fully quantified.

There remains a further question regarding the clinical applicability of the studies included in this assessment. All study participants underwent surgery (i.e. a histological confirmation of disease status was available). In practice, risk scores may be used in secondary care to triage women to surgery or surveillance/conservative management, as well as to guide decisions about when surgery should be undertaken (referral to a specialist gynaecological oncology unit). This potential mismatch between the study populations and real-world clinical practice is reflected in the relatively high estimate for the prevalence of malignancy (21.3%) derived from the studies included in our systematic review. It should be noted that a lower prevalence of malignancy may also affect risk score performance in practice.

Approximately half of the included published studies[17,42–46,48–50,52,58,60,62–64,66,76,78–81,83,86,97,98] (25/49) were conducted in Europe, but only six studies[45,60,62,66,78,79] were conducted solely in the UK, and a further two were multinational studies[17,42] that included a UK centre. There were no studies of the ROMA score or Overa (MIA2G) that included UK participants. The data included in this report may therefore have limited applicability to UK settings, particularly in relation to the performance estimates for the ROMA score and Overa (MIA2G).

Although the sample sizes of studies included in our systematic review were generally large for diagnostic accuracy studies (median $n = 277$, range 48–2445), it should be noted that the largest data sets were derived from the various phases of the IOTA study, and these tended to dominate the analyses for the ADNEX model and the IOTA group's simple ultrasound rules. Only one report per intervention (IOTA group's simple ultrasound rules of the ADNEX model) was included for each phase of the IOTA study.

Cost-effectiveness

Our cost-effectiveness analysis is the most comprehensive to date in terms of the costs and consequences considered, as well as the number of relevant risk scores considered. Moreover, the de novo probabilistic model was based on a previously published model for CG122.[1] For the present analysis, a number of adjustments were made to the model, mostly to update cost estimates, and most of the assumptions were maintained.

The model was also informed by a comprehensive, high-quality systematic review of diagnostic test accuracy. Additional parameters were either those from the original CG122 model[1] or any of the further assessments, or, when necessary, were based on a focused literature review, prioritising the key input parameters (e.g. the HR for SMDT referral vs. no SMDT referral). Such a review is standard practice in economic modelling, given the large number of parameters required.

As in any economic model, a number of major and minor assumptions had to be made (see *Chapter 4*, *Model parameters*). It is important to understand the impact of these assumptions in order to correctly interpret the results of the model. The impact of most assumptions has been explored in sensitivity and scenario analyses. These analyses underscored the robustness of the base-case results.

Uncertainties

Clinical effectiveness

There remain a number of areas of uncertainty in relation to the performance characteristics of risk scores for ovarian cancer in specific subgroups of women; no study reported data on the effects of other risk factors, such as family history of ovarian cancer, on the performance of any risk for ovarian malignancy.

There is uncertainty about the downstream consequences of using the various risk-scoring options available to select the most appropriate care pathway for women with an adnexal mass (management by a general gynaecologist or referral to a SMDT). The limited data available for the ROMA score do not suggest any substantial performance advantage over current practice (the RMI 1), particularly when the more inclusive definition of malignancy is used (target condition: all malignant tumours including borderline). Consideration of the data from studies that reported accuracy estimates for both the whole-study population (target condition: all malignant tumours including borderline) and for selected populations in which participants found to have borderline tumours and/or those with rare ovarian cancers or non-ovarian primaries were excluded, indicates that women with borderline tumours and those with rare ovarian cancers or non-ovarian primaries may be disproportionately represented among those with FN, low-risk ROMA scores. One comparative accuracy ROMA score study,[99] using Abbott Diagnostics' ARCHITECT tumour marker assays, reported test performance estimates for the target condition epithelial ovarian cancer, calculated both with and without the inclusion of participants with borderline tumours; these data indicated that around half of the FN risk scores were accounted for by women with borderline tumours, 3 out of 6 (50%) using the ROMA

score and 7 out of 13 (54%) using the RMI 1.[99] Similarly, a comparative accuracy ROMA score study,[89] using Roche Diagnostics' Elecsys tumour marker assays, reported test performance estimates for the whole-study population and for a selected population in which eight (3%) women with non-epithelial ovarian cancer and non-ovarian primaries were excluded from the analysis; women with malignancies other than epithelial ovarian cancer accounted for four (50%) of the FN results using the ROMA score and three (37.5%) of the FN results using the RMI 1 score. The potential to detect non-epithelial ovarian cancers by including other tests (e.g. to measure AFP and beta-hCG, as recommended in CG122[1] for women aged < 40 years with suspected ovarian cancer) in the standard work-up is unclear and was outside the scope of this assessment.

One further non-comparative ROMA score study,[97] using Roche Diagnostics' Elecsys tumour marker assays, reported test performance estimates calculated both with and without the inclusion of participants with borderline tumours and those with non-ovarian primaries; these data indicated that women with borderline tumours and those with non-ovarian primaries accounted for a high proportion, 12 out of 14 (86%), of the FN risk scores observed.[97] It should be noted that these observations are based on small numbers of women. Furthermore, although other risk scores [Overa (MIA2G), the IOTA group's simple ultrasound rules and the ADNEX model] appear to offer increased sensitivity, data were not available to explore the distribution of histological diagnoses among those women with FN low-risk classifications. The downstream consequences of a FN low-risk classification are likely to differ between women with different histological cancer types and between those with borderline tumours and higher-stage malignancies. A more complete exploration of the types of women who are likely to be misclassified as being at a low risk of developing ovarian cancer, using the various risk-scoring options available, as well as an investigation of the downstream clinical consequences for these patients, is therefore needed.

The results of comparative accuracy studies, as noted in *Statement of principal findings*, *Clinical effectiveness*, indicate that both the ADNEX model and the IOTA group's simple ultrasound rules (for which inconclusive results are assumed to be malignant) offer substantial increases in sensitivity for the prediction of malignancy relative to the RMI 1 at a decision threshold of 200 or 250. The introduction of these scores into routine practice would therefore be likely to reduce the numbers of women with malignancy who are falsely classified as being at low risk of developing ovarian cancer. However, this increased sensitivity is accompanied by a decrease in specificity, and, hence, an increase in the numbers of women with benign disease who would be unnecessarily referred to a specialist gynaecological oncology MDT, relative to the numbers associated with risk-scoring using the RMI 1. This trade-off can be illustrated using a hypothetical cohort of 1000 patients: assuming an overall prevalence of malignancy of 21.3% (the estimate used for the base case in our cost-effectiveness analysis), the numbers of women with malignancy who would not be referred to a SMDT would be 18, 33 and 154, based on the ADNEX model, the IOTA group's simple ultrasound rules and the RMI 1, respectively, and, conversely, the corresponding numbers of 'unnecessary' referrals of women with benign disease would be 181, 155 and 60, respectively. To achieve a similar level of sensitivity to that of the ADNEX model or the IOTA group's simple ultrasound rules, using the RMI 1 would require a very low decision threshold; for the same sample cohort of 1000 women, a RMI 1 threshold of 25 would result in 16 women with malignancy who would not be referred to a SMDT and 335 'unnecessary' referrals of women with benign disease.

It should also be noted that the performance of risk-scoring tools that include morphological features seen on ultrasound is likely to be affected by the level of skill and experience of the ultrasonographers. This is particularly the case when the method of applying the score includes an unspecified element of subjective judgement (e.g. the IOTA group's simple ultrasound rules with expert subjective assessment for inconclusive results). The effect of the ultrasonographer's experience on measures of test performance was considered in our systematic review, but very few data were found to inform this question. The majority of the studies of the ADNEX model and the IOTA group's simple ultrasound rules were derived from the IOTA cohort and tended to use experienced ultrasound examiners and/or provide tool-specific pre-study training. One study[49] explicitly assessed the effect of the training level of examiners on the diagnostic performance of the IOTA group's simple ultrasound rules and found no significant differences in test performance between EFSUMB level 2/3 examiners and EFSUMB level 1 examiners, but it should be noted that the

information value of this study is limited, as all examiners received one half-day of practical training in the IOTA group's simple ultrasound rules before the study. Perhaps more interestingly, two of the studies[52,62] evaluating the IOTA group's simple ultrasound rules explicitly reported using ultrasound operators with lower levels of experience: '63% of operators had performed fewer than 1000 scans, 24% were medical doctors and 76% were ultrasonographers';[62] 'Ultrasound examinations were performed by a fourth year trainee and junior staff in obstetrics and gynaecology who had less than one year of ultrasound experience, under the supervision of an expert examiner'.[52] Test performance estimates from both of these studies were similar to the overall summary estimates (see *Table 12* and *Chapter 3*, *Diagnostic performance of the Risk of Ovarian Malignancy Algorithm score*), providing some indication that the IOTA group's simple ultrasound rules may remain effective in the hands of less experienced operators. A more complete assessment of the levels of training and experience needed to achieve the required levels of test performance would inform implementation considerations (e.g. training requirements for secondary care ultrasonographers, increases in specialist cancer centre workload arising from the use of triage methods with higher sensitivity and the introduction of routine TVS in secondary care assessment).

Risk-scoring tools that are based solely on ultrasound also carry the inherent limitation that they cannot be used to assess women who are symptomatic but do not have a mass that is large enough to be detected as abnormal on an ultrasound scan; none of the studies identified by this systematic review included women in this group.

The ideal method of comparing the downstream resource use and clinical consequences of using the various risk-scoring options available would be a RCT comparing treatment pathways and subsequent clinical outcomes following risk-scoring by different methods. No randomised or non-randomised controlled trials that met the inclusion criteria for our systematic review were identified. A recently published RCT,[156] conducted in asymptomatic postmenopausal women with an incidentally detected adnexal mass on ultrasound, compared two risk assessment protocols based on the RMI 1 and on the IOTA group's simple ultrasound rules. This study found that more of the women who were assessed using the RMI 1 protocol than those assessed using the IOTA group's simple ultrasound rules protocol had surgery [18/68 (28.1%) vs. 7/68 (10.3%), with a relative risk of 2.57 (95% CI 1.15 to 5.76)]; there were no significant differences in rates of referral to a tertiary oncology unit or in delayed cancer diagnoses at 12 months. These findings are unlikely to be applicable to the population of interest in this assessment, as the prevalence of malignancy was much lower (2.7%) in this study population than in women referred to secondary care for the investigation of an adnexal mass, as seen in the studies included in our systematic review (median 29.9%, range 15–48.4%). The question of how different risk-scoring strategies affect referral rates and subsequent clinical outcomes in this population remains outstanding.

Cost-effectiveness

The economic analyses emphasise the importance of prioritising the sensitivity of the risk scores over the specificity. The benefits of referring as many women as possible or, if possible, all women with ovarian cancer to the SMDT outweigh the additional SMDT costs, even the additional SMDT costs related to unnecessary referrals (i.e. for FPs). More specifically, informal analyses using 100% sensitivity and 0% specificity indicate that a 'refer all to SMDT' strategy without risk scores might be cost-effective at the thresholds of £20,000 and £30,000 per QALY gained. It is, however, questionable as to whether or not such a strategy would be feasible for clinical practice, considering, among other factors, the potentially limited SMDT capacity. This is particularly because clinician opinion indicated that the real impact of FPs is probably not the additional cost but the time/resources taken away from TP women. This is an area of uncertainty, mainly because limited capacity is currently not considered in the economic model. Another logistic aspect that was not considered was a potential difference in time taken from entering the secondary care pathway to a confirmed diagnosis. Adding this might, for instance, favour the IOTA group's simple ultrasound rules (given that these are ultrasound only) and/or the ROMA score (if an ultrasound scan has already been done in primary care). However, no evidence was found to inform this potential difference in time between the strategies or to inform any possible consequences (e.g. utility increment associated with earlier diagnosis).

Other areas of uncertainty were those relating to risk-score costs. However, scenario analyses using equal risk-score costs indicated that this would not alter the conclusions. Other potentially relevant scenarios, such as (1) excluding the cost of CA125 testing from the ADNEX model (as the test can be used without CA125) and (2) adding the cost of CA125 testing to the IOTA group's simple ultrasound rules (in case the assay is still run in parallel), may be of interest. However, there are currently insufficient test accuracy data to support these analyses, and it is likely that these scenarios would result in improved cost-effectiveness for the ADNEX model. Furthermore, the handling of patients with malignancies other than ovarian cancer by assuming that all women have CRC in the model is a simplifying assumption made in line with CG122[1] to avoid additional complexity in the modelling. This simplifying assumption was shown not to be influential in a scenario in which women with other malignancies were not modelled, which is explained by the same approach being adopted for all risk scores, thus not affecting the results of the incremental analyses.

The main driver of the model results is the progression-free and overall survival HRs for SMDT referral versus no SMDT referral. This HR was obtained from a Cochrane review.[135] Although this Cochrane review[135] concluded that the evidence was consistent and stronger for ovarian cancer, it was stated to be low-quality evidence (because of the high risk of bias of the included retrospective observational studies). In addition, it is unclear whether or not this HR is representative of the difference between SMDT referral and no SMDT referral in the UK; this HR might be country specific as a result of differences between health systems. It is, however, reassuring that the review[135] included only studies that were performed in developed countries (Canada, the Netherlands, the UK and the USA). Nevertheless, given the aforementioned considerations, this HR should be considered an area of uncertainty. Furthermore, scenario analyses indicated that the SMDT (surgery) costs as well as a potential disutility for FP groups (favouring risk scores with higher specificity and hence informing the trade-off between sensitivity and specificity) are areas of uncertainty. With regard to the SMDT costs, feedback from clinicians highlighted that SMDT costs used in the model (obtained from NHS reference costs) were likely to be an underestimate and did not appropriately reflect the high costs associated with extensive surgery, which is performed in a proportion of women undergoing surgery in this setting. It should therefore be borne in mind that the RMI 1 (threshold of 250) was cost-effective at a threshold of £20,000 per QALY gained, whereas the IOTA group's simple ultrasound rules were cost-effective at a threshold of £30,000 per QALY gained in the scenario when higher SMDT costs were used. Unfortunately, there was no evidence to justify this increased cost of surgery.

Chapter 6 Conclusions

Implications for service provision

There is evidence to suggest that using either the IOTA group's ADNEX model or the IOTA group's simple ultrasound rules to assess the risk of malignancy in women with an adnexal mass may offer increased sensitivity relative to current practice (the RMI 1 at a decision threshold of 250 or 200), that is, a higher proportion of those women who have a malignant tumour would be referred to a SMDT. Both the IOTA group's ADNEX model and IOTA group's simple ultrasound rules have a lower specificity than the RMI 1 at a decision threshold of 250 or 200 and, hence, if the RMI 1 were replaced with either of these methods, it is also likely that more women with benign tumours would be 'unnecessarily' referred to a SMDT, with the attendant implications for an increased workload. However, to achieve a similar sensitivity using the RMI 1 would require a very low decision threshold (25) and hence a lower specificity and a greater number of unnecessary referrals than that achievable using either the IOTA group's ADNEX model or the IOTA group's simple ultrasound rules. The limited available evidence suggested that the ROMA score does not offer any clear performance advantage over the RMI 1. Although Overa (MIA2G) appeared to have a higher sensitivity than the ROMA score, there were no data to support a direct comparison between Overa (MIA2G) and the RMI 1.

In the base-case analysis, the IOTA group's ADNEX model (threshold of 10%) was considered to be cost-effective at thresholds of £20,000 and £30,000 per QALY gained. However, both cost and QALY differences between the strategies were small. This means that ICERs can change substantially, especially with small changes in either costs or QALYs. Therefore, it is difficult to be confident that other strategies, particularly the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant), which was cost-effective in some scenario analyses, might not be cost-effective. This is illustrated in the probabilities of being cost-effective for the IOTA group's simple ultrasound rules and the IOTA group's ADNEX model (threshold of 10%), which was 39% and 60%, respectively, at the £20,000 threshold and 23% and 75%, respectively, at the £30,000 threshold.

For the premenopausal and postmenopausal subgroups of women, the IOTA group's ADNEX model (threshold of 10%) remained cost-effective at the thresholds of £20,000 and £30,000 per QALY gained.

Overall, the model does provide evidence to strongly prioritise sensitivity over specificity. As a result, the IOTA group's ADNEX model (threshold of 10%), which had the highest sensitivity (96.3%) was considered to be cost-effective.

Suggested research priorities

In addition to information about the diagnostic performance of different risk-scoring methods, it is important to understand the consequences of applying these scores in practice:

- Further studies are required to explore the distribution of histological diagnoses among patients with FN, low-risk classifications. A more complete exploration of the types of women who are likely to be misclassified as having a low risk of developing ovarian cancer, using the various risk-scoring options available, as well as an investigation of the downstream clinical consequences for these women is required. If one or more of the risk scores evaluated in this assessment is introduced into routine practice, a postimplementation audit would be informative.

- Studies designed to capture the downstream resource use and clinical consequences of using the various risk-scoring options are likely to be informative. An example of such a study might be a cluster RCT, in which general gynaecology departments are randomised to use different risk-scoring methods to inform decisions about referral to a SMDT; outcomes could include rates of referral, staging investigations, surgery in a specialist setting (gynaecological oncologist), postsurgical outcomes and survival measures. Such a study could also inform issues around the costs and logistics of delivering various strategies in the NHS.

There are a number of areas that require further investigation if the introduction of ultrasound-based risk-scoring systems is being considered:

- An assessment of the levels of training and experience needed to achieve the required levels of test performance when using risk scores that include morphological features observed on ultrasound examination is required.
- Given the implementation issues around the use of risk-scoring systems that require ultrasound examination by expert or specifically trained personnel, research should consider whether or not implementation could be better delivered through one-stop clinics, similar to those used to assess postmenopausal bleeding. Such one-stop clinics, in which women could be seen by specialist gynaecologists and scanned by IOTA-trained personnel, may overcome some of the potential hurdles of implementing an imaging-based approach and interpreting imaging information in the context of other observations.
- Further studies or further analyses of the IOTA data set are needed to understand the role of menopausal status in the performance of both the IOTA and ADNEX model tests.
- Studies on the acceptability (the likely uptake) of TVS for women being assessed in general gynaecology (secondary care) settings may also be useful.

Further large diagnostic cohort studies are needed to fully evaluate the performance of the ROMA score (using different manufacturers' tumour marker assays) and Overa (MIA2G) compared with that of the RMI 1 at a decision threshold of 250 or 200, or at a lower threshold(s) if this is considered to be appropriate or if current guidance changes. These studies should be conducted in a population that includes the full spectrum of differential diagnoses that are likely to be present in women referred to secondary care for the investigation of an adnexal mass.

Further diagnostic cohort studies or subgroup analyses of existing data sets are needed to fully explore the possible variation in the accuracy of all risk scores in relevant subgroups (e.g. menopausal status and family history of ovarian cancer).

Given the areas of uncertainty highlighted in the *Chapter 5, Discussion*, the feasibility of a 'refer all to SMDT' strategy should be considered. If this strategy is not deemed feasible, the thresholds of the risk scores should examined, bearing in mind that sensitivity should be prioritised over specificity, and also bearing in mind the available SMDT capacity.

Acknowledgements

The authors would like to acknowledge the clinical advice and expert opinion provided by specialist members of the assessment subgroup: Dr Cathie Sturgeon, Consultant Clinical Scientist, Royal Infirmary of Edinburgh; Mr Jed Hawe, Consultant Gynaecologist, Countess of Chester Hospital NHS Foundation Trust; Dr Jurjees Hasan, Consultant Medical Oncologist, The Christie Hospital, Manchester; Dr Michael Weston, Consultant Radiologist, St James's University Hospital, Leeds; Professor Richard Edmonds, Professor of Gynaecological Oncology, University of Manchester; Dr Sudha Sundar, Senior Lecturer and Consultant in Gynaecological Oncology, City Hospital, Birmingham; and Dr Tracie Miles, Clinical Nurse Specialist, Royal United Hospital, Bath.

We are grateful to the following researchers for providing additional information and unpublished data to support this assessment: the IOTA study group; Mona Aarenstrup Karlsen, Herlev University Hospital; Dr Sharon O'Toole, University of Dublin; Dr Elisa Piovano, University Hospital, Turin; and Dr Julie Clemente, Department of Obstetrics and Gynaecology, Alfonso Specialist Hospital, Manila.

We would also like to thank the lay members of the NICE Diagnostics Advisory Committee and Assessment Subgroup for providing input on the patients' perspective at key stages of the assessment process.

Contributions of authors

Marie Westwood, Shona Lang and **Sohan Deshpande** planned and performed the systematic review and interpretation of the evidence.

Bram Ramaekers and **Sabine Grimm** planned and performed the cost-effectiveness analyses and interpreted the results.

Shelley de Kock devised and performed the literature searches and provided information support to the project.

Nigel Armstrong contributed to the planning and interpretation of the systematic review and cost-effectiveness analyses and the acquisition of input data, and conducted the model peer review.

Manuela Joore and **Jos Kleijnen** provided senior advice and support to the cost-effectiveness analyses and the systematic review, respectively.

All authors were involved in drafting and/or commenting on the report.

Data-sharing statement

All available data are contained within the report. All queries should be submitted to the corresponding author.

References

1. National Collaborating Centre for Cancer. *Ovarian Cancer: The Recognition and Initial Management of Ovarian Cancer*. Manchester: National Institute for Health and Care Excellence; 2011. URL: www.nice.org.uk/guidance/cg122/evidence/full-guideline-181688797 (accessed 22 September 2016).
2. Myers ER, Bastian LA, Havrilesky LJ, Kulasingam SL, Terplan MS, Cline KE, *et al.* Management of adnexal mass. *Evid Rep Technol Assess* 2006;**130**:1–145.
3. Prat J. FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication. *J Gynecol Oncol* 2015;**26**:87–9. https://doi.org/10.3802/jgo.2015.26.2.87
4. Office for National Statistics (ONS). *Cancer Registration Statistics, England: 2014. Cancer Diagnoses and Age-Standardised Incidence Rates for All Cancer Sites by Age, Sex and Region*. Newport: ONS; 2016. URL: www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2014/pdf (accessed 22 September 2016).
5. Cancer Research UK. *Ovarian Cancer (C56-C57.4): 2013. Number of New Cases, Crude And European Age-Standardised (AS) Incidence Rates per 100,000 Population, Females, UK*. London: Cancer Research UK; 2015. URL: www.cancerresearchuk.org/sites/default/files/cstream-node/cases_rates_ovary_1.pdf (accessed 22 September 2016).
6. Cancer Research UK. *Ovarian Cancer (C56-C57.4): 2012–2014. Average Number of Deaths Per Year and Age-Specific Mortality Rates per 100,000 Population, UK*. London: Cancer Research UK; 2015. URL: www.cancerresearchuk.org/sites/default/files/cstream-node/deaths_crude_ovary_M14.pdf (accessed 22 September 2016).
7. Cancer Research UK. *Ovarian Cancer (C56-C57.4): 2011–2013. Average Number of New Cases Per Year and Age-Specific Incidence Rates per 100,000 Population, Females, UK*. London: Cancer Research UK; 2015. URL: www.cancerresearchuk.org/sites/default/files/cstream-node/cases_crude_ovary_1.pdf (accessed 22 September 2016).
8. Cancer Research UK. *Ovarian Cancer (C56): 2014. Proportion of Cancers Diagnosed at Each Stage, All Ages, England*. London: Cancer Research UK; 2016. URL: www.cancerresearchuk.org/sites/default/files/cstream-node/inc_stage_ovarian_0.pdf (accessed 22 September 2016).
9. Wang J, Gao J, Yao H, Wu Z, Wang M, Qi J. Diagnostic accuracy of serum HE4, CA125 and ROMA in patients with ovarian cancer: a meta-analysis. *Tumour Biol* 2014;**35**:6127–38. https://doi.org/10.1007/s13277-014-1811-6
10. British Society for Gynaecological Endoscopy, Royal College of Obstetricians and Gynaecologists (RCOG). *Management of Suspected Ovarian Masses in Premenopausal Women*. London: RCOG; 2011. URL: www.ncin.org.uk/view?rid=1740 (accessed 22 September 2016).
11. Khoja L, Nolan K, Mekki R, Milani A, Mescallado N, Ashcroft L, *et al.* Improved survival from ovarian cancer in patients treated in Phase III trial active cancer centres in the UK. *Clin Oncol* 2016;**28**:760–5. https://doi.org/10.1016/j.clon.2016.06.011
12. Anton C, Carvalho FM, Oliveira EI, Maciel GA, Baracat EC, Carvalho JP. A comparison of CA125, HE4, risk ovarian malignancy algorithm (ROMA), and risk malignancy index (RMI) for the classification of ovarian masses. *Clinics* 2012;**67**:437–41. https://doi.org/10.6061/clinics/2012(05)06
13. Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;**112**:40–6. https://doi.org/10.1016/j.ygyno.2008.08.031

14. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I, International Ovarian Tumor Analysis (IOTA) Group. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. *Ultrasound Obstet Gynecol* 2000;**16**:500–5. https://doi.org/10.1046/j.1469-0705.2000.00287.x

15. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, *et al.* Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008;**31**:681–90. https://doi.org/10.1002/uog.5365

16. Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, *et al.* Predicting the risk of malignancy in adnexal masses based on the Simple Rules from the International Ovarian Tumor Analysis group. *Am J Obstet Gynecol* 2016;**214**:424–37. https://doi.org/10.1016/j.ajog.2016.01.007

17. Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, *et al.* Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ* 2014;**349**:g5920. https://doi.org/10.1136/bmj.g5920

18. Van Calster B, Van Hoorde K, Froyman W, Kaijser J, Wynants L, Landolfo C, *et al.* Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors. *Facts Views Vis Obgyn* 2015;**7**:32–41.

19. Scottish Intercollegiate Guidelines Network (SIGN), Healthcare Improvement Scotland. *Management of Epithelial Ovarian Cancer*. SIGN135. Edinburgh: SIGN; 2013. URL: http://sign.ac.uk/pdf/sign135.pdf (accessed 12 October 2016).

20. National Institute for Health and Care Excellence (NICE). *Suspected Cancer: Recognition and Referral*. NICE Guideline [NG12]. Manchester: NICE; 2015. URL: http://nice.org.uk/guidance/ng12 (accessed 4 October 2016).

21. Royal College of Radiologists (RCR). *About iRefer*. London: RCR; 2016. URL: www.rcr.ac.uk/clinical-radiology/being-consultant/rcr-referral-guidelines/about-irefer (accessed 12 October 2016).

22. National Institute for Health and Care Excellence (NICE). *Guidance on the Use of Palitaxel in the Treatment of Ovarian Cancer*. Technology appraisal guidance [TA55]. London: NICE; 2005. URL: http://nice.org.uk/guidance/ta55 (accessed 12 October 2016).

23. National Institute for Health and Care Excellence (NICE). *Bevacizumab in Combination with Paclitaxel and Carboplatin for First-Line Treatment of Advanced Ovarian Cancer*. Technology appraisal guidance [TA284]. London: NICE; 2013. URL: http://nice.org.uk/guidance/ta284 (accessed 12 October 2016).

24. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007;**7**:3. https://doi.org/10.1186/1471-2288-7-3

25. National Institute for Health and Care Excellence (NICE). *Olaparib for Maintenance Treatment of Relapsed, Platinum-Sensitive, BRCA Mutation-Poisitive Ovarian, Fallopian Tube and Peritoneal Cancer after Response to Second-Line or Subsequent Platinum-Based Chemotherapy*. Technology appraisal guidance [TA381]. London: NICE; 2016. URL: http://nice.org.uk/guidance/ta381 (accessed 12 October 2016).

26. National Institute for Health and Care Excellence (NICE). *Topotecan, Peylated Liposomal Doxorubicin Hydrochloride, Paclitaxel, Trabecteden and Gemcitabine for Treatment Recurrent Ovarian Cancer*. Technology appraisal guidance [TA389]. London: NICE; 2016. URL: http://nice.org.uk/guidance/ta389 (accessed 12 October 2016).

27. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York: University of York; 2009. URL: www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm (accessed 23 March 2011).

28. National Institute for Health and Care Excellence (NICE). *Diagnostics Assessment Programme Manual*. Manchester: NICE; 2011. URL: www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf (accessed 4 October 2016).

29. Cochrane Diagnostic Test Accuracy Working Group. *Handbook for DTA Reviews*. London: The Cochrane Collaboration; 2009. URL: http://srdta.cochrane.org/handbook-dta-reviews (accessed 23 March 2011).

30. Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol* 2011;**64**:602–7. https://doi.org/10.1016/j.jclinepi.2010.07.006

31. Canadian Agency for Drugs and Technologies in Health (CADTH). *CADTH Peer Review Checklist for Search Strategies*. Ottawa, ON: CADTH; 2013. URL: www.cadth.ca/en/resources/finding-evidence-is (accessed 17 July 2013).

32. Wright K, McDaid C. *Is the Retraction of Journal Articles in Electronic Journals and Databases Consistent and Timely? A Case Study*. Poster presented at the 19th Cochrane Colloquium, Madrid, Spain, October 2011.

33. Wright K, McDaid C. Reporting of article retractions in bibliographic databases and online journals. *J Med Libr Assoc* 2011;**99**:164–7. https://doi.org/10.3163/1536-5050.99.2.010

34. Royle P, Waugh N. Should systematic reviews include searches for published errata? *Health Info Libr J* 2004;**21**:14–20. https://doi.org/10.1111/j.1471-1842.2004.00459.x

35. Waffenschmidt S. *Assessing the Completeness of Systematic Reviews via the 'Related Articles' Function and/or a Simple Structured Boolean Search in PubMed – A Pilot Study (B202)*. Paper presented at the 19th Cochrane Colloquium, Madrid, Spain, October 2011.

36. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

37. Wolff R, Whiting P, Mallett S, Riley R, Westwood M, Kleijnen J. *PROBAST: Prediction Model Risk of Bias Assessment Tool*. Evidence Synthesis Network: systematic reviews of prognostic studies: new approaches to prognostic reviews and qualitative evidence synthesis, Manchester, UK, 27 May 2014.

38. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90. https://doi.org/10.1016/j.jclinepi.2005.02.022

39. Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A, Bachmann LM. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;**61**:1095–103. https://doi.org/10.1016/j.jclinepi.2007.09.013

40. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;**8**:239–51. https://doi.org/10.1093/biostatistics/kxl004

41. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;**6**:31. https://doi.org/10.1186/1471-2288-6-31

42. Szubert S, Wojtowicz A, Moszynski R, Zywica P, Dyczkowski K, Stachowiak A, *et al.* External validation of the IOTA ADNEX model performed by two independent gynecologic centers. *Gynecol Oncol* 2016;**142**:490–5. https://doi.org/10.1016/j.ygyno.2016.06.020

43. Joyeux E, Miras T, Masquin I, Duglet PE, Astruc K, Douvier S. Before surgery predictability of malignant ovarian tumors based on ADNEX model and its use in clinical practice. *Gynecol Obstet Fertil* 2016;**44**:557–64. https://doi.org/10.1016/j.gyobfe.2016.07.007

44. Meys EM, Jeelof LS, Achten NM, Slangen BF, Lambrechts S, Kruitwagen RF, *et al.* Estimating the risk of malignancy in adnexal masses: an external validation of the ADNEX model and comparison with other frequently used ultrasound methods. *Ultrasound Obstet Gynecol* 2016;**12**:12.

45. Moffatt J, Laver L, Dunford F, Cullimore J. Predicting the risk of malignancy of ovarian masses using the ADNEX (Assessment of Different NEoplasias in the adneXa) model: a retrospective study at the Great Western Hospital. *BJOG* 2016;**123**(Suppl. 2):103–4.

46. Sayasneh A, Ferrara L, De Cock B, Saso S, Al-Memar M, Johnson S, *et al.* Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study. *Br J Cancer* 2016;**115**:542–8. https://doi.org/10.1038/bjc.2016.227

47. Tinnangwattana D, Vichak-Ururote L, Tontivuthikul P, Charoenratana C, Lerthiranwong T, Tongsong T. IOTA simple rules in differentiating between benign and malignant adnexal masses by non-expert examiners. *Asian Pac J Cancer Prev* 2015;**16**:3835–8. https://doi.org/10.7314/APJCP.2015.16.9.3835

48. Abdalla N, Bachanek M, Trojanowski S, Cendrowski K, Sawicki W. Diagnostic value of ultrasound indicators of neoplastic risk in preoperative differentiation of adnexal masses. *J Ultrason* 2013;**13**:145–54. https://doi.org/10.15557/JoU.2013.0015

49. Knafel A, Banas T, Nocun A, Wiechec M, Jach R, Ludwin A, *et al.* The prospective external validation of International Ovarian Tumor Analysis (IOTA) simple rules in the hands of level I and II examiners. *Ultraschall Med* 2016;**37**:516–23. https://doi.org/10.1055/s-0034-1398773

50. Testa A, Kaijser J, Wynants L, Fischerova D, Van Holsbeke C, Franchi D, *et al.* Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. *Br J Cancer* 2014;**111**:680–8. https://doi.org/10.1038/bjc.2014.333

51. Tantipalakorn C, Wanapirak C, Khunamornpong S, Sukpan K, Tongsong T. IOTA simple rules in differentiating between benign and malignant ovarian tumors. *Asian Pac J Cancer Prev* 2014;**15**:5123–6. https://doi.org/10.7314/APJCP.2014.15.13.5123

52. Alcázar JL, Pascual MÁ, Olartecoechea B, Graupera B, Aubá M, Ajossa S, *et al.* IOTA simple rules for discriminating between benign and malignant adnexal masses: prospective external validation. *Ultrasound Obstet Gynecol* 2013;**42**:467–71. https://doi.org/10.1002/uog.12485

53. Weinberger V, Minar L. Diagnostics of malign ovarian tumors by ultrosound and CA 125 – our experience. *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):498.

54. Knafel A, Nocun A, Banas T, Wiechec M, Jach R, Pietrus M, *et al.* IOTA simple ultrasound-based rules: why do we have inconclusive results? *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):155–6.

55. Silvestre L, Martins WP, Candido-Dos-Reis FJ. Limitations of three-dimensional power Doppler angiography in preoperative evaluation of ovarian tumors. *J Ovarian Res* 2015;**8**:47. https://doi.org/10.1186/s13048-015-0174-y

56. Kaijser J, Van Gorp T, Van Holsbeke C, Sayasneh A, Vergote I, Bourne T, *et al.* IOTA simple descriptors (SD) or simple rules (SR) as a triage test in patients with ovarian tumours: subsequent value of CA125, HE4 or ROMA in clinical reality? *BJOG* 2013;**120**:371.

57. Abdalla N, Bachanek M, Kowalska J, Cendrowski K, Sawicki W. Role of HE4 and simple ultrasound rules proposed by iota group in preoperative evaluation of adnexal masses: a prospective study. *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):835.

58. Piovano E, Cavallero C, Fuso L, Viora E, Ferrero A, Gregori G, *et al.* Diagnostic accuracy and cost-effectiveness of different strategies to triage adnexal masses: a prospective study. *Ultrasound Obstet Gynecol* 2016;**5**:5.

59. Tongsong T, Tinnangwattana D, Vichak-Ururote L, Tontivuthikul P, Charoenratana C, Lerthiranwong T. Comparison of effectiveness in differentiating benign from malignant ovarian masses between IOTA simple rules and subjective sonographic assessment. *Asian Pac J Cancer Prev* 2016;**17**:4377–80.

60. Murala KS, Ma K, Henry RJW, Snape S, Dwivedi R. Performance of IOTA simple rules and RMI in preoperative classification of adnexal lesions in DGH setting. *BJOG* 2014;**121**(Suppl. 2):49–50.

61. Di Legge A, Testa AC, Ameye L, Van Calster B, Lissoni AA, Leone FP, *et al.* Lesion size affects diagnostic performance of IOTA logistic regression models, IOTA simple rules and risk of malignancy index in discriminating between benign and malignant adnexal masses. *Ultrasound Obstet Gynecol* 2012;**40**:345–54. https://doi.org/10.1002/uog.11167

62. Sayasneh A, Wynants L, Preisler J, Kaijser J, Johnson S, Stalder C, *et al.* Multicentre external validation of IOTA prediction models and RMI by operators with varied training. *Br J Cancer* 2013;**108**:2448–54. https://doi.org/10.1038/bjc.2013.224

63. Fathallah K, Huchon C, Bats AS, Metzger U, Lefrère-Belda MA, Bensaid C, Lécuru F. External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors. *Gynecol Obstet Fertil* 2011;**39**:477–81. https://doi.org/10.1016/j.gyobfe.2011.05.007

64. Ruiz de Gauna B, Rodriguez D, Olartecoechea B, Aubá M, Jurado M, Gómez Roig MD, Alcázar JL. Diagnostic performance of IOTA simple rules for adnexal masses classification: a comparison between two centers with different ovarian cancer prevalence. *Eur J Obstet Gynecol Reprod Biol* 2015;**191**:10–14. https://doi.org/10.1016/j.ejogrb.2015.05.024

65. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, *et al.* Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010;**341**:c6839. https://doi.org/10.1136/bmj.c6839

66. Baker C, Pasipanodya T, Dwivedi R. The management of suspected ovarian masses in premenopausal women in a DGH setting. *BJOG* 2013;**120**:372.

67. Ameye L, Timmerman D, Valentin L, Paladini D, Zhang J, Van Holsbeke C, *et al.* Clinically oriented three-step strategy for assessment of adnexal pathology. *Ultrasound Obstet Gynecol* 2012;**40**:582–91. https://doi.org/10.1002/uog.11177

68. Zhang Z, Pappas T, Sokoll LJ, Smith A, Chen L, Chan DW, *et al.* Derivation of a second generation multivariate index assay to improve specificity in presurgical evaluation of adnexal masses for risk of ovarian malignancy. *J Clin Oncol* 2015;**33**:5561.

69. Wolf J, Pappas T, Zhang Z, Smith A, Hudson T, Carpenter B, *et al.* Validation of a second-generation MIA (MIA2G) for triage of adnexal masses. *J Clin Oncol* 2015;**33**:A5556.

70. Coleman RL, Herzog TJ, Chan DW, Munroe DG, Pappas TC, Smith A, *et al.* Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses. *Am J Obstet Gynecol* 2016;**215**:82.e1–11. https://doi.org/10.1016/j.ajog.2016.03.003

71. Aktürk E, Karaca RE, Alanbay I, Dede M, Karaşahin E, Yenen MC, Başer I. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. *J Gynecol Oncol* 2011;**22**:177–82. https://doi.org/10.3802/jgo.2011.22.3.177

72. Yamamoto Y, Yamada R, Oguri H, Maeda N, Fukaya T. Comparison of four malignancy risk indices in the preoperative evaluation of patients with pelvic masses. *Eur J Obstet Gynecol Reprod Biol* 2009;**144**:163–7. https://doi.org/10.1016/j.ejogrb.2009.02.048

73. Lou HY, Meng H, Zhu QL, Zhang Q, Jiang YX. [Application values of four risk of malignancy indices in the preoperative evaluation of patients with adnexal masses.] *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 2010;**32**:297–302. https://doi.org/10.3881/j.issn.1000-503X.2010.03.013

74. Ulusoy S, Akbayir O, Numanoglu C, Ulusoy N, Odabas E, Gulkilik A. The risk of malignancy index in discrimination of adnexal masses. *Int J Gynaecol Obstet* 2007;**96**:186–91. https://doi.org/10.1016/j.ijgo.2006.10.006

75. Manjunath AP, Pratapkumar, Sujatha K, Vani R. Comparison of three risk of malignancy indices in evaluation of pelvic masses. *Gynecol Oncol* 2001;**81**:225–9. https://doi.org/10.1006/gyno.2001.6122

76. Tingulstad S, Hagen B, Skjeldestad FE, Onsrud M, Kiserud T, Halvorsen T, Nustad K. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. *Br J Obstet Gynaecol* 1996;**103**:826–31. https://doi.org/10.1111/j.1471-0528.1996.tb09882.x

77. Asif N, Sattar A, Dawood MM, Rafi T, Aamir M, Anwar M. Pre-operative evaluation of ovarian mass: risk of malignancy index. *J Coll Physicians Surg Pak* 2004;**14**:128–31. https://doi.org/03.2004/JCPSP.128131

78. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol* 1990;**97**:922–9. https://doi.org/10.1111/j.1471-0528.1990.tb02448.x

79. Davies AP, Jacobs I, Woolas R, Fish A, Oram D. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. *Br J Obstet Gynaecol* 1993;**100**:927–31. https://doi.org/10.1111/j.1471-0528.1993.tb15109.x

80. Morgante G, la Marca A, Ditto A, De Leo V. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. *Br J Obstet Gynaecol* 1999;**106**:524–7. https://doi.org/10.1111/j.1471-0528.1999.tb08318.x

81. Presl J, Kučera R, Topolčan O, Novotný Z, Vrzalová J, Fuchsova R, *et al.* [HE4 a biomarker of ovarian cancer.] *Ceska Gynekol* 2012;**77**:445–9

82. Chan KK, Chen CA, Nam JH, Ochiai K, Wilailak S, Choon AT, *et al.* The use of HE4 in the prediction of ovarian cancer in Asian women with a pelvic mass. *Gynecol Oncol* 2013;**128**:239–44. https://doi.org/10.1016/j.ygyno.2012.09.034

83. Karlsen MA, Sandhu N, Høgdall C, Christensen IJ, Nedergaard L, Lundvall L, *et al.* Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2012;**127**:379–83. https://doi.org/10.1016/j.ygyno.2012.07.106

84. Piovano E, Macchi C, Cavallero C, Fuso L, De Ruvo D, Viora E, *et al.* Preoperative management of the adnexal mass: a prospective multi-center study based on 391 patients. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):487.

85. Van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, *et al.* HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. *Br J Cancer* 2011;**104**:863–70. https://doi.org/10.1038/sj.bjc.6606092.

86. Novotny Z, Presl J, Kucera R, Topolcan O, Vrzalova J, Fuchsova R, *et al.* HE4 and ROMA index in Czech postmenopausal women. *Anticancer Res* 2012;**32**:4137–40.

87. Moore R, Miller C, DiSilvestro P, Landrum L, Gajewski W, Renneisen P, *et al.* Evaluation of the risk of ovarian malignancy algorithm in women with a pelvic mass presenting to general gynecologists. *Gynecol Oncol* 2011;**120**:S68. https://doi.org/10.1016/j.ygyno.2010.12.163

88. Moore R, Miller M, MacLaughlan S, Vucetic Z, Gajewski W, DiSilvestro P. The use of the Risk of Ovarian Malignancy Algorithm (ROMA) with clinical assessment improves ovarian cancer detection in women with a pelvic mass. *Gynecol Oncol* 2012;**125**(Suppl. 1):S38. https://doi.org/10.1016/j.ygyno.2011.12.087

89. Yanaranop M, Anakrat V, Siricharoenthai S, Nakrangsee S, Thinkhamrop B. Is the Risk of Ovarian Malignancy Algorithm better than other tests for predicting ovarian malignancy in women with pelvic masses? *Gynecol Obstet Invest* 2016;**21**:21.

90. Clemente J, Benitez G. A retrospective cohort study to validate CA125 and the combination of CA125 and HE4 using the ROMA in assessing the risk for ovarian malignancy in women diagnosed with an adnexal mass in Makati Medical Centre. *BJOG* 2015;**122**(Suppl. 1):133–4.

91. Kaijser J, Van Gorp T, Van Hoorde K, Van Holsbeke C, Sayasneh A, Vergote I, *et al.* A comparison between an ultrasound based prediction model (LR2) and the risk of ovarian malignancy algorithm (ROMA) to assess the risk of malignancy in women with an adnexal mass. *Gynecol Oncol* 2013;**129**:377–83. https://doi.org/10.1016/j.ygyno.2013.01.018

92. Kaijser J, Van Gorp T, Van Holsbeke C, Sayasneh A, Vergote I, Bourne T, *et al.* Diagnostic test performance of CA125, HE4, ROMA and IOTA's LR2 in adnexal tumours of different size. *BJOG* 2013;**120**:360–1.

93. Chen X, Zhou H, Chen R, He J, Wang Y, Huang L, *et al.* Development of a multimarker assay for differential diagnosis of benign and malignant pelvic masses. *Clin Chim Acta* 2015;**440**:57–63. https://doi.org/10.1016/j.cca.2014.11.013

94. Langhe R, Norris L, Petzold M, Abu Saadeh F, Kamran W, Ibrahim N, *et al.* ROMA index improves specificity of ovarian cancer diagnosis compared to CA125 alone. *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):588.

95. Xu Y, Zhong R, He J, Ding R, Lin H, Deng Y, *et al.* Modification of cut-off values for HE4, CA125 and the ROMA algorithm for early-stage epithelial ovarian cancer detection: results from 1021 cases in South China. *Clin Biochem* 2016;**49**:32–40. https://doi.org/10.1016/j.clinbiochem.2015.07.029

96. Li L, Wan J, Cai G, Yuan L, Liang J, Song J, *et al.* Value of serum human epididymis secretory protein 4 as a marker for differential diagnosis of malignant and benign gynecological diseases of patients in southern China. *Clin Chim Acta* 2016;**459**:170–6. https://doi.org/10.1016/j.cca.2016.06.010

97. Janas L, Głowacka E, Wilczyński JR, Malinowski A, Nowak M. [Evaluation of applicability of HE4 and ROMA in the preoperative diagnosis of adnexal masses.] *Ginekol Pol* 2015;**86**:193–7. https://doi.org/10.17772/gp/2062

98. Van Gorp T, Veldman J, Van Calster B, Cadron I, Leunen K, Amant F, *et al.* Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses. *Eur J Cancer* 2012;**48**:1649–56. https://doi.org/10.1016/j.ejca.2011.12.003

99. Winarto H, Laihad BJ, Nuranna L. Modification of cutoff values for HE4, CA125, the Risk of Malignancy Index, and the Risk of Malignancy Algorithm for ovarian cancer detection in Jakarta, Indonesia. *Asian Pac J Cancer Prev* 2014;**15**:1949–53. https://doi.org/10.7314/APJCP.2014.15.5.1949

100. Kaijser J, Van Gorp T, Smet ME, Van Holsbeke C, Sayasneh A, Epstein E, *et al.* Are serum HE4 or ROMA scores useful to experienced examiners for improving characterization of adnexal masses after transvaginal ultrasonography? *Ultrasound Obstet Gynecol* 2014;**43**:89–97. https://doi.org/10.1002/uog.12551

101. Moore RG, Miller MC, Disilvestro P, Landrum LM, Gajewski W, Ball JJ, *et al.* Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstet Gynecol* 2011;**118**:280–8. https://doi.org/10.1097/AOG.0b013e318224fce2

102. Zhang P, Wang C, Cheng L, Zhang P, Guo L, Liu W, *et al.* Comparison of HE4, CA125, and ROMA diagnostic accuracy: a prospective and multicenter study for Chinese women with epithelial ovarian cancer. *Medicine* 2015;**94**:e2402. https://doi.org/10.1097/MD.0000000000002402

103. Al Musalhi K, Al Kindi M, Al Aisary F, Ramadhan F, Al Rawahi T, Al Hatali K, Mula-Abed WA. Evaluation of HE4, CA-125, Risk of Ovarian Malignancy Algorithm (ROMA) and Risk of Malignancy Index (RMI) in the preoperative assessment of patients with adnexal mass. *Oman Med J* 2016;**31**:336–44. https://doi.org/10.5001/omj.2016.68

104. Shulman LP, Smith A, Pappas T, Bonato V, Munroe D, Wolf J. *Clinical Performance Comparison of Two IVDMIAs for Pre-Surgical Assessment of Ovarian Cancer Risk*. Annual Meeting of the American College of Obstetricians and Gynecologists, Washington, DC, USA, 14–17 May 2016. https://doi.org/10.1097/01.AOG.0000483382.09130.2b

105. Gizzo S, Berretta R, Noventa M, Nardelli GB, Silvio Patrelli T. Borderline ovarian tumors and diagnostic dilemma of intraoperative diagnosis: could preoperative HE4 assay and ROMA score assessment increase the frozen section accuracy? A multicenter case-control study. *Int J Gynecol Cancer* 2014;**24**:21. https://doi.org/10.1155/2014/803598

106. Ali M, Jabeen R, Khurshid M. Preoperative diagnosis of malignancy in ovarian masses: a comparison of risk of malignancy indices I and II. *Med Forum Mon* 2010;**21**:19–23.

107. Abudia S. Evaluation of the risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic mass. *Jamahiriya Med J* 2010;**10**:286–9.

108. Shahzad N, Rashid N, Zahra S, Malik A. Role of malignancy index in prediction of malignancy in ovarian masses preoperative. *Med Forum Mon* 2015;**26**:44–7.

109. Rossi A, Forzano L, Romanello I, Ambrosini G, Iuri V, Marchesoni D. Comparison of pelvic masses score (PMS) and Risk of Malignancy Index (RMI 3) in the evaluation of pelvic masses. *Eur J Gynaecol Oncol* 2014;**35**:421–4.

110. Kucera R, Topolcan O, Presl J, Novotny Z, Svobodova S, Vrzalova J, *et al.* Clinical benefit of ROMA index in the diagnosis of ovarian cancer. *Klin Biochem Metabol* 2012;**20**:248–51.

111. Woman's Health University Hospital, Egypt. *IOTA Versus Pattern Recognition Method in Diagnosis of Ovarian Masses*. ClinicalTrials.gov. 2000. URL: https://ClinicalTrials.gov/show/NCT02800031 (accessed 9 March 2017).

112. Sundar S, Rick C, Dowling F, Au P, Snell K, Rai N, *et al.* Refining ovarian cancer test accuracy scores (ROCkeTS) trial update. *BJOG* 2016;**123**(Suppl. 2):102.

113. Sundar S, Rick C, Dowling F, Au P, Snell K, Rai N, *et al.* Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. *BMJ Open* 2016;**6**:e010333. https://doi.org/10.1136/bmjopen-2015-010333

114. Meys EM, Rutten IJ, Kruitwagen RF, Slangen BF, Bergmans MG, Mertens HJ, *et al.* Investigating the performance and cost-effectiveness of the simple ultrasound-based rules compared to the risk of malignancy index in the diagnosis of ovarian cancer (SUBSONiC-study): protocol of a prospective multicenter cohort study in the Netherlands. *BMC Cancer* 2015;**15**:482. https://doi.org/10.1186/s12885-015-1319-5

115. Ikiz N, Guvenal T, Taneli F, Koyuncu FM, Kandiloglu AR, Bilge S, *et al.* Comparison of ROMA (Risk of Ovarian Malignancy Algorithm), RMI (Risk of Malignancy Index) and OTI (Ovarian Tumor Index) in patients with adnexal mass. *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):905.

116. Keogh F, Tan AL, Eva LJ. HE4 as a tumour marker for the prediction of ovarian carcinoma. *BJOG* 2015;**122**:137–8.

117. Romagnolo C, Leon AE, Fabricio ASC, Del Pup L, Papadakis C, Odicino FE, *et al.* HE4, CA125 and risk of ovarian malignancy algorithm (Roma) as diagnostic tools of ovarian cancer in patients with pelvic mass: an Italian multicenter prospective study. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):528–9.

118. Martin Rodriguez S, Ascorbe Salcedo P, Jareno Blanco MS. Diagnostic accuracy of HE4, CA125 and Roma for women with suspected ovarian cancer. *Clin Chem Lab Med* 2015;**53**:S424.

119. Tingulstad S, Hagen B, Skjeldestad FE, Halvorsen T, Nustad K, Onsrud M. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. *Obstet Gynecol* 1999;**93**:448–52. https://doi.org/10.1097/00006250-199903000-00028

120. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd edn. Oxford: Oxford University Press; 2005.

121. Ding Y, Hay J. Cost-effectiveness analysis of multimodal screening for ovarian cancer. *Value Health* 2010;**13**:A37. https://doi.org/10.1016/S1098-3015(10)72162-8

122. Forde GK, Hornberger J, Michalopoulos S, Bristow RE. Cost-effectiveness analysis of a multivariate index assay compared to modified American College of Obstetricians and Gynecologists criteria and CA-125 in the triage of women with adnexal masses. *Curr Med Res Opin* 2016;**32**:321–9. https://doi.org/10.1185/03007995.2015.1123679

123. Havrilesky LJ, Dinan M, Sfakianos GP, Curtis LH, Barnett JC, Van Gorp T, Myers ER. Costs, effectiveness, and workload impact of management strategies for women with an adnexal mass. *J Natl Cancer Inst* 2015;**107**:322. https://doi.org/10.1093/jnci/dju322

124. Drescher CW, Hawley S, Thorpe JD, Marticke S, McIntosh M, Gambhir SS, Urban N. Impact of screening test performance and cost on mortality reduction and cost-effectiveness of multimodal ovarian cancer screening. *Cancer Prev Res* 2012;**5**:1015–24. https://doi.org/10.1158/1940-6207.CAPR-11-0468

125. Kearns B, Chilcott J, Whyte S, Preston L, Sadler S. Cost-effectiveness of screening for ovarian cancer amongst postmenopausal women: a model-based economic evaluation. *BMC Med* 2016;**14**:200. https://doi.org/10.1186/s12916-016-0743-y

126. Cancer Research UK. *Ovarian Cancer Incidence Statistics*. London: Cancer Research UK; 2017. URL: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One (accessed 13 April 2017).

127. Colombo N, Peiretti M, Castiglione M, ESMO Guidelines Working Group. Non-epithelial ovarian cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol* 2009;**20**(Suppl. 4):24–6. https://doi.org/10.1093/annonc/mdp118

128. Bell R, Petticrew M, Luengo S, Sheldon T. Screening for ovarian cancer: a systematic review. *Health Technol Assess* 1998;**2**(2).

129. National Cancer Registration and Analysis Service. *Colorectal Cancer Survival by Stage – NCIN Data Briefing*. London: National Cancer Intelligence Network; 2010. URL: www.ncin.org.uk/publications/data_briefings/colorectal_cancer_survival_by_stage (accessed 13 April 2017).

130. Venesmaa P, Ylikorkala O. Morbidity and mortality associated with primary and repeat operations for ovarian cancer. *Obstet Gynecol* 1992;**79**:168–72. https://doi.org/10.1016/0020-7292(92)90941-B

131. Gerestein CG, Damhuis RA, Burger CW, Kooi GS. Postoperative mortality after primary cytoreductive surgery for advanced stage epithelial ovarian cancer: a systematic review. *Gynecol Oncol* 2009;**114**:523–7. https://doi.org/10.1016/j.ygyno.2009.03.011

132. Loft A, Andersen TF, Brønnum-Hansen H, Roepstorff C, Madsen M. Early postoperative mortality following hysterectomy. A Danish population based study, 1977–1981. *Br J Obstet Gynaecol* 1991;**98**:147–54. https://doi.org/10.1111/j.1471-0528.1991.tb13360.x

133. Collinson F, Qian W, Fossati R, Lissoni A, Williams C, Parmar M, *et al.* Optimal treatment of early-stage ovarian cancer. *Ann Oncol* 2014;**25**:1165–71. https://doi.org/10.1093/annonc/mdu116

134. International, Collaborative Ovarian Neoplasm Group. Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. *Lancet* 2002;**360**:505–15. https://doi.org/10.1016/S0140-6736(02)09738-6

135. Woo YL, Kyrgiou M, Bryant A, Everett T, Dickinson HO. Centralisation of services for gynaecological cancer. *Cochrane Database Syst Rev* 2012;**3**:CD007945. https://doi.org/10.1002/14651858.CD007945.pub2

136. Westwood M, Corro Ramos I, Lang S, Luyendijk M, Zaim R, Stirk L, *et al. Faecal Immunochemical Tests to Triage Patients with Lower Abdominal Symptoms for Suspected Colorectal Cancer Referrals in Primary Care: A Systematic Review and Cost-Effectiveness Analysis. A diagnostic assessment report*. York: Kleijnen Systematic Reviews; 2016.

137. Tappenden P, Chilcott J, Eggington S, Patnick J, Sakai H, Karnon J. Option appraisal of population-based colorectal cancer screening programmes in England. *Gut* 2007;**56**:677–84. https://doi.org/10.1136/gut.2006.095109

138. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**:509–18. https://doi.org/10.1111/j.1524-4733.2010.00700.x

139. Havrilesky LJ, Broadwater G, Davis DM, Nolte KC, Barnett JC, Myers ER, Kulasingam S. Determination of quality of life-related utilities for health states relevant to ovarian cancer diagnosis and treatment. *Gynecol Oncol* 2009;**113**:216–20. https://doi.org/10.1016/j.ygyno.2008.12.026

140. Grann VR, Panageas KS, Whang W, Antman KH, Neugut AI. Decision analysis of prophylactic mastectomy and oophorectomy in BRCA1-positive or BRCA2-positive patients. *J Clin Oncol* 1998;**16**:979–85. https://doi.org/10.1200/JCO.1998.16.3.979

141. Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. *Am J Gastroenterol* 1999;**94**:1650–7. https://doi.org/10.1111/j.1572-0241.1999.01157.x

142. Curtis L. *Unit Costs of Health and Social Care 2014*. Canterbury: Personal Social Services Research Unit, University of Kent; 2014.

143. Department of Health and Social Care (DHSC). *NHS Reference Costs 2015–2016*. London: DHSC; 2016. URL: www.gov.uk/government/publications/nhs-reference (accessed 29 November 2016).

144. Curtis L, Burns A. *Unit Costs of Health and Social Care 2016*. Canterbury: Personal Social Services Research Unit, University of Kent; 2016. URL: www.pssru.ac.uk/project-pages/unit-costs/2016/index.php (accessed 29 November 2016).

145. Joint Formulary Committee. *British National Formulary*. London: BMJ Group and Pharmaceutical Press; 2016. URL: https://bnf.nice.org.uk/ (accessed 29 November 2016).

146. Kumpulainen S, Sankila R, Leminen A, Kuoppala T, Komulainen M, Puistola U, *et al.* The effect of hospital operative volume, residual tumor and first-line chemotherapy on survival of ovarian cancer – a prospective nation-wide study in Finland. *Gynecol Oncol* 2009;**115**:199–203. https://doi.org/10.1016/j.ygyno.2009.07.011

147. Hinde S, Epstein D, Cook A, Embleton A, Perren T, Sculpher M. The cost-effectiveness of bevacizumab in advanced ovarian cancer using evidence from the ICON7 trial. *Value Health* 2016;**19**:431–9. https://doi.org/10.1016/j.jval.2016.01.013

148. Oza AM, Cook AD, Pfisterer J, Embleton A, Ledermann JA, Pujade-Lauraine E, *et al.* Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. *Lancet Oncol* 2015;**16**:928–36. https://doi.org/10.1016/S1470-2045(15)00086-8

149. National Institute for Health and Care Excellence (NICE). *Ultra-Radical (Extensive) Surgery for Advanced Ovarian Cancer*. Interventional procedures guidance [IPG470]. London: NICE; 2013.

150. Meys EM, Kaijser J, Kruitwagen RF, Slangen BF, Van Calster B, Aertgeerts B, *et al.* Subjective assessment versus ultrasound models to diagnose ovarian cancer: a systematic review and meta-analysis. *Eur J Cancer* 2016;**58**:17–29. https://doi.org/10.1016/j.ejca.2016.01.007

151. Dayyani F, Uhlig S, Colson B, Simon K, Rolny V, Morgenstern D, Schlumbrecht M. Diagnostic performance of risk of ovarian malignancy algorithm against CA125 and HE4 in connection with ovarian cancer: a meta-analysis. *Int J Gynecol Cancer* 2016;**26**:1586–93. https://doi.org/10.1097/IGC.0000000000000804

152. Li F, Tie R, Chang K, Wang F, Deng S, Lu W, *et al.* Does risk for ovarian malignancy algorithm excel human epididymis protein 4 and CA125 in predicting epithelial ovarian cancer: a meta-analysis. *BMC Cancer* 2012;**12**:258. https://doi.org/10.1186/1471-2407-12-258

153. Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M, Jurkovic D. Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis. *Ultrasound Obstet Gynecol* 2014;**44**:503–14. https://doi.org/10.1002/uog.13437

154. Rai N, Champaneria R, Snell K, Mallett S, Bayliss SE, Neal RD, *et al.* Symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer. *Cochrane Database Syst Rev* 2015;**12**:CD011964. https://doi.org/10.1002/14651858.cd011964

155. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;**58**:882–93. https://doi.org/10.1016/j.jclinepi.2005.01.016

156. Nunes N, Ambler G, Foo X, Naftalin J, Derdelis G, Widschwendter M, Jurkovic D. Comparison of two protocols for the management of asymptomatic postmenopausal women with adnexal tumours – a randomised controlled trial of RMI/RCOG vs Simple Rules. *Br J Cancer* 2017;**116**:584–91. https://doi.org/10.1038/bjc.2017.17

157. Centre for Reviews and Dissemination (CRD). *Search Strategies: NHS EED MEDLINE Using OvidSP (Economics Filter)*. York: CRD; 2014. URL: www.crd.york.ac.uk/crdweb/searchstrategies.asp (accessed 2 June 2014).

158. Centre for Reviews and Dissemination (CRD). *Search Strategies: NHS EED EMBASE Using OvidSP (Economics Filter)*. York: CRD; 2014. URL: www.crd.york.ac.uk/crdweb/searchstrategies.asp (accessed 4 April 2016).

Appendix 1 Literature search strategies

Clinical effectiveness searches

MEDLINE (via Ovid)

Date range searched: 1946 to week 2 November 2016.

Date searched: 24 November 2016.

Records found: 644.

Search strategy

1. exp Ovarian Neoplasms/ (79,417)
2. Fallopian Tube Neoplasms/ (2722)
3. Uterine Neoplasms/ (40,421)
4. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (6088)
5. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (104,209)
6. or/1-5 (151,393)
7. Peritoneal Neoplasms/ (13,589)
8. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (337,303)
9. or/7-8 (348,068)
10. ovar$.ti,ab,ot. (225,015)
11. 9 and 10 (23,572)
12. 6 or 11 (155,473)
13. ((risk adj4 malignan$ adj4 index) or (risk adj4 malignan$ adj4 indice$) or RMI).ti,ab,ot. (815)
14. (menopau$ or perimenopaus$ or premenopaus$ or postmenopaus$ or POF or climacteric or (change adj2 life)).ti,ab,ot. (91,426)
15. exp Menopause/ (55,959)
16. 14 or 15 (102,167)
17. (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. (315,798)
18. Ultrasonography/ (67,496)
19. 17 or 18 (333802)
20. (CA125$ or CA 125$ or ca 12-5$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (7731)
21. CA-125 Antigen/ (4333)
22. 20 or 21 (8485)
23. 16 and 19 and 22 (316)
24. 13 or 23 (1059)
25. (ROMA or (Ovar$ adj5 Algor$)).ti,ab,ot. (1670)
26. (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (469)

27. 16 and 22 and 26 (91)
28. 25 or 27 (1704)
29. (IOTA or international ovarian tumo?r analysis).ti,ab,ot. (1360)
30. ((Simple adj3 rules) or (simple adj3 descriptors) or SRrisk or b-rules or m-rules).ti,ab,ot. (1493)
31. 19 or 29 (335,099)
32. 30 and 31 (38)
33. (adnex$ adj8 (model$ or score$ or assess$)).ti,ab,ot. (287)
34. (ova2 or overa).ti,ab,ot. (25)
35. Follicle Stimulating Hormone/ (35,545)
36. (Follicle stimulat$ hormone$ or FSH or follitropin or fertiline fertinom p or follicotropin folliculostimulating hormone$ or follitrophin or follitropin$ or folltropin$ or 9002-68-0).ti,ab,ot,rn. (49,152)
37. 35 or 36 (49,152)
38. Apolipoprotein A-I/ (8782)
39. (apolipoprotein A1 or apo a1 or apo hdl 3 or apo hdl iii or apo high density lipoprotein 3 or apolipoprotein a 1 or apolipoprotein a i or apoprotein a1 or apoprotein ai or apoprotein a 1 or apoprotein a i).ti,ab,ot. (8019)
40. 38 or 39 (12379)
41. Transferrin/ (17,226)
42. (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1).ti,ab,ot,rn. (34,367)
43. 41 or 42 (34,367)
44. 22 and 26 and 37 and 40 and 43 (0)
45. 34 or 44 (25)
46. 24 or 28 or 32 or 33 or 45 (3019)
47. 12 and 46 (644)

MEDLINE In-Process & Other Non-Indexed Citations (via Ovid), MEDLINE Daily Update (via Ovid), MEDLINE Epub Ahead of Print (via Ovid)

MEDLINE In-Process & Other Non-Indexed Citations (via Ovid): searched to 22 November 2016.

MEDLINE Daily Update (via Ovid): searched to 22 November 2016.

MEDLINE Epub Ahead of Print (via Ovid): searched to 23 November 2016.

Date searched: 24 November 2016.

Records found: 83.

Search strategy

1. exp Ovarian Neoplasms/ (0)
2. Fallopian Tube Neoplasms/ (0)
3. Uterine Neoplasms/ (0)
4. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (904)
5. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (9696)

6. or/1-5 (9801)
7. Peritoneal Neoplasms/ (0)
8. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (25,858)
9. or/7-8 (25,858)
10. ovar$.ti,ab,ot. (18,312)
11. 9 and 10 (2204)
12. 6 or 11 (10,115)
13. ((risk adj4 malignan$ adj4 index) or (risk adj4 malignan$ adj4 indice$) or RMI).ti,ab,ot. (89)
14. (menopau$ or perimenopaus$ or premenopaus$ or postmenopaus$ or POF or climacteric or (change adj2 life)).ti,ab,ot. (7942)
15. exp Menopause/ (0)
16. 14 or 15 (7942)
17. (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. (41,009)
18. Ultrasonography/ (0)
19. 17 or 18 (41,009)
20. (CA125$ or CA 125$ or ca 12-5$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (849)
21. CA-125 Antigen/ (0)
22. 20 or 21 (849)
23. 16 and 19 and 22 (32)
24. 13 or 23 (110)
25. (ROMA or (Ovar$ adj5 Algor$)).ti,ab,ot. (144)
26. (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (185)
27. 16 and 22 and 26 (22)
28. 25 or 27 (149)
29. (IOTA or international ovarian tumo?r analysis).ti,ab,ot. (93)
30. ((Simple adj3 rules) or (simple adj3 descriptors) or SRrisk or b-rules or m-rules).ti,ab,ot. (347)
31. 19 or 29 (41,084)
32. 30 and 31 (11)
33. (adnex$ adj8 (model$ or score$ or assess$)).ti,ab,ot. (35)
34. (ova2 or overa).ti,ab,ot. (3)
35. Follicle Stimulating Hormone/ (0)
36. (Follicle stimulat$ hormone$ or FSH or follitropin or fertiline fertinom p or follicotropin folliculostimulating hormone$ or follitrophin or follitropin$ or folltropin$ or 9002-68-0).ti,ab,ot,rn. (2146)
37. 35 or 36 (2146)
38. Apolipoprotein A-I/ (0)
39. (apolipoprotein A1 or apo a1 or apo hdl 3 or apo hdl iii or apo high density lipoprotein 3 or apolipoprotein a 1 or apolipoprotein a i or apoprotein a1 or apoprotein ai or apoprotein a 1 or apoprotein a i).ti,ab,ot. (440)
40. 38 or 39 (440)
41. Transferrin/ (0)
42. (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1).ti,ab,ot,rn. (1465)
43. 41 or 42 (1465)
44. 22 and 26 and 37 and 40 and 43 (1)
45. 34 or 44 (4)
46. 24 or 28 or 32 or 33 or 45 (287)
47. 12 and 46 (83)

EMBASE (via Ovid)

Date range searched: 1974 to 23 November 2016.

Date searched: 24 November 2016.

Records found: 1185.

Search strategy

1. exp ovary cancer/ (97,370)
2. uterine tube tumor/ (1263)
3. uterine tube carcinoma/ (1899)
4. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (9245)
5. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (136,192)
6. peritoneum cancer/ (3891)
7. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (405888)
8. or/6-7 (408,204)
9. ovar$.ti,ab,ot. (278,995)
10. 8 and 9 (30,110)
11. 1 or 2 or 3 or 4 or 5 or 10 (168,296)
12. ((risk adj4 malignan$ adj4 index) or (risk adj4 malignan$ adj4 indice$) or RMI).ti,ab,ot. (1394)
13. risk of malignancy index/ (46)
14. 12 or 13 (1396)
15. (menopau$ or perimenopaus$ or premenopaus$ or postmenopaus$ or POF or climacteric or (change adj2 life)).ti,ab,ot. or menopause/ (136,468)
16. (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. or ultrasound/ or sonography/ (591,945)
17. (CA125$ or CA 125$ or ca 12-5$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (11,962)
18. CA 125 antigen/ (13,650)
19. 17 or 18 (16,956)
20. 15 and 16 and 19 (627)
21. 14 or 20 (1886)
22. ovarian malignancy algorithm/ (1)
23. (ROMA or (Ovar$ adj5 Algor$)).ti,ab,ot. (2502)
24. (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (956)
25. human epididymis protein 4/ (507)
26. or/24-25 (1036)
27. 15 and 19 and 26 (237)
28. 22 or 23 or 27 (2593)

29. (IOTA or international ovarian tumo?r analysis).ti,ab,ot. (846)
30. ((Simple adj3 rules) or (simple adj3 descriptors) or SRrisk or b-rules or m-rules).ti,ab,ot. (1796)
31. 16 or 29 (592,674)
32. 30 and 31 (66)
33. (adnex$ adj8 (model$ or score$ or assess$)).ti,ab,ot. (466)
34. (ova2 or overa).ti,ab,ot. (78)
35. follitropin/ (56,500)
36. (Follicle stimulat$ hormone$ or FSH or follitropin or fertiline fertinom p or follicotropin folliculostimulating hormone$ or follitrophin or follitropin$ or folltropin$ or 9002-68-0).ti,ab,ot,rn. (67,340)
37. or/35-36 (67,499)
38. apolipoprotein A1/ (16,294)
39. (apolipoprotein A1 or apo a1 or apo hdl 3 or apo hdl iii or apo high density lipoprotein 3 or apolipoprotein a 1 or apolipoprotein a i or apoprotein a1 or apoprotein ai or apoprotein a 1 or apoprotein a i).ti,ab,ot. (9160)
40. or/38-39 (18,737)
41. transferrin/ (27,791)
42. (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1).ti,ab,ot,rn. (42,260)
43. or/41-42 (42,344)
44. 19 and 26 and 37 and 40 and 43 (3)
45. 34 or 44 (81)
46. 21 or 28 or 32 or 33 or 45 (4880)
47. 11 and 46 (1185)

Cochrane Database of Systematic Reviews (via Wiley Online Library), Database of Abstracts of Reviews of Effects (via Wiley Online Library), Cochrane Central Register of Controlled Trials (via Wiley Online Library), Health Technology Assessment Database (via Wiley Online Library)

Cochrane Database of Systematic Reviews (via Wiley Online Library): Issue 11 of 12, November 2016.

Database of Abstracts of Reviews of Effects (via Wiley Online Library): Issue 2 of 4, April 2015.

Cochrane Central Register of Controlled Trials (via Wiley Online Library): Issue 10 of 12, October 2016.

Health Technology Assessment Database (via Wiley Online Library): Issue 4 of 4, October 2016.

Date searched: 24 November 2016.

Records found: 43.

- Cochrane Database of Systematic Reviews: 1.
- Database of Abstracts of Reviews of Effects: 5.
- Cochrane Central Register of Controlled Trials: 37.
- Health Technology Assessment: 0.

Search strategy

#1 MeSH descriptor: [Ovarian Neoplasms] explode all trees (1511)
#2 MeSH descriptor: [Fallopian Tube Neoplasms] this term only (45)
#3 MeSH descriptor: [Uterine Neoplasms] this term only (691)
#4 (AOSCa* or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA* or dysgerminom*):ti,ab,kw (231)

#5 ((ovar* or "high-grade serous" or "low-grade serous" or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) near/5 (cancer* or adenocarcin* or adeno-carcin* or tumo?r* or sarcoma* or neoplas* or metasta* or meta-sta* or carcino* or oncogenesis or malignan* or choriocarcinom* or teratoma* or cystadenocarcin* or rhabdomyosarcom* or rhabdo-myosarcom* or rhabdosarcom* or leiomyosarcoma* or leio-myosarcom* or androblastom* or arrhenoblastom* or adenoma* or lesion* or oncolo*)):ti,ab,kw (7371)
#6 #1 or #2 or #3 or #4 or #5 (7441)
#7 MeSH descriptor: [Peritoneal Neoplasms] this term only (213)
#8 (peritoneum or borderline or epithelial or primary peritoneal):ti,ab,kw (7882)
#9 #7 or #8 (8005)
#10 ovar*:ti,ab,kw (9974)
#11 #9 and #10 (1073)
#12 #6 or #11 (7490)
#13 ((risk near/4 malignan* near/4 index) or (risk near/4 malignan* near/4 indice*) or RMI):ti,ab,kw (53)
#14 (menopau* or perimenopaus* or premenopaus* or postmenopaus* or POF or climacteric or (change near/2 life)):ti,ab,kw (18,330)
#15 MeSH descriptor: [Menopause] explode all trees (6396)
#16 #14 or #15 (18,330)
#17 (ultraso* or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph* or doptone* or echograph* or echogram* or echosound*):ti,ab,kw (21,215)
#18 MeSH descriptor: [Ultrasonography] this term only (956)
#19 #17 or #18 (21,215)
#20 (CA125* or "CA 125*" or "CA 12-5*" or (antigen near/2 125) or (mucin near/1 16) or mucin16 or (muc near/1 16) or muc16):ti,ab,kw (473)
#21 MeSH descriptor: [CA-125 Antigen] this term only (157)
#22 #20 or #21 (473)
#23 #16 and #19 and #22 (21)
#24 #13 or #23 (73)
#25 (ROMA or (ovar* near/5 algor*)):ti,ab,kw (67)
#26 ("human epididymis protein 4" or "human epididymal protein 4" or "WAP four disulfide core domain protein 2" or "wap 4 disulfide core domain protein 2" or WFCD2 or EDDM4 or WAP5 or "wap four disulfide core domain 2" or "wap 4 disulfide core domain 2" or "HE 4" or HE4):ti,ab,kw (35)
#27 #16 and #22 and #26 (5)
#28 #25 or #27 (69)
#29 (IOTA or "international ovarian tumo?r analysis"):ti,ab,kw (22)
#30 ((simple near/3 rule*) or (simple near/3 descriptor*) or SRrisk or b-rule* or m-rule*):ti,ab,kw (44)
#31 #19 or #29 (21,234)
#32 #30 and #31 (2)
#33 (adnex* near/8 (model* or score* or assess*)):ti,ab,kw (17)
#34 (ova2 or overa):ti,ab,kw (4)
#35 MeSH descriptor: [Follicle Stimulating Hormone] this term only (1700)
#36 ("Follicle stimulat* hormone*" or FSH or follitropin or fertiline or "fertinom p" or follicotropin or "folliculostimulating hormone*" or follitrophin or follitropin* or folltropin* or 9002-68-0):ti,ab,kw (4010)
#37 #35 or #36 (4010)
#38 MeSH descriptor: [Apolipoprotein A-I] this term only (444)
#39 ("apolipoprotein A1" or "apo a1" or "apo hdl 3" or "apo hdl iii" or "apo high density lipoprotein 3" or "apolipoprotein a 1" or "apolipoprotein a I" or "apoprotein a1" or "apoprotein ai" or "apoprotein a 1" or "apoprotein a I"):ti,ab,kw (1334)
#40 #38 or #39 (1334)
#41 MeSH descriptor: [Transferrin] this term only (343)

#42 (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1):ti,ab,kw (1467)
#43 #41 or #42 (1467)
#44 #22 and #26 and #37 and #40 and #43 (0)
#45 #34 or #44 (4)
#46 #24 or #28 or #32 or #33 or #45 (155)
#47 #12 and #46 (44)

International Network of Agencies for Health Technology Assessment Publications (via the internet: www.inahta.org/publications/)

Date searched: 25 November 2016.

Records found: none.

Search strategy

(ovar* OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum) AND (RMI OR "risk of malignancy" OR ROMA OR "Ovarian Malignancy Algorithm" OR IOTA OR "International ovarian tumor analysis" OR "simple ultrasound rule" OR "simple rule" OR SRrisk OR b-rule OR m-rule OR Adnex OR OVA2 OR Overa OR HE4 OR "HE 4" OR epididy* OR "WAP 4" OR WAP4 OR "WAP four" OR WAP5 OR WFCD2 OR EDDM4 OR CA125 OR "CA 125" OR "CA 12-5" OR "antigen 125" OR "mucin 16" OR mucin16 OR "muc 16" OR muc16)

National Institute for Health Research Health Technology Assessment Journals Library (via the internet: www.journalslibrary.nihr.ac.uk/#/)

Date searched: 25 November 2016.

Records found: 23.

Search terms	Journal reports, (*n*)	Research projects, (*n*)
ovarian	12	44
ovary	0	5
ovaries	0	6
fallopian	0	7
oviduct	0	0
Total	**12**	**62**
Total after removal of duplicates	**12**	**43**
Combined total	**55**	
Total after removal of irrelevant studies	**23**	

Aggressive Research Intelligence Facility database (via the internet: www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/index.aspx)

Date searched: 25 November 2016.

Records found: 25.

Advanced search.

All published libraries

Search terms	Results
Abstract: ovarian	18
AND	
Abstract: rmi	
Abstract: ovarian	2
AND	
Abstract: malignancy index	
Abstract: ovarian	0
AND	
Abstract: malignancy indices	
Abstract: ovarian	3
AND	
Abstract: ROMA	
Abstract: ovarian	0
AND	
Abstract: malignancy algorithm	
Abstract: ovarian	1
AND	
Abstract: IOTA	
Abstract: International ovarian tumor analysis	1
Abstract: ovarian	1
AND	
Abstract: simple rule	
Abstract: SRrisk	0
OR	
Abstract: b-rule	
OR	
Abstract: m-rule	
Abstract: ovarian	4
AND	

Search terms	Results
Abstract: adnex	
Abstract: OVA2	2
OR	
Abstract: HE4	
OR	
Abstract: human epididymis protein 4	
OR	
Abstract: human epididymal protein 4	
Abstract: WAP4	0
OR	
Abstract: WAP 4	
OR	
Abstract: WAP four	
OR	
Abstract: WAP5	
Abstract: CA125	4
OR	
Abstract: CA-125	
OR	
Abstract: CA 12-5	
OR	
Abstract: antigen 125	
Total	**36**
Total after removal of duplicates	**25**

PROSPERO (via the internet: www.crd.york.ac.uk/prospero/)

Date searched: 25 November 2016.

Records found: four.

Search strategy

#1 MeSH DESCRIPTOR Ovarian Neoplasms EXPLODE ALL TREES (38)
#2 MeSH DESCRIPTOR Fallopian Tube Neoplasms EXPLODE ALL TREES (0)
#3 MeSH DESCRIPTOR Uterine Neoplasms EXPLODE ALL TREES (80)

#4 (ovar* or "high-grade serous" or "low-grade serous" or "sertoli-leydig cell" or fallopian or oviduct or uterine or uterus or tubal) near (cancer* or adenocarcin* or adeno-carcin* or tumor* or tumour* or sarcoma* or neoplas* or metasta* or meta-sta* or carcino* or oncogenesis or malignan* or choriocarcinom* or teratoma* or cystadenocarcin* or rhabdomyosarcom* or rhabdo-myosarcom* or rhabdosarcom* or leiomyosarcoma* or leio-myosarcom* or androblastom* or arrhenoblastom* or adenoma* or lesion* or oncolo*) (99)
#5 (AOSCa* or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA* or dysgerminom*) (9)
#6 #5 OR #4 OR #3 OR #2 OR #1 (160)
#7 MeSH DESCRIPTOR Peritoneal Neoplasms EXPLODE ALL TREES (4)
#8 peritoneum or borderline or epithelial or "primary peritoneal" (130)
#9 #7 OR #8 (134)
#10 ovar* (224)
#11 #9 AND #10 (24)
#12 #9 AND #10 (24)
#13 #6 OR #11 (164)
#14 (risk near malignan* near index) or (risk near malignan* near indice*) or RMI (7)
#15 menopau* or perimenopaus* or premenopaus* or postmenopaus* or POF or climacteric or (change near life) (372)
#16 MeSH DESCRIPTOR Menopause EXPLODE ALL TREES (30)
#17 #16 OR #15 (373)
#18 MeSH DESCRIPTOR Ultrasonography EXPLODE ALL TREES (98)
#19 (ultraso* or phonophoresis or sonication or sonification or "ultra sound" or ultrashell or sonograph* or doptone* or echograph* or echogram* or echosound*) (625)
#20 #19 OR #18 (653)
#21 MeSH DESCRIPTOR CA-125 Antigen EXPLODE ALL TREES (2)
#22 (CA125* or "CA 125*" or ca 12-5* or (antigen near "125") or (mucin near "16") or mucin16 or (muc near "16") or muc16) (8)
#23 #22 OR #21 (9)
#24 #17 AND #20 AND #23 (4)
#25 #14 OR #24 (9)
#26 ROMA or (Ovar* near Algor*) (30)
#27 ("human epididymis protein 4" or "human epididymal protein 4" or "WAP four" or "wap 4" or wap4 or WFCD2 or EDDM4 or WAP5 or "HE 4" or HE4) (3)
#28 #17 AND #23 AND #27 (2)
#29 #26 OR #28 (32)
#30 IOTA or "international ovarian tumor analysis" or "international ovarian tumour analysis" (3)
#31 ((Simple near rules) or (simple near descriptors) or SRrisk or b-rules or m-rules) (3)
#32 #20 OR #30 (653)
#33 #32 AND #31 (2)
#34 (adnex* near (model* or score* or assess*)) (0)
#35 ova2 or overa (0)
#36 MeSH DESCRIPTOR Follicle Stimulating Hormone EXPLODE ALL TREES (2)
#37 ("follicle stimulat* hormone*" or FSH or follitropin or fertiline or "fertinom p" or follicotropin or "folliculostimulating hormone*" or follitrophin or follitropin* or folltropin*) (31)
#38 #37 OR #36 (32)
#39 MeSH DESCRIPTOR Apolipoprotein A-I EXPLODE ALL TREES (0)
#40 ("apolipoprotein A1" or "apo a1" or "apo hdl 3" or "apo hdl iii" or "apo high density lipoprotein 3" or "apolipoprotein a 1" or "apolipoprotein a i" or "apoprotein a1" or "apoprotein ai" or "apoprotein a 1" or "apoprotein a i") (8)
#41 #40 OR #39 (8)
#42 MeSH DESCRIPTOR Transferrin EXPLODE ALL TREES (0)
#43 (transferrin or siderophilin or transferrinemia or transferrinaemia or transferrins) (21)
#44 #42 OR #43 (21)

#45 #23 AND #27 AND #38 AND #41 AND #44 (0)
#46 #35 OR #45 (0)
#47 #25 OR #29 OR #33 OR #34 OR #46 (40)
#48 #13 AND #47 (4)

ClinicalTrials.gov (via the internet: http://clinicaltrials.gov/ct2/search/advanced)

Date searched: 24 November 2016.

Records found: 269.

Expert search option.

Search strategy

(ovarian OR ovary OR ovaries OR "high-grade serous" OR "low-grade serous" OR "sertoli-leydig cell" OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum) AND (cancer OR adenocarcinoma OR tumor OR tumour OR sarcoma OR neoplasm OR neoplasia OR metastatic OR metastasis OR metastases OR carcinoma OR oncogenesis OR malignancy OR malignancies OR choriocarcinoma OR teratoma OR cystadenocarcinoma OR rhabdomyosarcoma OR rhabdosarcoma OR leiomyosarcoma OR androblastoma OR arrhenoblastoma OR adenoma OR lesion OR oncology OR oncologic) AND (RMI OR "risk of malignancy index" OR "risk of malignancy indices" OR ROMA OR "Risk of Ovarian Malignancy Algorithm" OR IOTA OR "International ovarian tumor analysis" OR "simple ultrasound rule" OR "simple rule" OR SRrisk OR b-rule OR m-rule OR Adnex OR OVA2 OR Overa OR HE4 OR "HE 4" OR "human epididymis protein 4" OR "human epididymal protein 4" OR "WAP 4" OR WAP4 OR "WAP four" OR WAP5 OR WFCD2 OR EDDM4 OR CA125 OR "CA 125" OR "CA 12-5" OR "antigen 125" OR "mucin 16" OR mucin16 OR "muc 16" OR muc16)European Union Clinical Trials Register (via the internet: www.clinicaltrialsregister.eu/ctr-search/search)

Date searched: 25 November 2016.

Records found: 122.

Search strategy

(ovarian OR ovary OR ovaries OR "high-grade serous" OR "low-grade serous" OR "sertoli-leydig cell" OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum) AND (cancer OR adenocarcinoma OR tumor OR tumour OR sarcoma OR neoplasm OR neoplasia OR metastatic OR metastasis OR metastases OR carcinoma OR oncogenesis OR malignancy OR malignancies OR choriocarcinoma OR teratoma OR cystadenocarcinoma OR rhabdomyosarcoma OR rhabdosarcoma OR leiomyosarcoma OR androblastoma OR arrhenoblastoma OR adenoma OR lesion OR oncology OR oncologic) AND (RMI OR "risk of malignancy index" OR "risk of malignancy indices" OR ROMA OR "Risk of Ovarian Malignancy Algorithm" OR IOTA OR "International ovarian tumor analysis" OR "simple ultrasound rule" OR "simple rule" OR SRrisk OR b-rule OR m-rule OR Adnex OR OVA2 OR Overa OR HE4 OR "HE 4" OR "human epididymis protein 4" OR "human epididymal protein 4" OR "WAP 4" OR WAP4 OR "WAP four" OR WAP5 OR WFCD2 OR EDDM4 OR CA125 OR "CA 125" OR "CA 12-5" OR "antigen 125" OR "mucin 16" OR mucin16 OR "muc 16" OR muc16)

World Health Organization International Clinical Trials Register Portfolio (via the internet: http://apps.who.int/trialsearch/)

Date searched: 24 November 2016.

Records found: 51.

Advanced search option	Results
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(2 records for) **2** trials found
Intervention: risk of malignancy index	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(5 records for) **5** trials found
Title: risk of malignancy index	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(0 records for) **0** trials found
Intervention: Risk of Ovarian Malignancy Algorithm	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(1 records for) **1** trial found
Title: Risk of Ovarian Malignancy Algorithm	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(1 records for) **1** trial found
Intervention: IOTA OR International ovarian tumor analysis OR simple ultrasound rule OR simple rule OR SRrisk OR b-rule OR m-rule	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(6 records for) **6** trials found
Title: IOTA OR International ovarian tumor analysis OR simple ultrasound rule" OR simple rule OR SRrisk OR b-rule OR m-rule	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(4 records for) **4** trials found
Intervention: HE4 OR human epididymis protein 4 OR human epididymal protein 4	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(10 records for) **10** trials found
Title: HE4 OR human epididymis protein 4 OR human epididymal protein 4	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(23 records for) **20** trials found
Intervention: CA125 OR CA-125	
Condition: ovarian OR ovary OR ovaries OR high-grade serous OR low-grade serous OR sertoli-leydig cell OR fallopian OR oviduct OR uterine OR uterus OR tubal OR peritoneal OR peritoneum	(27 records for) **21** trials found
Title: CA125 OR CA-125	
Standard search	***Results***
RMI AND ovarian	(4 records for) **4** trials found
ROMA AND ovarian	(2 records for) **2** trials found
adnex AND ovarian	(2 records for) **2** trials found
Total	**78**
Total after removal of duplicates	**51**

N.B. It was not possible to search for terms such as 'RMI' and 'ROMA' using the advanced search option. For example, the search term 'RMI' retrieves records containing words such as indeter**rmi**nate, hyperther**rmi**c, Metfo**rmi**n, etc., whereas the search term 'ROMA' retrieves **Roma**nian, st**roma**l, fib**roma**, a**roma**tase, a**roma**therapy, etc. Instead, searches were conducted using the standard search interface for these terms.

Radiological Society of North America (via the Internet: www.rsna.org/Past_Meetings.aspx)

Plenary Sessions, Science Sessions.

Date searched: 2 February 2017.

Records found: 45.

Filters: Biomarkers/Quantitative Imaging; Obstetric/Gynecologic Radiology; Radiation Oncology; Genitourinary Radiology; Oncologic Imaging; Ultrasound.

2016

Text words	Hits
Ovar*	5
Serous	0/1 duplicate removed
Sertoli-leydig	0
Fallopian	1
Oviduct	0
Uterine	10
Uterus	0
tubal	1
Total	**17**

2015

Text words	Hits
Ovar*	4
Serous	1/2 duplicates removed
Sertoli-leydig	0
Fallopian	0
Oviduct	0
Uterine	7
Uterus	0
tubal	1
Total	**6**

2014

Filter: meeting program.

Text words	Hits
Ovar*	4
Serous	1/3 duplicates removed
Sertoli-leydig	0
Fallopian	1
Oviduct	0
Uterine	16/17 duplicates removed
Uterus	0
Tub*	0/2 duplicates removed
Total	**22**

American Society of Clinical Oncology annual conference (via the internet http://meetinglibrary.asco.org/abstracts)

Date searched: 2 February 2017.

Records found: 603.

2016

Text words	Hits
(ovar* OR fallopian OR uter* OR tubal) (diagnos* OR predict* OR sensitiv* OR specific* OR likel* OR accura*)	217

2015

Text words	Hits
(ovar* OR fallopian OR uter* OR tubal) (diagnos* OR predict* OR sensitiv* OR specific* OR likel* OR accura*)	210

2014

Text words	Hits
(ovar* OR fallopian OR uter* OR tubal) (diagnos* OR predict* OR sensitiv* OR specific* OR likel* OR accura*)	176

Society of Gynecologic Oncology (via the internet)

Date searched: 2 February 2017.

Records found: 108.

2016 (www.sgo.org/2016-annual-meeting-archives/)

Text words in title/abstract (abstracts scanned for ovarian or uterine or gynaecologic cancer)	Hits
Risk of malignancy	2
RMI	0
Ultrasound	5
Ultra sound	0
CA125	1
CA 125	22/25 duplicates removed
ROMA	0
HE4	0/2 duplicates removed
Human epididymis protein 4	0/1 duplicate removed
Iota	0/1 duplicate removed
International ovarian	0
Simple rules	0
SRrisk	0
Ova2	0
Overa	0
Adnex	0/1 duplicate removed
Total	**30**

2015 (www.gynecologiconcology-online.net/issue/S0090–8258(15)X0005–9)

Text words	Hits
"Risk of malignancy"	5
RMI	0
Ultrasound	14
"Ultra sound"	0
CA125	29/30 duplicates removed
"CA 125"	0/30 duplicates removed
ROMA	0/1 duplicates removed
HE4	1/3 duplicates removed
Human epididymis protein 4	0/2 duplicates removed
Iota	0
International ovarian	0
Simple rules	0
SRrisk	0
Ova2	0
Overa	0
Adnex	0
Total	**49**

2014 (www.sgo.org/wp-content/uploads/2014/07/YGYNO_133_S1_compressed.pdf)

Text words in title/abstract (abstracts scanned for ovarian or uterine or gynaecologic cancer)	Hits
Risk of malignancy	3
RMI	0/3 duplicates removed
Ultrasound	2/4 duplicates removed
Ultra sound	0
CA125	0
CA 125	22/25 duplicates removed
ROMA	0
HE4	2/4 duplicates removed
Human epididymis protein 4	0
Iota	0
International ovarian	0
Simple rules	0
SRrisk	0
Ova2	0
Overa	0
Adnex	0
Total	**29**

The National Cancer Research Institute (via the internet)

Date searched: 2 February 2017.

Records found: 132.

2016 (http://abstracts.ncri.org.uk/year_published/2016/)

Text words: abstracts scanned for ovarian/uterine/gynaecological cancer	Hits
Ovar*	25

2015 (http://abstracts.ncri.org.uk/year_published/2015/)

Text words: abstracts scanned for ovarian/uterine/gynaecological cancer	Hits
Ovar*	61

2014 (http://abstracts.ncri.org.uk/year_published/2014/)

Text words: abstracts scanned for ovarian/uterine/gynaecological cancer	Hits
Ovar*	46

European Society of Radiology

Date searched: 2 February 2017.

Records found: 25.

2016: Scientific sessions and clinical trials in radiology (www.myesr.org/congress/about-ecr/past-congresses/ecr-2016)

Text words	Hits
Ovar*	5

2015: scientific sessions and late-breaking clinical trials (www.myesr.org/congress/about-ecr/past-congresses/ecr-2015)

Text words	Hits
Ovar*	13

2014: scientific sessions (www.myesr.org/congress/about-ecr/past-congresses/ecr-2014)

Text words	Hits
Ovar*	7

Cost-effectiveness searches

MEDLINE (via Ovid)

Date range searched: 1946 to week 2 November 2016.

Date searched: 23 November 2016.

Records found: 370.

Search strategy

1. exp Ovarian Neoplasms/ (79,388)
2. Fallopian Tube Neoplasms/ (2721)
3. Uterine Neoplasms/ (40,416)
4. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (6079)
5. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (104,167)
6. or/1-5 (151,338)
7. Peritoneal Neoplasms/ (13,578)
8. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (337,128)
9. or/7-8 (347,883)
10. ovar$.ti,ab,ot. (224,941)
11. 9 and 10 (23,562)
12. 6 or 11 (155,417)

13. (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. (315,645)
14. Ultrasonography/ (67,487)
15. 13 or 14 (333,646)
16. (CA125$ or CA 125$ or ca 12-5$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (7726)
17. CA-125 Antigen/ (4329)
18. 16 or 17 (8480)
19. (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (467)
20. Biomarkers, Tumor/ (118,213)
21. (Tumo?r marker$ or biomarker$ or bio-marker$ or cancer marker$ or neoplasm marker$).ti,ab,ot. (159,924)
22. 20 or 21 (246,574)
23. 15 or 18 or 19 or 22 (580,167)
24. 12 and 23 (20,146)
25. economics/ (28,593)
26. exp "costs and cost analysis"/ (216,876)
27. economics, dental/ (1917)
28. exp "economics, hospital"/ (23,025)
29. economics, medical/ (9388)
30. economics, nursing/ (4000)
31. economics, pharmaceutical/ (2804)
32. (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. (541,112)
33. (expenditure$ not energy).ti,ab. (21,940)
34. (value adj1 money).ti,ab. (29)
35. budget$.ti,ab. (20,740)
36. or/25-35 (682,801)
37. ((energy or oxygen) adj cost).ti,ab. (3151)
38. (metabolic adj cost).ti,ab. (1034)
39. ((energy or oxygen) adj expenditure).ti,ab. (20,576)
40. or/37-39 (23,929)
41. 36 not 40 (677,656)
42. letter.pt. (943,994)
43. editorial.pt. (416,892)
44. historical article.pt. (507,294)
45. or/42-44 (1,844,260)
46. 41 not 45 (644,021)
47. 24 and 46 (370)

Economics terms based on costs filter.[157]

MEDLINE In-Process & Other Non-Indexed Citations (via Ovid), MEDLINE Daily Update (via Ovid) and MEDLINE Epub Ahead of Print (via Ovid)

MEDLINE In-Process & Other Non-Indexed Citations (via Ovid): to 22 November 2016.

MEDLINE Daily Update (via Ovid): to 22 November 2016.

MEDLINE Epub Ahead of Print (via Ovid): to 23 November 2016.

Date searched: 24 November 2016.

Records found: 31.

Search strategy

1. exp Ovarian Neoplasms/ (0)
2. Fallopian Tube Neoplasms/ (0)
3. Uterine Neoplasms/ (0)
4. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (904)
5. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (9696)
6. or/1-5 (9801)
7. Peritoneal Neoplasms/ (0)
8. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (25,858)
9. or/7-8 (25,858)
10. ovar$.ti,ab,ot. (18,312)
11. 9 and 10 (2204)
12. 6 or 11 (10,115)
13. (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. (41,009)
14. Ultrasonography/ (0)
15. 13 or 14 (41,009)
16. (CA125$ or CA 125$ or ca 12-5$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (849)
17. CA-125 Antigen/ (0)
18. 16 or 17 (849)
19. (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (185)
20. Biomarkers, Tumor/ (0)
21. (Tumo?r marker$ or biomarker$ or bio-marker$ or cancer marker$ or neoplasm marker$).ti,ab,ot. (32,099)
22. 20 or 21 (32,099)
23. 15 or 18 or 19 or 22 (73,106)
24. 12 and 23 (1440)
25. economics/ (0)
26. exp "costs and cost analysis"/ (1)
27. economics, dental/ (0)
28. exp "economics, hospital"/ (0)
29. economics, medical/ (0)
30. economics, nursing/ (0)
31. economics, pharmaceutical/ (0)
32. (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. (91,939)
33. (expenditure$ not energy).ti,ab. (2877)
34. (value adj1 money).ti,ab. (5)
35. budget$.ti,ab. (3723)
36. or/25-35 (95,781)
37. ((energy or oxygen) adj cost).ti,ab. (467)
38. (metabolic adj cost).ti,ab. (159)

39. ((energy or oxygen) adj expenditure).ti,ab. (2166)
40. or/37-39 (2715)
41. 36 not 40 (95,025)
42. letter.pt. (38,204)
43. editorial.pt. (27,650)
44. historical article.pt. (0)
45. or/42-44 (65,854)
46. 41 not 45 (94,317)
47. 24 and 46 (31)

Economics terms based on costs filter.[157]

EMBASE (via Ovid)

Date range searched: 1974 to 23 November 2016.

Date searched: 24 November 2016.

Records found: 665.

Search strategy

1. exp ovary cancer/ (97,370)
2. uterine tube tumor/ (1263)
3. uterine tube carcinoma/ (1899)
4. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (9245)
5. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (136,192)
6. peritoneum cancer/ (3891)
7. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (405,888)
8. or/6-7 (408,204)
9. ovar$.ti,ab,ot. (278,995)
10. 8 and 9 (30,110)
11. 1 or 2 or 3 or 4 or 5 or 10 (168,296)
12. (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. or ultrasound/ or sonography/ (591,945)
13. (CA125$ or CA 125$ or ca 12-5$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (11,962)
14. CA 125 antigen/ (13,650)
15. 13 or 14 (16,956)
16. (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (956)
17. human epididymis protein 4/ (507)
18. or/16-17 (1036)
19. tumor marker/ (62,368)
20. (Tumo?r marker$ or biomarker$ or bio-marker$ or cancer marker$ or neoplasm marker$).ti,ab,ot. (260,389)
21. 19 or 20 (294,218)

22. 12 or 15 or 18 or 21 (887,954)
23. 11 and 22 (25,903)
24. health-economics/ (37,185)
25. exp economic-evaluation/ (262,667)
26. exp health-care-cost/ (249,052)
27. exp pharmacoeconomics/ (184,843)
28. or/24-27 (566,720)
29. (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. (778,313)
30. (expenditure$ not energy).ti,ab. (30,211)
31. (value adj2 money).ti,ab. (1844)
32. budget$.ti,ab. (29,697)
33. or/29-32 (806,888)
34. 28 or 33 (1,107,372)
35. letter.pt. (963,779)
36. editorial.pt. (523,590)
37. note.pt. (663,117)
38. or/35-37 (2,150,486)
39. 34 not 38 (1,007,210)
40. (metabolic adj cost).ti,ab. (1124)
41. ((energy or oxygen) adj cost).ti,ab. (3573)
42. ((energy or oxygen) adj expenditure).ti,ab. (25,235)
43. or/40-42 (29,019)
44. 39 not 43 (1,001,250)
45. exp animal/ (22,704,681)
46. exp animal-experiment/ (2,060,346)
47. nonhuman/ (4,993,357)
48. (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh. (5,418,904)
49. or/45-48 (24,242,905)
50. exp human/ (18,234,517)
51. exp human-experiment/ (393,716)
52. 50 or 51 (18,236,045)
53. 49 not (49 and 52) (6,007,829)
54. 44 not 53 (926,312)
55. 23 and 54 (665)

Economics terms based on costs filter.[158]

NHS Economic Evaluation Database (via Wiley Online Library)

Issue 2 of 4, April 2015.

Date searched: 24 November 2016.

Records found: 11.

Search strategy

#1 MeSH descriptor: [Ovarian Neoplasms] explode all trees (1511)
#2 MeSH descriptor: [Fallopian Tube Neoplasms] this term only (45)
#3 MeSH descriptor: [Uterine Neoplasms] this term only (691)
#4 (AOSCa* or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA* or dysgerminom*):ti,ab,kw (231)

#5 ((ovar* or "high-grade serous" or "low-grade serous" or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) near/5 (cancer* or adenocarcin* or adeno-carcin* or tumo?r* or sarcoma* or neoplas* or metasta* or meta-sta* or carcino* or oncogenesis or malignan* or choriocarcinom* or teratoma* or cystadenocarcin* or rhabdomyosarcom* or rhabdo-myosarcom* or rhabdosarcom* or leiomyosarcoma* or leio-myosarcom* or androblastom* or arrhenoblastom* or adenoma* or lesion* or oncolo*)):ti,ab,kw (7371)
#6 #1 or #2 or #3 or #4 or #5 (7441)
#7 MeSH descriptor: [Peritoneal Neoplasms] this term only (213)
#8 (peritoneum or borderline or epithelial or primary peritoneal):ti,ab,kw (7882)
#9 #7 or #8 (8005)
#10 ovar*:ti,ab,kw (9974)
#11 #9 and #10 (1073)
#12 #6 or #11 (7490)
#13 (ultraso* or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph* or doptone* or echograph* or echogram* or echosound*):ti,ab,kw (21,215)
#14 MeSH descriptor: [Ultrasonography] this term only (956)
#15 #13 or #14 (21,215)
#16 (CA125* or "CA 125*" or "CA 12-5*" or (antigen near/2 125) or (mucin near/1 16) or mucin16 or (muc near/1 16) or muc16):ti,ab,kw (473)
#17 MeSH descriptor: [CA-125 Antigen] this term only (157)
#18 #16 or #17 (473)
#19 ("human epididymis protein 4" or "human epididymal protein 4" or "WAP four disulfide core domain protein 2" or "wap 4 disulfide core domain protein 2" or WFCD2 or EDDM4 or WAP5 or "wap four disulfide core domain 2" or "wap 4 disulfide core domain 2" or "HE 4" or HE4):ti,ab,kw (35)
#20 MeSH descriptor: [Biomarkers, Tumor] this term only(1866)
#21 ("tumo?r marker*" or biomarker* or bio-marker* or "cancer marker*" or "neoplasm marker*"):ti,ab,kw (19,067)
#22 #20 or #21 (19,071)
#23 #15 or #18 or #19 or #22 (40,201)
#24 #12 and #23 (708)

EconLit (via EBSCOhost)

Date range searched: 1966 to 25 November 2016.

Date searched: 25 November 2016.

Records found: 1.

Search strategy

S10	S4 AND S9	1
S9	S5 OR S6 OR S7 OR S8	144
S8	((tumo?r N3 marker*) or biomarker* or bio-marker* or (cancer N3 marker*) or (neoplas* N3 marker*))	92
S7	("human epididymis protein 4" or "human epididymal protein 4" or "WAP four disulfide core domain protein 2" or "wap 4 disulfide core domain protein 2" or WFCD2 or EDDM4 or WAP5 or "wap four disulfide core domain 2" or "wap 4 disulfide core domain 2" or "HE 4" or HE4)	0
S6	(CA125* or CA 125* or ca 12-5* or (antigen N2 "125") or (mucin N1 "16") or mucin16 or (muc N1 "16") or muc16)	4
S5	(ultraso* or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph* or doptone* or echograph* or echogram* or echosound*)	48
S4	S1 or S2 or S3	32
S3	(peritoneum or epithelial or primary peritoneal)	2

S10	**S4 AND S9**	**1**
S2	(AOSCa* or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA* or dysgerminom*)	10
S1	(ovar* or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) N5 (Cancer* or adenocarcin* or adeno-carcin* or tumo?r* or sarcoma* or neoplas* or metasta* or meta-sta* or carcino* or oncogenesis or malignan* or choriocarcinom* or teratoma* or cystadenocarcin* or rhabdomyosarcom* or rhabdo-myosarcom* or rhabdosarcom* or leiomyosarcoma* or leio-myosarcom* or androblastom* or arrhenoblastom* or adenoma* or lesion* or oncolo*)	21

Cost-Effectiveness Analysis Registry (via the internet: www.cearegistry.org)

Date searched: 25 November 2016.

Records found: two.

Search strategy

Search terms	Results
CA125	0
CA-125	1
CA 125	1
Antigen 125	1
Mucin 16	0
Mucin16	0
epidiymis	0
epididymal	0
HE4	0
HE-4	0
WAP 4	0
WAP4	0
WAP four	0
WAP5	0
EDDM4	0
WFCD2	0
Ovarian ultrasound	0
Ovarian ultrasonography	0
Ovarian biomarker	0
Ovarian biomarkers	0
tumor marker	0
tumour marker	0
tumor markers	0
tumour markers	0
Total	**3**
Total after removal of duplicates	**2**

Research Papers in Economics (via the internet: http://repec.org/)

Date searched: 25 November 2016.

Records found: six.

IDEAS search interface

Search strategy

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine luterus | tubal | peritoneum | borderline | epithelial | "primary peritoneal" | AOSCa | HGSC | EOC | HGSOC | LGSC | LGSOC | OVCA | dysgerminoma) + (ultrasound | phonophoresis | sonication | sonification | "ultra sound" | ultrashell | sonograph | doptone | echograph | echogram | echosound)

Records retrieved: one.

Search strategy

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine luterus | tubal | peritoneum | borderline | epithelial | "primary peritoneal" | AOSCa | HGSC | EOC | HGSOC | LGSC | LGSOC | OVCA | dysgerminoma) + (CA125 | "CA 125" | "CA 12-5" | "antigen 125" | "mucin 16" | mucin16 | "muc 16" | muc16)

Records retrieved: three.

Search strategy

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine luterus | tubal | peritoneum | borderline | epithelial | "primary peritoneal" | AOSCa | HGSC | EOC | HGSOC | LGSC | LGSOC | OVCA | dysgerminoma) + ("human epididymis" | "human epididymal" | WAP4 | "WAP 4" | "WAP four" | WFCD2 | EDDM4 | WAP5 | "HE 4" | HE4)

Records retrieved: none.

Search strategy

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine luterus | tubal | peritoneum | borderline | epithelial | "primary peritoneal" | AOSCa | HGSC | EOC | HGSOC | LGSC | LGSOC | OVCA | dysgerminoma) + ("tumor marker" | "tumor markers" | "tumour marker" | "tumour markers" | biomarker | biomarkers | bio-marker | bio-markers | "cancer marker" | "cancer markers | "neoplasm marker" | "neoplasm markers")

Records retrieved: three.

Records retrieved in total: seven.

Records retrieved after duplicates: six.

Key

| OR

+ AND

"..." phrase search

Focused outcomes searches

MEDLINE (via Ovid)

Date range searched: 1946 to week 3 January 2017.

Date searched: 31 January 2017.

Records found: 205.

Search strategy

1. Specialization/ (22,763)
2. Surgical Oncology/ (9)
3. ((medical or surg$ or gyn?ecolog$ or physician$) adj1 (speciali$ or oncolog$)).ti,ab,ot. (18,188)
4. ((special$ or tertiary) adj5 (hospital$ or care$ or healthcare or centre$ or center$ or facility or facilities)).ti,ab,ot. (102,854)
5. (central$ adj5 (hospital$ or care$ or healthcare$ or facility or facilities)).ti,ab,ot. (11,444)
6. exp Tertiary Healthcare/ (601)
7. or/1-6 (148,856)
8. ((general$ or obstetric$ or secondary or regular) adj1 (care or healthcare or surg$ or gyn?ecolog$)).ti,ab,ot. (30,791)
9. exp Secondary Care/ (274)
10. or/8-9 (30,863)
11. 7 and 10 (3252)
12. exp Gynecologic Surgical Procedures/ (74,324)
13. (gyn?ecolog$ adj2 surger$).ti,ab,ot. (5371)
14. or/12-13 (77,049)
15. exp Ovarian Neoplasms/ (73,422)
16. Fallopian Tube Neoplasms/ (2583)
17. Uterine Neoplasms/ (38,852)
18. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (5388)
19. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (96,213)
20. or/15-19 (141,283)
21. Peritoneal Neoplasms/ (13,029)
22. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (298,409)
23. or/21-22 (308,780)
24. ovar$.ti,ab,ot. (208,481)
25. 23 and 24 (21,354)
26. 20 or 25 (145,013)
27. exp Cystadenoma/ (5907)
28. (cystadenoma$ or cystoma$ or cyst$ adenoma$).ti,ab,ot. (5325)
29. Fibroma/ (11,012)
30. (Fibroma$ or acrochordon$ or fibroepithelial or fibrous tumo?r$).ti,ab,ot. (12,675)
31. exp Teratoma/ (19,759)
32. (teratoma$ or dermoid$ or dentigerous cyst$ or dysembryoplastic anomal$ or goiter$ or goitre$ or struma$ or sacrococcygeal fistle$ or teratodermoid cyst$ or teratoid tumo?r$).ti,ab,ot. (34,781)
33. or/27-32 (68,835)
34. Pelvis/ (20,123)

35. exp Adnexa Uteri/ (96,410)
36. (pelvi$ or ovar$ or adnexa$).ti,ab,ot. (312,267)
37. or/34-36 (357,008)
38. 33 and 37 (8928)
39. ((pelvi$ or adnexa$ or ovar$) adj6 (mass or masses)).ti,ab,ot. (7720)
40. 14 or 26 or 38 or 39 (213,422)
41. 11 and 40 (205)

MEDLINE Epub Ahead of Print (via Ovid), MEDLINE In-Process & Other Non-Indexed Citations and MEDLINE Daily Update

MEDLINE Epub Ahead of Print (via Ovid): to 30 January 2017.

MEDLINE In-Process & Other Non-Indexed Citations: to 30 January 2017.

MEDLINE Daily Update: to 30 January 2017.

Date searched: 31 January 2017.

Records found: 29.

Search strategy

1. Specialization/ (12)
2. Surgical Oncology/ (2)
3. ((medical or surg$ or gyn?ecolog$ or physician$) adj1 (speciali$ or oncolog$)).ti,ab,ot. (3288)
4. ((special$ or tertiary) adj5 (hospital$ or care$ or healthcare or centre$ or center$ or facility or facilities)).ti,ab,ot. (20,479)
5. (central$ adj5 (hospital$ or care$ or healthcare$ or facility or facilities)).ti,ab,ot. (1681)
6. exp Tertiary Healthcare/ (4)
7. or/1-6 (24,851)
8. ((general$ or obstetric$ or secondary or regular) adj1 (care or healthcare or surg$ or gyn?ecolog$)).ti,ab,ot. (4328)
9. exp Secondary Care/ (2)
10. or/8-9 (4328)
11. 7 and 10 (484)
12. exp Gynecologic Surgical Procedures/ (87)
13. (gyn?ecolog$ adj2 surger$).ti,ab,ot. (683)
14. or/12-13 (766)
15. exp Ovarian Neoplasms/ (331)
16. Fallopian Tube Neoplasms/ (5)
17. Uterine Neoplasms/ (57)
18. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (1232)
19. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (12,027)
20. or/15-19 (12,208)
21. Peritoneal Neoplasms/ (41)
22. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (28,926)
23. or/21-22 (28,956)
24. ovar$.ti,ab,ot. (20,892)

25. 23 and 24 (2883)
26. 20 or 25 (12,554)
27. exp Cystadenoma/ (7)
28. (cystadenoma$ or cystoma$ or cyst$ adenoma$).ti,ab,ot. (486)
29. Fibroma/ (5)
30. (Fibroma$ or acrochordon$ or fibroepithelial or fibrous tumo?r$).ti,ab,ot. (1688)
31. exp Teratoma/ (31)
32. (teratoma$ or dermoid$ or dentigerous cyst$ or dysembryoplastic anomal$ or goiter$ or goitre$ or struma$ or sacrococcygeal fistle$ or teratodermoid cyst$ or teratoid tumo?r$).ti,ab,ot. (3571)
33. or/27-32 (5709)
34. Pelvis/ (13)
35. exp Adnexa Uteri/ (82)
36. (pelvi$ or ovar$ or adnexa$).ti,ab,ot. (33,945)
37. or/34-36 (33,961)
38. 33 and 37 (792)
39. ((pelvi$ or adnexa$ or ovar$) adj6 (mass or masses)).ti,ab,ot. (1272)
40. 14 or 26 or 38 or 39 (14,145)
41. 11 and 40 (29)

EMBASE (via Ovid)

Date range searched: 1974 to 30 January 2017.

Date searched: 31 January 2017.

Records found: 524.

Search strategy

1. medical specialist/ (102,067)
2. ((medical or surg$ or gyn?ecolog$ or physician$) adj1 (speciali$ or oncolog$)).ti,ab,ot. (40,274)
3. ((special$ or tertiary) adj5 (hospital$ or care$ or healthcare or centre$ or center$ or facility or facilities)).ti,ab,ot. (189,219)
4. (central$ adj5 (hospital$ or care$ or healthcare$ or facility or facilities)).ti,ab,ot. (18,866)
5. exp tertiary healthcare/ (71,024)
6. or/1-5 (326,375)
7. ((general$ or obstetric$ or secondary or regular) adj1 (care or healthcare or surg$ or gyn?ecolog$)).ti,ab,ot. (47054)
8. exp secondary healthcare/ (4685)
9. or/7-8 (48,456)
10. 6 and 9 (7200)
11. (gyn?ecolog$ adj2 surger$).ti,ab,ot. (8507)
12. exp gynecologic surgery/ (132,958)
13. or/11-12 (135,722)
14. exp ovary cancer/ (99,193)
15. uterine tube tumor/ (1280)
16. uterine tube carcinoma/ (1938)
17. (AOSCa$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA$ or dysgerminom$).ti,ab,ot. (9541)
18. ((ovar$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer$ or adenocarcin$ or adeno-carcin$ or tumo?r$ or sarcoma$ or neoplas$ or metasta$ or meta-sta$ or carcino$ or oncogenesis or malignan$ or choriocarcinom$ or teratoma$ or cystadenocarcin$ or rhabdomyosarcom$ or rhabdo-myosarcom$ or rhabdosarcom$ or leiomyosarcoma$ or leio-myosarcom$ or androblastom$ or arrhenoblastom$ or adenoma$ or lesion$ or oncolo$)).ti,ab,ot. (138,404)

19. exp peritoneum cancer/ (13,274)
20. (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (412,063)
21. or/19-20 (422,066)
22. ovar$.ti,ab,ot. (283,143)
23. 21 and 22 (31,658)
24. 14 or 15 or 16 or 17 or 18 or 23 (171,059)
25. cystadenoma/ (7719)
26. (cystadenoma$ or cystoma$ or cyst$ adenoma$).ti,ab,ot. (7325)
27. fibroma/ (11,802)
28. (Fibroma$ or acrochordon$ or fibroepithelial or fibrous tumo?r$).ti,ab,ot. (16,228)
29. teratoma/ (25,747)
30. ovary teratoma/ (2771)
31. (teratoma$ or dermoid$ or dentigerous cyst$ or dysembryoplastic anomal$ or goiter$ or goitre$ or struma$ or sacrococcygeal fistle$ or teratodermoid cyst$ or teratoid tumo?r$).ti,ab,ot. (43,888)
32. or/25-30 (58,645)
33. pelvis/ (67,918)
34. ovary/ (67,924)
35. (pelvi$ or ovar$ or adnexa$).ti,ab,ot. (445,224)
36. or/33-35 (462,883)
37. 32 and 36 (10,871)
38. ((pelvi$ or adnexa$ or ovar$) adj6 (mass or masses)).ti,ab,ot. (12,936)
39. 13 or 24 or 37 or 38 (295,584)
40. 10 and 39 (524)

Cochrane Database of Systematic Reviews (via Wiley Online Library), Database of Abstracts of Reviews of Effects (via Wiley Online Library), Cochrane Central Register of Controlled Trials (via Wiley Online Library), Health Technology Assessment Database (via Wiley Online Library) and NHS Economic Evaluation Database (via Wiley Online Library)

Cochrane Database of Systematic Reviews (via Wiley Online Library): Issue 1, January 2017.

Database of Abstracts of Reviews of Effects (via Wiley Online Library): Issue 2, April 2015.

Cochrane Central Register of Controlled Trials (via Wiley Online Library): Issue 11, November 2016.

Health Technology Assessment Database (via Wiley Online Library): Issue 4 of 4, October 2016.

NHS Economic Evaluation Database (via Wiley Online Library): Issue 2 of 4, April 2015.

Date searched: 31 January 2017.

Records found: 23.

Cochrane Database of Systematic Reviews: six.

Database of Abstracts of Reviews of Effects: none.

Cochrane Central Register of Controlled Trials: 17.

NHS Economic Evaluation Database: none.

Health Technology Assessment Database: none.

Search strategy

#1 MeSH descriptor: [Specialization] explode all trees (107)
#2 ((medical or surg* or gynaecolog* or gynecolog* or physician*) near/1 (speciali* or oncolog*)):ti,ab,kw (2745)
#3 ((special* or tertiary) near/5 (hospital* or care* or healthcare or centre* or center* or facility or facilities)):ti,ab,kw (8821)
#4 (central* near/5 (hospital* or care* or healthcare* or facility or facilities)):ti,ab,kw (1006)
#5 MeSH descriptor: [Tertiary Healthcare] explode all trees (7)
#6 #1 or #2 or #3 or #4 or #5 (12,086)
#7 ((general* or obstetric* or secondary or regular) near/1 (care or healthcare or surg* or gynaecolog* or gynecolog*)):ti,ab,kw (3877)
#8 MeSH descriptor: [Secondary Care] explode all trees (22)
#9 #7 or #8 (3877)
#10 #6 and #9 (306)
#11 MeSH descriptor: [Gynecologic Surgical Procedures] explode all trees (4254)
#12 ((gynaecolog* or gynecolog*) near/2 surger*):ti,ab,kw (1907)
#13 #11 or #12 (5608)
#14 MeSH descriptor: [Ovarian Neoplasms] explode all trees (1513)
#15 MeSH descriptor: [Fallopian Tube Neoplasms] explode all trees (45)
#16 MeSH descriptor: [Uterine Neoplasms] explode all trees (3024)
#17 (AOSCa* or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA* or dysgerminom*):ti,ab,kw (235)
#18 ((ovar* or high-grade-serous or low-grade-serous or sertoli-leydig-cell or fallopian or oviduct or uterine or uterus or tubal) near/5 (Cancer* or adenocarcin* or adeno-carcin* or tumor* or tumour* or sarcoma* or neoplas* or metasta* or meta-sta* or carcino* or oncogenesis or malignan* or choriocarcinom* or teratoma* or cystadenocarcin* or rhabdomyosarcom* or rhabdo-myosarcom* or rhabdosarcom* or leiomyosarcoma* or leio-myosarcom* or androblastom* or arrhenoblastom* or adenoma* or lesion* or oncolo*)):ti,ab,kw (7517)
#19 #14 or #15 or #16 or #17 or #18 (7909)
#20 MeSH descriptor: [Peritoneal Neoplasms] explode all trees (213)
#21 (peritoneum or borderline or epithelial or primary peritoneal):ti,ab,kw (7962)
#22 #20 or #21 (8085)
#23 ovar*:ti,ab,kw (10,032)
#24 #22 and #23 (1080)
#25 #19 or #24 (7947)
#26 MeSH descriptor: [Cystadenoma] explode all trees (4)
#27 (cystadenoma* or cystoma* or cyst* adenoma*):ti,ab,kw (101)
#28 MeSH descriptor: [Fibroma] explode all trees (8)
#29 (Fibroma* or acrochordon* or fibroepithelial or fibrous-tumour* or fibrous-tumor*):ti,ab,kw (57)
#30 MeSH descriptor: [Teratoma] explode all trees (29)
#31 (teratoma* or dermoid* or dentigerous-cyst* or dysembryoplastic-anomal* or goiter* or goitre* or struma* or sacrococcygeal-fistle* or teratodermoid-cyst* or teratoid-tumour* or teratoid-tumor*):ti,ab,kw (580)
#32 #26 or #27 or #28 or #29 or #30 or #31 (728)
#33 MeSH descriptor: [Pelvis] explode all trees (815)
#34 MeSH descriptor: [Adnexa Uteri] explode all trees (1277)
#35 (pelvi* or ovar* or adnexa*):ti,ab,kw (17,164)
#36 #33 or #34 or #35 (17,379)
#37 #32 and #36 (55)
#38 ((pelvi* or adnexa* or ovar*) near/6 (mass or masses)):ti,ab,kw (268)
#39 #13 or #25 or #37 or #38 (12,832)
#40 #10 and #39 (23)

Appendix 2 Excluded studies

To be included in the review, studies had to fulfil the following criteria:

- population: people of any age with suspected ovarian cancer
- setting: secondary care
- index test: ROMA score, simple ultrasound rules (IOTA group), ADNEX model (IOTA group), Overa (MIA2G), RMI 1 (using decision thresholds other than 250)
- reference standard: histological examination of a surgically resected of biopsy sample; studies that used follow-up as the reference standard for some or all test negative patients were also eligible for inclusion
- outcome: sufficient data to construct 2 × 2 table of test performance or clinical outcomes.

The following table summarises the studies that were screened for inclusion based on full-text publication, but did not fulfil one or more of the above criteria. Studies were assessed sequentially against the criteria; the first criterion failed is classified as the reason for exclusion. The table shows which of the criteria each study fulfilled ('Yes') and on which items it failed ('No'), as well as any that were 'Unclear'. Articles that did not report primary research were not assessed further. Any criteria that are not applicable to a study are marked as NA.

Details of excluded studies with rationale for exclusion

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Abbott Laboratories (Singapore). Evaluation of HE4 and CA125 Serum Markers to Improve the Risk Determination of Ovarian Cancer in Malaysian women. In *WHO International Clinical Trials Registry Platform (ICTRP)* [Internet]. Geneva: World Health Organization; 2014. URL: http://isrctn.com/ISRCTN45238573 (accessed 24 November 2016)	No	Yes	Yes	No	Unclear	No	Trial registry entry for completed study, no results or publications posted
Abdalla N, Bachanek M, Winiarek J, Cendrowski K, Sawicki W. Analysis of the diagnostic value of logistic regression model and HE4 in the presurgical assessment of adnexal masses. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):379	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention IOTA group's regression model (not the IOTA group's simple rules or the ADNEX model)
Abdulrahman GO Jr, McKnight L, Lutchman Singh K. Risk of malignancy index in women with adnexal masses – comparing RMI 1, 2 and 3 in the Welsh population. *Int J Gynecol Cancer* 2012;**22**:E411	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Accuracy of RMI 1 at a threshold of 200, in a tertiary care setting
Abdulrahman GO Jr, McKnight L, Lutchman Singh K. The risk of malignancy index (RMI) in women with adnexal masses in Wales. *Taiwan J Obstet Gynecol* 2014;**53**:376–81	Yes	Yes	No	No	No	Yes	No relevant intervention Accuracy of RMI at a threshold of 200, in a tertiary care setting
Akdeniz N, Kuyumcuoglu U, Kale A, Erdemoglu M, Caca F. Risk of malignancy index for adnexal masses. *Eur J Gynaecol Oncol* 2009;**30**:178–80	Yes	Yes	Unclear	No	Yes	No	No relevant outcomes Insufficient information to calculate sensitivity and specificity
Alanbay I, Akturk E, Coksuer H, Ercan M, Karasahin E, Dede M, *et al.* Comparison of risk of malignancy index (RMI), CA125, CA 19–9, ultrasound score, and menopausal status in borderline ovarian tumor. *Gynecol Endocrinol* 2012;**28**:478–82	Yes	No	Yes	No	Unclear	Yes	Case–control study (benign *vs.* borderline) No relevant intervention RMI 4

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Alcazar JL, Pascual MA, Graupera B, Auba M, Errasti T, Olartecoechea B, *et al.* External validation of IOTA simple descriptors and simple rules for classifying adnexal masses. *Ultrasound Obstet Gynecol* 2016;**48**:397–402	Yes	No	Yes	No	Yes	Yes	No relevant intervention IOTA group's simple ultrasound rules in combination with other rules not included in this assessment Selected population (not classifiable using the IOTA group's simple descriptors)
Al Musalhi K, Al-Kindi M, Ramadhan F, Al-Rawahi T, Al-Hatali K, Mula-Abed WA. Validity of cancer antigen-125 (CA-125) and risk of malignancy index (RMI) in the diagnosis of ovarian cancer. *Oman Med J* 2015;**30**:428–34	Yes	Yes	Yes	No	No	No	No relevant intervention Study of RMI 2
Andersen ES, Knudsen A, Rix P, Johansen B. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *Gynecol Oncol* 2003;**90**:109–12	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Anton C, Carvalho FM, Oliveira EI, Maciel G, Baracat EC, Carvalho JP. Comparison of four methods for classification of ovarian masses using CA125, HE4, risk of malignancy index, and ROMA. *Int J Gynecol Cancer* 2011;**21**(Suppl. 3):658	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA score using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Anton C, Carvalho FM, Oliveira EI, Maciel GA, Baracat EC, Carvalho JP. A comparison of CA125, HE4, risk ovarian malignancy algorithm (ROMA), and risk malignancy index (RMI) for the classification of ovarian masses. *Clinics (São Paulo)* 2012;**67**:437–41	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA score using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Antovska V, Dimitrov G, Aleksioska N. Our modification of risk of malignancy index in patients with ovarian malignancy. *Int J Gynecol Cancer* 2011;**21**(Suppl. 3):820	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Accuracy of RMI at a threshold of 200

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Araujo KG, Jales RM, Pereira PN, Yoshida A, de Angelo Andrade L, Sarian LO, *et al.* Performance of the IOTA ADNEX model in the preoperative discrimination of adnexal masses in a gynecologic oncology centre. *Ultrasound Obstet Gynecol* 2016;**19**:19	Yes	Yes	No	Yes	Yes	Yes	Tertiary care gynaecological oncology centre Threshold optimisation study
Arun-Muthuvel V, Jaya V. Pre-operative evaluation of ovarian tumors by risk of malignancy index, CA125 and ultrasound. *Asian Pac J Cancer Prev* 2014;**15**:2929–32	Yes	Yes	No	Yes	Yes	Yes	Tertiary care setting (women scheduled for surgery in a gynaecological oncology department)
Ashrafgangooei T, Rezaeezadeh M. Risk of malignancy index in preoperative evaluation of pelvic masses. *Asian Pac J Cancer Prev* 2011;**12**:1727–30	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Threshold optimisation study for RMI, data for cut-off value of 238
Ashrafganjooei T. Risk of malignancy index in evaluation of pelvic masses. *Int J Gynecol Cancer* 2011;**21**(Suppl. 3):673	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Optimised RMI threshold (238)
Ashrafganjooei T. Risk of malignancy index in evaluation of pelvic masses. *Int J Gynecol Cancer* 2011;**21**(Suppl. 2):96	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention RMI (unspecified threshold)
Aslam N, Banerjee S, Carr JV, Savvas M, Hooper R, Jurkovic D. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. *Obstet Gynecol* 2000;**96**:75–80	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Validation of regression models (not interventions included in this assessment)
Auge JM, Molina R, Escudero JM, Foj L, Filella X, Fuste P. HE-4 utility to increase efficiency in patients with abdominal masses. *Clin Chem Lab Med* 2014;**52**:S365	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Bailey J, Tailor A, Naik R, Lopes A, Godfrey K, Hatem HM, *et al.* Risk of malignancy index for referral of ovarian cancer cases to a tertiary centre: does it identify the correct cases? *Int J Gynecol Cancer* 2006;**16**(Suppl. 1):30–4	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200 in a tertiary care setting
Bensaid C, Le Frere Belda MA, Metzger U, Larousserie F, Clement D, Chatellier G, *et al.* Performance of laparoscopy in identifying malignant ovarian cysts. *Surg Endosc* 2006;**20**:1410–4	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200 in a tertiary care setting

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Braicu E, Torsten U, Richter R, Zimmermann M, Chekerov R, Kronenberger C, *et al.* Value of biomarkers and sonography in predicting malignancy in pelvic mass patients. Preliminary results from prospective, multicentric, ongoing study. *Int J Gynecol Cancer* 2014;**24**(Suppl. 4):366–7	Yes	Yes	Unclear	Yes	Yes	No	No relevant outcomes Insufficient information to calculate sensitivity and specificity
Braicu EI, Torsten U, Mecke H, Richter R, Ames K, Hellmeyer L, *et al.* Role of HE4, CA125, and ultrasound in risk assessment in pelvic mass patients: results from a prospective, multicentric study. *J Clin Oncol* 2015;**33**:5535	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Braicu EI, Torsten U, Mecke H, Richter R, Hellmeyer L, Nohe G, *et al.* HE4 performs better than CA125 as a diagnostic biomarker in premenopausal pelvic mass patients. Final results from a prospective, multicentric study. *Int J Gynecol Cancer* 2016;**26**:21–2	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Blontzos N, Vorgias G, Papatheodorou D, Vylliotou V, Novkovic N, Diakosavas M, *et al.* The clinical value of adding HE4 and ROMA index to CA-125 in the preoperative workout of adnexal masses. *Int J Gynecol Cancer* 2016;**26**:172	Yes	Yes	Unclear	No	Yes	No	No relevant intervention ROMA assays and threshold NR
Bristow RE, Hodeib M, Smith A, Chan DW, Zhang Z, Fung ET, *et al.* Impact of a multivariate index assay on referral patterns for surgical management of an adnexal mass. *Am J Obstet Gynecol* 2013;**209**:581	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention
Cacho R, Sia Su L. Distinguishing the benign and malignant adnexal mass: a prospective external validation of a risk of malignancy index (RMI) based on intra-operative features. *Int J Gynaecol Obstet* 2009;**107**:S136	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Validation of an unspecified RMI scoring system
Campos C, Sarian LO, Jales RM, Hartman C, Araujo KG, Pitta D, *et al.* Performance of the Risk of Malignancy Index for Discriminating Malignant Tumours in Women With Adnexal Masses. *J Ultrasound Med* 2016;**35**:143–52	Yes	Yes	No	Yes	Yes	Yes	Accuracy of RMI 1 in a tertiary care setting

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Chia YN, Marsden DE, Robertson G, Hacker NF. Triage of ovarian masses. *Aust N Z J Obstet Gynaecol* 2008;**48**:322–8	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200, in a tertiary care setting
Chopra S, Vaishya R, Kaur J. An Evaluation of the applicability of the risk of malignancy index for adnexal masses to patients seen at a tertiary hospital in Chandigarh, India. *J Obstet Gynaecol India* 2015;**65**:405–10	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 2
Chudecka-Glaz A, Cymbaluk-Ploska A, Jastrzebska J, Menkiszak J. Can ROMA algorithm stratify ovarian tumor patients better when being based on specific age ranges instead of the premenopausal and postmenopausal status? *Tumour Biol* 2016;**37**:8879–87	Yes	Yes	Unclear	No	Unclear	Accuracy	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Chudecka-Glaz A, Cymbaluk-Ploska A, Luterek-Puszynska K, Menkiszak J. Diagnostic usefulness of the Risk of Ovarian Malignancy Algorithm using the electrochemiluminescence immunoassay for HE4 and the chemiluminescence microparticle immunoassay for CA125. *Oncol Lett* 2016;**12**:3101–14	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Clarke SE, Grimshaw R, Rittenberg P, Kieser K, Bentley J. Risk of Malignancy Index in the Evaluation of Patients With Adnexal Masses. *J Obstet Gynaecol Can* 2009;**31**:440–5	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 120, in a tertiary care setting
Daemen A, Valentin L, Fruscio R, Van Holsbeke C, Melis GB, Guerriero S, *et al.* Improving the preoperative classification of adnexal masses as benign or malignant by second-stage tests. *Ultrasound Obstet Gynecol* 2011;**37**:100–6	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA group data set used to evaluate performance of different regression models
Dasari P, Pannirselvan PCL, Sridhar MG. Ultrasonographic scoring and risk of malignancy index in preoperative prediction of ovarian malignancy. *J Gynecol Surg* 2013;**29**:61–4	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 2

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Ellerbrock J, Mertens H, Engelen M, Bergmans M, Nolting E, Kruitwagen R. Evaluation of the risk of malignancy index performance for referral in the south-eastern part of the Netherlands. *Int J Gynecol Cancer* 2011;**21**(Suppl. 3):1269	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200, in a tertiary care setting
Elsawy MM, Meleiss M, Abdel Sattar HR, Abo Ollo M. Prospective study using the risk of ovarian malignancy algorithm for detection of ovarian cancer in Egypt. *Int J Gynecol Cancer* 2012;**22**:E317	Yes	No	Unclear	No	Unclear	Yes	Case–control study No relevant intervention ROMA assays and threshold NR
Enakpene CA, Omigbodun AO, Goecke TW, Odukogbe AT, Beckmann MW. Preoperative evaluation and triage of women with suspicious adnexal masses using risk of malignancy index. *J Obstet Gynaecol Res* 2009;**35**:131–8	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 250, in a tertiary care setting
Ertas S, Vural F, Vural F, Tufekci EC, Ertas AC, Kose G, *et al.* Predictive value of malignancy risk indices for ovarian masses in premenopausal and postmenopausal women. *Asian Pac J Cancer Prev* 2016;**17**:2177–83	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Evelyne M, Jeroen K, Roy K, Arnold-Jan K, Brigitte S, Ben Van C, *et al.* Subjective assessment of grey scale and colour Doppler ultrasound features versus the International Ovarian Tumour Analysis (IOTA) logistic regression (LR2) model versus simple ultrasound rules versus Risk of Malignancy Index (RMI) for diagnosing ovarian cancer in women with an adnexal mass. 2013	No	Yes					PROSPERO registration for a relevant systematic review
Farzaneh F, Honarvar Z, Yaraghi M, Yaseri M, Arab M, Hosseini M, *et al.* Preoperative evaluation of risk of ovarian malignancy algorithm index in prediction of malignancy of adnexal masses. *Iran Red Crescent Med J* 2014;**16**:e17185	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA score using different manufacturers' assays for CA125 and HE4 measurements measurements (not a valid CE-marked intervention)

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Froyman W, Landolfo C, Bourne T, Cock BD, Testa A, Valentin L, *et al.* Performance of the RMI and IOTA ADNEX and simple rules risk model in the evaluation of adnexal masses not classifiable using the easy descriptors as first step. *BJOG* 2016;**123**:83–4	Yes	No	Unclear	Yes	Yes	No	Selected, 'difficult to diagnose' tumours No relevant outcomes Senisitivity and specificity data not fully reported
Fujirebio Diagnostics I. *New Biomarkers Evaluating Ovarian Cancer*. 2014. URL: https://ClinicalTrials.gov/show/NCT01466049 (accessed 5 July 2018)	No	Yes	Yes	No	Unclear	No	Trial registry entry for completed study, no results or publications posted
Gasparov AS, Zhordania, Paianidi Iu G, Dubinskaia ED. [Oncogynecological aspects of adnexal masses.] *Vestn Ross Akad Med Nauk* 2013;**8**:9–13	Yes	Yes	Unclear	No	Unclear	No	No relevant outcomes
Gramellini D, Fieni S, Sanapo L, Casilla G, Verrotti C, Nardelli GB. Diagnostic accuracy of IOTA ultrasound morphology in the hands of less experienced sonographers. *Aust N Z J Obstet Gynaecol* 2008;**48**:195–201	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention
Grenache DG, Vucetic Z. Comparison of two multimarker serum tests for the prediction of ovarian cancer in women with a pelvic mass. *J Clin Oncol* 2013;**31**(Suppl.):A5555	Yes	No	Unclear	Yes	Yes	No	Not patients with suspected ovarian cancer
Grenache DG, Heichman KA, Werner TL, Vucetic Z. Clinical performance of two multi-marker blood tests for predicting malignancy in women with an adnexal mass. *Clin Chim Acta* 2015;**438**:358–63	Yes	No	Unclear	Yes	Yes	No	Not patients with suspected ovarian cancer
Guerriero S, Saba L, Ajossa S, Peddes C, Sedda F, Piras A, *et al.* Assessing the reproducibility of the IOTA simple ultrasound rules for classifying adnexal masses as benign or malignant using stored 3D volumes. *Eur J Obstet Gynecol Reprod Biol* 2013;**171**:157–60	Yes	No	No	Yes	NA	Yes	Not a clinical study in patients with suspected ovarian cancer IOTA group's training study, using video clips
Gulati A, Sharma A, Suneja A, Vaid NB, Sharma S, Yadav P. Comparison of ovarian crescent sign & risk of malignancy index in prediction of ovarian malignancy. *Int J Gynecol Cancer* 2011;**21**(Suppl. 2):117	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI (unspecified threshold)

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Hagen B, Tingulstad S, Onsrud M, Moen M, Kiserud T, Eik-Nes S, *et al.* [Preoperative identification of malignancy among women with a pelvic mass. Evaluation of a risk index based on ultrasound findings. CA 125 in serum and menopausal status.] *Tidsskr Nor Laegeforen* 1995;**115**:820–2	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Harry VN, Narayansingh GV, Parkin DE. The risk of malignancy index for ovarian tumours in Northeast Scotland – a population based study. *Scott Med J* 2009;**54**:21–3	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 2
He G, Holcroft CA, Beauchamp MC, Yasmeen A, Ferenczy A, Kendall-Dupont J, *et al.* Combination of serum biomarkers to differentiate malignant from benign ovarian tumours. *J Obstet Gynaecol Can* 2012;**34**:567–74	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Regression model, including multiple biomarkers and RMI
Hodeib M, Bristow RE, Smith A, Zhang Z, Chan DW, Fung ET, *et al.* Impact of a multivariate index assay on referral patterns for surgical management of an adnexal mass. *Gynecol Oncol* 2013;**131**:258	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Multivariate index assay (MIA), Overa 1
Hogdall E, Karlesn MA, Christensen IJ, Lundvall L, Engelholm SA, Nedergaard L, *et al.* Diagnostic value of HE4, CA125 and the ROMA index in ovarian cancer patients from a tertiary centre. *Int J Gynecol Cancer* 2012;**22**:S42	Yes	Yes	No	No	Unclear	No	No relevant intervention Tertiary care, ROMA assays and threshold NR, RMI threshold NR
Ikiz N, Guvenal T, Taneli F, Koyuncu FM, Kandiloglu AR, Bilge S, *et al.* Comparison of ROMA (risk of ovarian malignancy algorithm), RMI (risk of malignancy index) and OTI (ovarian tumour index) in patients with adnexal mass. *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):905	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Imperial NA, Angeli N, Rivera W, Wilhelmina. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *J Obstet Gynaecol Res* 2015;**41**:77	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Optimised RMI threshold (273)
Imperial NA, Rivera W. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *BJOG* 2015;**122**:137	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200 and optimised threshold (273)

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Imperial NA, Rivera W. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *Int J Gynaecol Obstet* 2015;**131**:E412	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Irshad F, Irshad M, Naz M, Asim Ikram M. Accuracy of 'risk of malignancy index' in the preoperative diagnosis of Zovarian malignancy in post menopausal women. *Rawal Med J* 2013;**38**:266–70	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 250
Jabeen R, Khan SA, Naveed S. Risk of Malignancy Index in the preoperative evaluation of patients with ovarian masses. *Rawal Med J* 2015;**40**:78–80	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Jacob F, Meier M, Caduff R, Goldstein D, Pochechueva T, Hacker N, *et al.* No benefit from combining HE4 and CA125 as ovarian tumour markers in a clinical setting. *Gynecol Oncol* 2011;**121**:487–91	Yes	No	Unclear	Yes	Yes	Yes	Not patients with suspected ovarian cancer
Jarvis S. The ROMA (Risk of Ovarian Malignancy Algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: is it really useful? *Ann Clin Biochem* 2011;**48**:392	No	Yes					Not a primary study
Javdekar R, Maitra N. Risk of Malignancy Index (RMI) in evaluation of adnexal mass. *J Obstet Gynaecol India* 2015;**65**:117–21	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 2
Kaijser J, Van Gorp T, Van Hoorde K, Van Holsbeke C, Bourne T, Vergote I, *et al.* Serum CA-125 and HE-4 versus an ultrasound based predictive model to assess risk of malignancy in women with adnexal masses. *Int J Gynecol Cancer* 2012;**22**:E149–50	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Kaijser J, Van Gorp T, Sayasneh A, Vergote I, Bourne T, Van Calster B, *et al.* Differentiating stage I epithelial ovarian cancer from benign disease in women with adnexal tumors using biomarkers or the ROMA algorithm. *Gynecol Oncol* 2013;**130**:398–9	No	Yes					Not a primary study
Kalapotharakos G, Asciutto C, Henic E, Casslen B, Borgfeldt C. High preoperative blood levels of HE4 predicts poor prognosis in patients with ovarian cancer. *J Ovarian Res* 2012;**5**:20	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA score using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Kader Ali Mohan GR, Jaaback K, Proietto A, Robertson R, Angstetra D. Risk Malignancy Index (RMI) in patients with abnormal pelvic mass: Comparing RMI 1, 2 and 3 in an Australian population. *Aust N Z J Obstet Gynaecol* 2010;**50**:77–80	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Kadija S, Stefanovic A, Jeremic K, Radojevic MM, Nikolic L, Markovic I, *et al.* The utility of human epididymal protein 4, cancer antigen 125, and risk for malignancy algorithm in ovarian cancer and endometriosis. *Int J Gynecol Cancer* 2012;**22**:238–44	Yes	Yes	Yes	No	Yes	No	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Karimi-Zarchi M, Mojaver SP, Rouhi M, Hekmatimoghaddam SH, Moghaddam RN, Yazdian-Anari P, *et al.* Diagnostic value of the Risk of Malignancy Index (RMI) for detection of pelvic malignancies compared with pathology. *Electron Physician* 2015;**7**:1505–10	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 70
Karlsen MA, Hogdall EV, Christensen IJ, Borgfeldt C, Kalapotharakos G, Zdrazilova-Dubska L, *et al.* A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer – an international multicenter study in women with an ovarian mass. *Gynecol Oncol* 2015;**138**:640–6	Yes	Unclear	No	No	No	No	No relevant intervention Risk model development (Copenhagen Index) using data from existing studies and stored blood samples
Keogh F, Tan AL, Eva LJ. HE4 as a tumour marker for the prediction of ovarian carcinoma. *BJOG* 2015;**122**:137–8	Yes	Yes	Unclear	No	Yes	No	No relevant intervention ROMA assays and threshold NR
Kho CZB, Chong YW, Lee YT, Krishnaswamy G, Ong CL, Lam SL, *et al.* Preoperative evaluation of paediatric adnexal masses with paediatric risk of malignancy index improves ovarian conservation and surgical morbidity. *Pediatr Blood Cancer* 2015;**62**:S187	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Paediatric version of RMI (not a specified intervention)
Ko HS, Kim N, Park YG. Re: interobserver agreement in describing adnexal masses using the International Ovarian Tumour Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. *Ultrasound Obstet Gynecol* 2015;**45**:238	No	Yes					Not a primary study

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Kondalsamy-Chennakesavan S, Obermair A. Differentiating stage I epithelial ovarian cancer from benign disease in women with adnexal tumours using biomarkers or the ROMA algorithm. *Gynecol Oncol* 2013;**130**:400	No	Yes					Not a primary study
Kondalsamy-Chennakesavan S, Hackethal A, Bowtell D, Australian Ovarian Cancer Study Group, Obermair A. Differentiating stage 1 epithelial ovarian cancer from benign ovarian tumours using a combination of tumour markers HE4, CA125, and CEA and patient's age. *Gynecol Oncol* 2013;**129**:467–71	Yes	No	Unclear	Yes	Yes	Yes	Diagnostic case–control study
Lasho MA, Algeciras-Schimnich A. Determination of ROMA score performance using the roche elecsys HE4 and CA 125 immunoassays. *Clin Chem* 2014;**60**(Suppl. 1):S12–13	Yes	Yes	Unclear	Yes	Unclear	Yes	Reference standard unspecified
Leelahakorn S, Tangjitgamol S, Manusirivithaya S, Thongsuksai P, Jaroenchainon P, Jivangkul C. Comparison of ultrasound score, CA125, menopausal status, and risk of malignancy index in differentiating between benign and borderline or malignant ovarian tumors. *J Med Assoc Thai* 2005;**88**(Suppl. 2):22–30	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Not RMI 1
Li AJ. New biomarkers for the diagnosis of ovarian carcinoma: OVA1 and ROMA. [Italian.] *G Ital Ostet Ginecol* 2012;**34**:409–14	No	Yes					Not a primary study
Li ZQ, Smalley RJ, Glover CL, Raju S, Falcone K, Fegely M, *et al.* Comparison of serum CYFRA 21–1 and ROMA in distinguishing ovarian cancer from benign pelvic masses. *J Clin Oncol* 2012	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Loh AHP, Ong CL, Lam SL, Chua JHY, Chui CH. Risk of malignancy index for preoperative evaluation of paediatric ovarian tumors. *Pediatr Blood Cancer* 2010;**55**:785	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Development of new paediatric risk indices
Lokich E, Palisoul M, Romano N, Craig Miller M, Robison K, Stuckey A, *et al.* Assessing the risk of ovarian malignancy algorithm for the conservative management of women with a pelvic mass. *Gynecol Oncol* 2015;**139**:248–52	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Lokich E, Palisoul M, Romano N, Stuckey AR, Robison KM, DiSilvestro PA, *et al.* ROMA guided conservative management for women diagnosed with an ovarian cyst or pelvic mass. *Gynecol Oncol* 2015;**137**:21	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Tertiary care, ROMA assays not specified
Longoria T, Ueland F, Zhang Z, Chan D, Smith A, Fung E, *et al.* Clinical performance of a multivariate index assay for detecting early-stage ovarian cancer. *Gynecol Oncol* 2013;**131**:259	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention MIA, Overa 1
Ma S, Shen K, Lang J. [Effect of a risk of malignancy index in preoperative diagnosis of ovarian cancer.] *Zhonghua Fu Chan Ke Za Zhi* 2001;**36**:162–4	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. *Chin Med J* 2003;**116**:396–9	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention Data for various RMI thresholds (50 to 1000, not including 250)
Maitra NK, Javadekar R. Risk of malignancy index in the evaluation of adnexal mass. *BJOG* 2014;**121**:206	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 2
Manegold-Brauer G, Schoetzau A, Hacker N, Lapaire O, Heinzelmann-Schwarz V. Proposal of a new two-step use of the risk of malignancy index in a general gynecological outpatient setting as compared to a gynecological cancer center. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):223	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Accuracy of RMI at a threshold of 200
Mansour GM, El-Lamie IK, El-Sayed HM, Ibrahim AM, Laban M, Abou-Louz SK, *et al.* Adnexal mass vascularity assessed by 3-dimensional power Doppler: does it add to the risk of malignancy index in prediction of ovarian malignancy?: four hundred-case study. *Int J Gynecol Cancer* 2009;**19**:867–72	Yes	Yes	No	No	Yes	Yes	No relevant intervention RMI threhold optimisation in a tertiary care setting
Martin Rodriguez S, Ascorbe Salcedo P, Jareno Blanco MS. Diagnostic accuracy of HE4, CA125 and Roma for women with suspected ovarian cancer. *Clin Chem Lab Med* 2015;**53**:S424	Yes	Yes	Unclear	Yes	Unclear	No	No relevant outcomes Insufficient data to determine accuracy measures

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Martra F, Tripodi E, Modaffari P, Zanfagnin V, Fuso L, De Sanso G, *et al.* Ultrasound score versus experienced ultrasound examiner interpretation: are both necessary to improve the management of ovarian masses? *Int J Gynecol Cancer* 2011;**21**(Suppl. 3):385	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention IOTA score (no details)
Meray O, Turkcuoglu I, Meydanli MM, Kafkasli A. Risk of malignancy index is not sensitive in detecting non-epithelial ovarian cancer and borderline ovarian tumor. *J Turkishgerman Gynecol Assoc* 2010;**11**:22–6	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Mills P, Court S, Giamougiannis P, Daines L. Is the risk of malignancy (RMI) score useful in deciding management when below 250? A 2-year retrospective surgical study. *BJOG* 2015;**122**:144–5	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention
Mohammed ABF, Ahuga VK, Taha M. Validation of the Risk of Malignancy Index in primary evaluation of ovarian masses. *Middle East Fertil Soc J* 2014;**19**:324–8	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Study of RMI 3 and 4
Mol BW, Boll D, De Kanter M, Heintz AP, Sijmons EA, Oei SG, *et al.* Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. *Gynecol Oncol* 2001;**80**:162–7	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Validation study of 21 published models (not interventions included in this assessment)
Molina R, Escudero JM, Fuste P. HE-4 levels in gynaecological patients undergoing surgical treatment for suspected malignancies. Systems to increase efficiency. *Tumor Biol* 2014;**35**:S9	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Molina R, Escudero JM, Fuste P. HE-4 utility to increase efficiency in patients with abdominal masses. *Tumor Biol* 2014;**35**:S6	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Moolthiya W, Yuenyao P. The risk of malignancy index (RMI) in diagnosis of ovarian malignancy. *Asian Pac J Cancer Prev* 2009;**10**:865–8	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Moore RG, Jabre-Raughley M, Brown AK, Robison KM, Miller CM, Allard JW, *et al.* Comparison of a novel multiple marker assay versus the risk of malignancy index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;**112**(Suppl. 1):25–6	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention RMI (unspecified threshold)

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Moore RG, Jabre-Raughley M, Brown AK, Robison KM, Miller MC, Allard WJ, *et al.* Comparison of a novel multiple marker assay vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am J Obstet Gynecol* 2010;**203**:228	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Moore EK, Iavazzo C, Argent V, Leung E, Pitkin S, Benton S, *et al.* Does the risk of malignancy algorithm have a role in triaging symptomatic women for further investigation? Results of a pilot 'real world' study. *Int J Gynecol Cancer* 2013;**23**(Suppl. 1):64	Yes	Yes	No	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Moore RG, Hawkins DM, Miller MC, Landrum LM, Gajewski W, Ball JJ, *et al.* Combining clinical assessment and the Risk of Ovarian Malignancy Algorithm for the prediction of ovarian cancer. *Gynecol Oncol* 2014;**135**:547–51	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Moszynski R, Zywica P, Wojtowicz A, Szubert S, Sajdak S, Stachowiak A, *et al.* Menopausal status strongly influences the utility of predictive models in differential diagnosis of ovarian tumors: an external validation of selected diagnostic tools. *Ginekol Pol* 2014;**85**:892–9	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention) The IOTA data are for models other than simple ultrasound rules or ADNEX model
Nahar S, Shamsuddin L, Faruqui M, Ara G. Sonographic prediction of ovarian malignancy in adnexal mass. *Bangladesh J Obstet Gynecol* 2012;**27**:67–71	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention
Numanoglu C, Kuru O, Sakinci M, Akbayir O, Ulker V. Ovarian fibroma/fibrothecoma: retrospective cohort study shows limited value of risk of malignancy index score. *Aust N Z J Obstet Gynaecol* 2013;**53**:287–92	Yes	Yes	Unclear	No	Yes	Yes	RMI at a threshold of 200, data for a small subgroup of patients with fibroma/fibrothecoma

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Ong C, Biswas A, Choolani M, Low JJ. Comparison of risk of malignancy indices in evaluating ovarian masses in a Southeast Asian population. *Singapore Med J* 2013;**54**:136–9	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Data for various RMI thresholds (from 50 to 1000, not including 250)
Ozbay PO, Ekinci T, Caltekin MD, Yilmaz HT, Temur M, Yilmaz O, *et al.* Comparative evaluation of the risk of malignancy index scoring systems (1–4) used in differential diagnosis of adnexal masses. *Asian Pac J Cancer Prev* 2015;**16**:345–9	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 250
Park JY, Park YR, Choe JW, Chun SI, Kim DY, Suh DS, *et al.* Human epididymis secretory protein 4 (HE4) versus cancer antigen 125 (CA125) in the diagnosis of malignant ovarian tumor. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):511	Yes	No	Unclear	Yes	Unclear	No	Diagnostic case–control study
Partheen K, Kristjansdottir B, Sundfeldt K. Evaluation of ovarian cancer biomarkers HE4 and CA-125 in women presenting with a suspicious cystic ovarian mass. *J Gynecol Oncol* 2011;**22**:244–52	Yes	Yes	No	Yes	Yes	Yes	Tertiary care setting gynaecologic oncology surgery
Peces Rama A, Llanos Llanos MC, Sanchez Ferrer ML, Alcazar Zambrano JL, Martinez Mendoza A, Nieto Diaz A. Simple descriptors and simple rules of the International Ovarian Tumour Analysis (IOTA) Group: a prospective study of combined use for the description of adnexal masses. *Eur J Obstet Gynecol Reprod Biol* 2015;**195**:7–11	Yes	No	Yes	Yes	Yes	No	No relevant outcomes Selected population (unclassifiable using IOTA group's simple descriptors)
Pineda L, Salcedo E, Vilhena C, Juez L, Alcazar JL. Interobserver agreement in assigning IOTA colour score to adnexal masses using three-dimensional volumes or digital videoclips: potential implications for training. *Ultrasound Obstet Gynecol* 2014;**44**:361–4	Yes	No	No	No	NA	No	Not a clinical study in patients with suspected ovarian cancer IOTA training study, using video clips
Pitta Dda R, Sarian LO, Barreta A, Campos EA, Andrade LL, Fachini AM, *et al.* Symptoms, CA125 and HE4 for the preoperative prediction of ovarian malignancy in Brazilian women with ovarian masses. *BMC Cancer* 2013;**13**:423	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA score using different manufacturers' assays for CA125 and HE4 measurements (not a valid CE-marked intervention)
Putri I, How JA, Marino J, Villegas R, McNally O, Grover S, *et al.* A 32 year review of clinical presentation and the use of risk of malignancy index (RMI2) in diagnosis of ovarian malignancies in children and adolescents. *Int J Gynecol Cancer* 2014;**24**(Suppl. 4):211–12	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 2

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Ratnavelu N, Founta C, Addison C, Bradbury M, Handley G, Das M, *et al.* The role of adding HE4 to CA125 for women referred to secondary care with suspected ovarian cancer in facilitating management decision making: a prospective pilot study. *Int J Gynecol Cancer* 2014;**24**(Suppl. 4):486–7	Yes	Yes	Unclear	No	No	Yes	No relevant intervention ROMA assays and threshold NR
Raza A, Mould T, Wilson M, Burnell M, Bernhardt L. Increasing the effectiveness of referral of ovarian masses from cancer unit to cancer center by using a higher referral value of the risk of malignancy index. *Int J Gynecol Cancer* 2010;**20**:552–4	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 450
Richards A, Herbst U, Pather S, Saidi S, Tejada-Berges T, Williams P, *et al.* HE4, CA125, the Risk of Malignancy Algorithm (ROMA) and the Risk of Malignancy Index (RMI) and complex pelvic masses – a prospective comparison in the preoperative evaluation of adnexal and pelvic masses in an Australian population. *BJOG* 2015;**122**:150	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Assay details NR
Richards A, Herbst U, Manalang J, Pather S, Saidi S, Tejada-Berges T, *et al.* HE4, CA125, the Risk of Malignancy Algorithm and the Risk of Malignancy Index and complex pelvic masses – a prospective comparison in the pre-operative evaluation of pelvic masses in an Australian population. *Aust Z J Obstet Gynaecol* 2015;**55**:493–7	Yes	Yes	No	Yes	Yes	No	Tertiary care setting
Rogulski L, Strzelczyk J. Simple ultrasound rules used by general gynecologists supplemented with ROMA assessment in differentiating malignant and benign adnexal masses. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):1479	Yes	Yes	Unclear	Yes	Unclear	No	No relevant outcomes
Romagnolo C, Leon AE, Fabricio ASC, Del Pup L, Papadakis C, Odicino FE, *et al.* HE4, CA125 and risk of ovarian malignancy algorithm (ROMA) as diagnostic tools of ovarian cancer in patients with pelvic mass: an Italian multicenter prospective study. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):528–9	Yes	Yes	Unclear	Yes	Unclear	No	No relevant outcomes
Rossi A, Braghin C, Soldano F, Isola M, Capodicasa V, Londero AP, *et al.* A proposal for a new scoring system to evaluate pelvic masses: Pelvic Masses Score (PMS). *Eur J Obstet Gynecol Reprod Biol* 2011;**157**:84–8	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Ruiz de Gauna B, Sanchez P, Pineda L, Utrilla-Layna J, Juez L, Alcazar JL. Interobserver agreement in describing adnexal masses using the International Ovarian Tumor Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. *Ultrasound Obstet Gynecol* 2014;**44**:95–9	Yes	No	No	Yes	No	No	Not a clinical study in patients with suspected ovarian cancer IOTA training study, using video clips
Sandri MT, Bottari F, Franchi D, Boveri S, Candiani M, Ronzoni S, *et al.* Comparison of HE4, CA125 and ROMA algorithm in women with a pelvic mass: correlation with pathological outcome. *Gynecol Oncol* 2013;**128**:233–8	Yes	Yes	Yes	Yes	Yes	No	No relevant outcomes Data for specificity at a fixed sensitivity
Sayasneh A, Kaijser J, Preisler J, Johnson S, Stalder C, Husicka R, *et al.* A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. *Gynecol Oncol* 2013;**130**:140–6	Yes	No	Unclear	No	Yes	Yes	No relevant intervention No data for IOTA group's simple ultrasound rules alone Selected population (unclassifiable using the IOTA group's simple descriptors)
Sayasneh A, Preisler J, Stlader C, Husicka R, Naji O, Kaijser J, *et al.* A randomised controlled trial to compare the clinical impact of RMI versus LR2 to characterise adnexal masses: interim analysis of phase 4 IOTA study. *BJOG* 2013;**120**:357–358	Yes	Yes	No	No	NA	Yes	No relevant intervention IOTA group's regression model (not simple ultrasound rules or ADNEX model)
Sayasneh A, Kaijser J, Preisler J, Smith AA, Raslan F, Johnson S, *et al.* Accuracy of ultrasonography performed by examiners with varied training and experience in predicting specific pathology of adnexal masses. *Ultrasound Obstet Gynecol* 2015;**45**:605–12	Yes	Yes	No	No	Yes	Yes	No relevant intervention
Senel SA, Ozcam H, Ateser GB, Vatansever D. Risk of malignancy indices in differentiation of malignant adnexal masses from the benign adnexal masses. *Int J Gynecol Cancer* 2015;**25**(Suppl. 1):1006	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Accuracy of RMI 4
Shimada K, Matsumoto K, Mimura T, Ishikawa T, Hirose Y, Shimizu H, *et al.* Ultrasound-based logistic regression modelling versus magnetic resonance imaging for discriminating between benign and malignant adnexal masses: a prospective study. *Int J Gynecol Cancer* 2016;**26**:820	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA group's regression model (not simple ultrasound rules or ADNEX model)

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Simsek HS, Tokmak A, Ozgu E, Doganay M, Danisman N, Erkaya S, *et al.* Role of a risk of malignancy index in clinical approaches to adnexal masses. *Asian Pac J Cancer Prev* 2014;**15**:7793–7	Yes	Yes	No	No	Yes	Yes	No relevant intervention Optimised RMI threshold (163.5) in a tertiary care setting
Simsek S, Tokmak A, Ozgu E, Doganay M, Danisman N, Erkaya S, *et al.* The role of risk of malignancy index (RMI) in clinical approach to adnexial masses. *Int J Gynecol Cancer* 2014;**24**(Suppl. 4):348	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Optimised RMI threshold (163.85)
Sladkevicius P, Valentin L. Intra- and interobserver agreement when describing adnexal masses using the International Ovarian Tumour Analysis terms and definitions: a study on three-dimensional ultrasound volumes. *Ultrasound Obstet Gynaecol* 2013;**41**:318–27	Yes	Yes	Unclear	No	Yes	No	No relevant intervention IOTA group models (not simple ultrasound rules or ADNEX model)
Sole-Sedeno J, Agramunt S, Mancebo G, Rueda C, Sastre M, Alameda F, *et al.* Risk malignancy index in the evaluation of the adnexal masses. *Int J Gynecol Cancer* 2012;**22**:E967–8	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Unspecified RMI threshold
Stiekema A, Van De Vrie R, Lok C, Van Driel W, Korse T, Buist M, *et al.* Serum HE4 as additional step to the RMI 1 improves the diagnosis of patients with a pelvic mass. *Int J Gynecol Cancer* 2016;**26**:169	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200 and HE4
Sumpaico WW. Comparison of ROMA to RMI for ovarian carcinoma in Asia. *Int J Gynaecol Obstet* 2012;**119**:S248–9	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold NR
Tanriverdi HA, Sade H, Akbulut V, Barut A, Bayar U. [Clinical and ultrasonographic evaluation of pelvic masses.] *J Turk German Gynecol Assoc* 2007;**8**:67–70	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Terzic M, Dotlic J, Ladjevic IL, Atanackovic J, Ladjevic N. Evaluation of the risk malignancy index diagnostic value in patients with adnexal masses. *Vojnosanit Pregl* 2011;**68**:589–93	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Terzic M, Dotlic J, Likic I, Brndusic N, Pilic I, Ladjevic N, *et al.* Risk of malignancy index validity assessment in premenopausal and postmenopausal women with adnexal tumours. *Taiwan J Obstet Gynecol* 2013;**52**:253–7	Yes	Yes	Yes	No	Yes	No	No relevant intervention
Thompson R, Dempsey A, Abdel-Aty M. Which risk of malignancy index (RMI) calculation is a better predictor of malignancy, and at what level should we refer to the cancer centre? A retrospective observational study conducted at East Lancashire Hospitals NHS Trust. *BJOG* 2014;**121**:9	Yes	No	Unclear	Yes	No	Yes	Diagnostic case–control study
Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I, *et al.* Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound Obstet Gynecol* 1999;**13**:17–25	Yes	Yes	Unclear	No	Unclear	Accuracy	No relevant intervention
Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, *et al.* Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. *J Clin Oncol* 2005;**23**:8794–801	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA group models (not simple ultrasound rules or ADNEX model)
Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, *et al.* Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumours. *J Clin Oncol* 2007;**25**:4194–200	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA group model (not simple ultrasound rules or ADNEX model)
Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, *et al.* Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010;**36**:226–34	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Validation of other IOTA group's models (not simple ultrasound rules or ADNEX model)
Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, *et al.* Predicting the risk of malignancy in adnexal masses based on the simple rules from the International Ovarian Tumour Analysis group. *Am J Obstet Gynecol* 2016;**214**:424–37	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Development and validation of the IOTA group's simple ultrasound rules risk model
Tingulstad S, Hagen B, Skjeldestad FE, Halvorsen T, Nustad K, Onsrud M. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. *Obstet Gynecol* 1999;**93**:448–52	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention RMI 2

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Toledo KL, Audifred JR, Topete RE, Niebla DC, Hernandez SE, Morales L. Comparison between histopathological results and malignancy index risk in adnexal complex cysts treated by laparoscopic surgery. *J Minim Invasive Gynecol* 2016;**23**(Suppl. 1):217–18	Yes	Yes	Yes	No	Yes	No	No relevant outcomes
Torres JC, Derchain SF, Faundes A, Gontijo RC, Martinez EZ, Andrade LA. Risk-of-malignancy index in preoperative evaluation of clinically restricted ovarian cancer. *São Paulo Med J* 2002;**120**:72–6	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Not RMI 1
Trevino-Baez JD, Cantu-Cruz JA, Medina-Mercado J, Abundis A. [Diagnostic accuracy of malignancy risk index II in post-menopausal women with adnexal tumour.] *Cir Cir* 2016;**84**:109–14	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Study of RMI 2
University of South F, Universitaire Ziekenhuizen L. *International Ovarian Tumour Analysis (IOTA) Phase 5*. URL: https://ClinicalTrials.gov/show/NCT01698632 (accessed 5 July 2018)	No	Yes	Unclear	No	NA	No	Trial registry entry
Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. *Ultrasound Obstet Gynecol* 2001;**18**:357–65	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention
Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, *et al.* Adnexal masses difficult to classify as benign or malignant using subjective assessment of grey-scale and Doppler ultrasound findings: logistic regression models do not help. *Ultrasound Obstet Gynecol* 2011;**38**:456–65	Yes	No	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200, for masses unclassifiable by CA125
Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, *et al.* Unilocular adnexal cysts with papillary projections but no other solid components: is there a diagnostic method that can classify them reliably as benign or malignant before surgery? *Ultrasound Obstet Gynecol* 2013;**41**:570–81	Yes	Yes	Unclear	No	Yes	No	No relevant intervention Development of an IOTA group model to predict malignancy in unilocular cysts with papillations
Van Calster B, Timmerman D, Bourne T, Testa AC, Van Holsbeke C, Domali E, *et al.* Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007;**99**:1706–14	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention

continued

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Van Calster B, Timmerman D, Valentin L, McIndoe A, Ghaem-Maghami S, Testa AC, *et al.* Triaging women with ovarian masses for surgery: observational diagnostic study to compare RCOG guidelines with an International Ovarian Tumour Analysis (IOTA) group protocol. *BJOG* 2012;**119**:662–71	Yes	Yes	Yes	No	Yes	No	No relevant intervention IOTA group model (not simple ultrasound rules or ADNEX model)
van den Akker PA, Aalders AL, Snijders MP, Kluivers KB, Samlal RA, Vollebergh JH, *et al.* Evaluation of the Risk of Malignancy Index in daily clinical management of adnexal masses. *Gynecol Oncol* 2010;**116**:384–8	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Van Holsbeke C, Van Calster B, Valentin L, Testa AC, Ferrazzi E, Dimou I, *et al.* External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. *Clin Cancer Res* 2007;**13**:4440–7	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, *et al.* Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumour analysis study. *Clin Cancer Res* 2009;**15**:684–91	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Validation of IOTA group models (not simple ultrasound rules or ADNEX model)
Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, *et al.* External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. *Clin Cancer Res* 2012;**18**:815–25	Yes	Yes	Unclear	No	Yes	No	No relevant intervention Accuracy of RMI at a threshold of 200
Villiotou V, Vorgias G, Lekka I, Karampelas A, Dertimas V. Evaluation of HE4, CA 125 and ROMA predictive index in patients with gynaecological diseases. *Clin Chem Lab Med* 2014;**52**:S479	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA assays and threshold NR
Wang LM, Song H, Song X, Zhou XB. An improved risk of malignancy index in diagnosis of adnexal mass. *Chin Med J* 2012;**125**:533–5	Yes	Yes	Yes	No	Yes	No	No relevant intervention Unspecified RMI threshold
Wilailak S, Chan KK, Chen CA, Nam JH, Ochiai K, Aw TC, *et al.* Distinguishing benign from malignant pelvic mass utilizing an algorithm with HE4, menopausal status, and ultrasound findings. *Journal of Gynecologic Oncology* 2015;**26**:46–53	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention

Study	Criteria						Reason for exclusion
	Primary study	Population	Setting	Index test	Reference standard	Outcome	
Winarto H, Bismarck JL, Purbadi S, Nuranna L. Is ROMA scoring systems really better than RMI for indonesian patients, in DR. ciptomangunkusumo hospital. *Int J Gynecol Cancer* 2011;**21**(Suppl. 3):S403	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention
Yamamoto Y, Tsuchida A, Ushiwaka T, Nagai R, Matsumoto M, Komatsu J, *et al.* Comparison of 4 risk-of-malignancy indexes in the preoperative evaluation of patients with pelvic masses: a prospective study. *Clin Ovarian Other Gynecol Cancer* 2014;**7**:8–12	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200
Yavuzcan A, Caglar M, Ozgu E, Ustun Y, Dilbaz S, Ozdemir I, *et al.* Should cut-off values of the risk of malignancy index be changed for evaluation of adnexal masses in Asian and Pacific populations? *Asian Pac J Cancer Prev* 2013;**14**:5455–9	Yes	No	Unclear	No	Yes	Yes	Not patients with suspected ovarian cancer
Yazbek J, Aslam N, Tailor A, Hillaby K, Raju KS, Jurkovic D. A comparative study of the risk of malignancy index and the ovarian crescent sign for the diagnosis of invasive ovarian cancer. *Ultrasound Obstet Gynecol* 2006;**28**:320–4	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI at a threshold of 200 in a tertiary care setting
Yoshida A, Derchain SF, Pitta DR, Andrade LA, Sarian LO. Comparing the Copenhagen Index (CPH-I) and Risk of Ovarian Malignancy Algorithm (ROMA): two equivalent ways to differentiate malignant from benign ovarian tumors before surgery? *Gynecol Oncol* 2016;**140**:481–5	Yes	Yes	No	Yes	Yes	No	Tertiary care setting
Zannoni L, Savelli L, Jokubkiene L, Di Legge A, Condous G, Testa AC, *et al.* Intra- and interobserver agreement with regard to describing adnexal masses using International Ovarian Tumour Analysis terminology: reproducibility study involving seven observers. *Ultrasound Obstet Gynecol* 2014;**44**:100–8	No	No	No	No	NA	No	Not a clinical study in patients with suspected ovarian cancer IOTA group's training study, using video clips
Zhang S. Performance of ovarian malignancy algorithm in predicting pelvic mass in patients at risk of ovarian cancer. *Chin J Clin Oncol* 2014;**41**:513–17	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA assays and threshold NR

NA, not applicable; NR, not reported.

Appendix 3 Example assessments of study quality

Example QUADAS-2 assessment

Van Calster et al. *(2014)*[17]

Domain 1: patient selection

A. Risk of bias	
Consecutive patients with at least one adnexal mass selected for surgical intervention, referred for IOTA group phase 3 study	
Was a consecutive or random sample of patients enrolled?	Yes
Was a case–control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	**Risk: low**
B. Applicability	
Women referred for evaluation of an adnexal mass. Secondary or tertiary care referral, but ADNEX model includes a term for type of referral centre	
Do the included patients match the question?	**Concerns: low**

Domain 2: index test(s)

A. Risk of bias	
ADNEX validation data set. No details regarding who performed tests, whether or not they were blind, or when they were performed	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it prespecified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	**Risk: low**
B. Applicability	
Are there concerns that the index test, its conduct or interpretation differ from the review question?	**Concerns: low**

Domain 3: reference standard

A. Risk of bias	
Histology of resected mass (no further details). Performed without knowledge of ultrasound	
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk: low**
B. Applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns: high**

Domain 4: flow and timing

A. Risk of bias	
Calculation of 2 × 2 data from reported sensitivity and specificity values resulted in non-whole numbers for some analyses. The time from index test to surgery was ≤ 120 days	
Was there an appropriate time interval between the index test and reference standard?	Yes
Did all patients receive a reference standard?	Yes
For comparative accuracy studies, did all patients receive all index tests?	NA
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	**Risk: unclear**

NA, not applicable.

Example PROBAST assessment

Van Calster et al. *(2014)*[17]

Domain 1: participant selection

A. Risk of bias

Describe the sources of data and criteria for participant selection:

Data were derived from an international, multicentre, prospective cohort study (the IOTA study) of consecutive women with at least one adnexal mass that was clinically judged to require surgery. Participants were excluded if they refused transvaginal ultrasonography, were pregnant at the time of presentation or received surgery > 120 days after the ultrasound examination. The IOTA group was established to develop and validate diagnostic models for adnexal masses, based on large multicentre data sets, using a standardised ultrasound examination protocol, terms and definitions

The ADNEX model was developed using data collected in IOTA study phases 1, 1b and 2 (1999–2007) and validated using data collected in phase 3 (2009–12); inclusion criteria remained the same throughout

		Dev	Val
1.1 Were participant selection criteria similar to the model development study?			Yes
1.2 Were appropriate data sources used (e.g. cohort, RCT or nested case–control study data?)		Yes	Yes
1.3 Were all inclusions and exclusions of participants appropriate?		Yes	Yes
Risk of bias introduced by selection of participants	**Risk:** *(low/high/unclear)*	**Low**	**Low**

Rationale of bias rating:

B. Applicability

Describe included participants, setting and dates:

Women with at least one adnexal mass requiring surgery. Women were evaluated in a mixture of secondary care settings and gynaecological oncology tertiary referral centres

Concern that the included participants and setting do not match the review question	**Concern:** *(low/high/unclear)*	**High**	**High**

Rationale of applicability rating:

The study setting is not a complete match for that specified in the scope for this assessment

Dev, model development study; Val, model validation study.
Note
Shading indicates that this particular question should not be applied.

Domain 2: predictors

A. Risk of bias

List and describe predictors included in the final model (e.g. definition and timing of assessment):

Age, serum CA125 level (log-transformed), type of centre (tertiary referral gynaecological oncology centre vs. other centres), maximum diameter of the lesion (log-transformed), proportion of solid tissue (with quadratic term), number of papillary projections, > 10 cyst locules, acoustic shadows and ascites were included in the final ADNEX model. Family history of ovarian cancer was dropped by the variable selection analysis. Predictors were assessed prior to surgery and histological evaluation. Participating centres used one of four manufacturers' immunoradiometric assay kits to measure CA125; all kits used the OC125 antibody

		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		No	No
2.2 Were predictors defined and assessed in a similar way to predictors in the development model?			Yes
2.3 Were predictor assessments made without knowledge of outcome data?		Yes	Yes
2.4 Are all predictors available at the time the model is intended to be used?		Yes	Yes
Risk of bias introduced by predictors or their assessment	**Risk:** *(low/high/unclear)*	Low	Low

Rationale of bias rating:

Study centres used different CA125 assays; however, all assays used the same antibody and, therefore, the effects of this variation are likely to be minimal

B. Applicability

Concern that the definition, assessment or timing of predictors in the model do not match the review question	**Concern:** *(low/high/unclear)*	**Low**	**Low**

Rationale of applicability rating:

The inclusion of CA125 assays from a variety of manufacturers reflects the reality of clinical practice

Dev, model development study; Val, model validation study.
Note
Shading indicates that this particular question should not be applied.

Domain 3: outcome

A. Risk of bias

Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:

The outcome was determined by histopathological analysis of the mass after surgical removal by laparotomy or laparoscopy (as considered appropriate by the surgeon). The stage of malignant tumours was recorded using the FIGO classification system. Excised tissue was examined locally at each study centre. The histological classification was performed without knowledge of the ultrasound results, but it was not clear whether or not the pathologists were aware of other predictor information. The final diagnosis was divided into five types: benign, borderline, stage I invasive, stages II–IV invasive and secondary metastatic cancer

	Dev	Val
3.1 Was the outcome determined appropriately?	Yes	Yes
3.2 Was a prespecified or standard outcome definition used?	Yes	Yes
3.3 Were predictors excluded from the outcome definition?	Yes	Yes
3.4 Was the outcome defined and determined in a similar way for all participants?	Yes	Yes
3.5 Was the outcome defined and determined in a similar way to the outcome in the model development study?		Yes
3.6 Was the outcome determined without knowledge of predictor information?	Unclear	Unclear
3.7 Was the time interval between predictor assessment and outcome determination appropriate?	Yes	Yes

Risk of bias introduced by the outcome or its determination	**Risk:** *(low/high/unclear)*	**Unclear**	**Unclear**

Rationale of bias rating:

It was not clear whether or not pathologists were blinded to the CA125 results

B. Applicability

At what time point was the outcome determined:

All surgery was performed within 120 days of ultrasound examination

If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:

NA

Concern that the outcome, its definition, timing or determination do not match the review question	**Concern:** *(low/high/unclear)*	**Low**	**Low**

Rationale of applicability rating:

Dev, model development study; NA, not applicable; Val, model validation study.
Note
Shading indicates that this particular question should not be applied.

Domain 4: analysis

Risk of bias

Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:

The development data set included 3506 women and the validation data set included 2403 women. There were 10 candidate predictors. The development data set included 949 (27%) women with malignancies (including borderline tumours) and the validation data set included 980 (41%) women with malignancies (including borderline tumours)

Describe how the model was developed (predictor selection, optimism, risk groups, model performance):

To avoid overfitting, 10 candidate predictors were selected; selection was based on topic expertise and stability of predictors across centres. Furthermore, data-driven selection used a method based on multivariable fractional polynomials; the variable selection procedure is a variant of the standard backward selection procedure. Age and type of centre were forced into the model

To acknowledge the variability between centres, multinomial logistic regression with random centre intercepts was used to construct a polytomous model. Predictor coefficients were multiplied with uniform shrinkage factors to avoid exaggerated model coefficients

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross-validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):

The model was validated using data collected, by the same criteria, in a later phase of the IOTA study (temporal validation). Discriminatory performance was assessed using diagnostic accuracy measures, with histological diagnosis as the reference standard and by calculating a polytomous discrimination index. Calibration of predicted probabilities was assessed using calibration plots showing the relationship between predicted and observed probabilities for each type of tumour. The plots were based on a parametric, multinomial logistic *n* calibration analysis, using random centre intercepts

Describe the performance measures of the model [e.g. (re)calibration, discrimination, (re)classification, net benefit]:

Discrimination measures and calibration plots were reported

Describe any participants who were excluded from the analysis:

All participants who met the study inclusion criteria appear to have been included in the analysis

Describe missing data on predictors and outcomes as well as methods used for missing data:

The CA125 data were missing for 31% of participants. Predictive mean-matching regression, using variables that were related to either the level of CA125 itself or to the unavailability of CA125 (i.e. a binary indicator indicating for each woman whether or not CA125 was missing) was used to estimate missing values. This was repeated 100 times to generate multiple imputations of the missing values, resulting in 100 completed data sets

	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?	Yes	Yes
4.2 Were continuous and categorical predictors handled appropriately?	Yes	
4.3 Were all enrolled participants included in the analysis?	Yes	
4.4 Were participants with missing data handled appropriately?	Yes	
4.5 Was selection of predictors based on univariable analysis avoided?	Yes	
4.6 Were important complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	Yes	Yes
4.7 Were relevant model performance measures evaluated (e.g. calibration and discrimination)?	Yes	Yes
4.8 Was model overfitting and optimism in model performance accounted for?	Unclear	
4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Unclear	
Risk of bias introduced by the analysis **Risk:** *(low/high/unclear)*	**Unclear**	**Low**

Rationale of bias rating:

Some aspects of model development were not fully reported

Dev, model development study; Val, model validation study.
Note
Shading indicates that this particular question should not be applied.

Overall judgement about risk of bias and applicability of the prediction model evaluation

Overall judgement of risk of bias	**Risk:** *(low/high/unclear)*	**Unclear**

Summary of sources of potential bias:

Some aspects of model development were not fully reported

Overall judgement of applicability	**Concern:** *(low/high/unclear)*	**Low**

Summary of applicability concerns:

The study setting is not a complete match for that specified in the scope for this assessment; however, the final ADNEX model includes a variable for centre type (general secondary care vs. tertiary referral gynaecological oncology setting); the model should therefore be usable in either setting

Appendix 4 Full study details

TABLE 33 Study details and baseline participant characteristics

Study details	Selection criteria	Participant details	Test(s)
Abdalla *et al.* (2013)[48] Country: Poland Funding: NR Recruitment start–end: January 2011–December 2011	Inclusion criteria: women admitted with adnexal mass Exclusion criteria: ultrasound examination > 90 days before surgery; no CA125 level Study setting: mixed Point in care pathway at which index test is given: following referral to the Department of Clinical Obstetrics, Women's Diseases and Gynaecological Oncology, in a university hospital	*All* • Number tested: 87 • Age (years): mean 44.5 (SD 16.6); range 17–79 • % premenopausal : % postmenopausal: 60.9 : 39.1 • Definition of postmenopausal: amenorrhoea for ≥ 1 year and aged ≥ 50 years in patients with a history of hysterectomy • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules, and RMI 1
Aktürk *et al.* (2011)[71] Country: Turkey Funding: NR Recruitment start–end: October 2008–February 2010	Inclusion criteria: women with pelvic masses scheduled for laparotomy or laparoscopy Exclusion criteria: NR Study setting: secondary care (Department of Obstetrics and Gynaecology) Point in care pathway at which index test is given: following referral to secondary care (Department of Obstetrics and Gynaecology)	*Benign* • Number tested: 80 • Age (years): NR • % premenopausal : % postmenopausal: 80 : 20 • Definition of postmenopausal: > 1 year of amenorrhoea or an age of > 50 years in women who have had a hysterectomy • CA125 (U/ml): mean 28 (SD 23.8); range 3–120 • HE4 (pmol/l): NR *Malignant* • Number tested: 20 • Age (years): NR • % premenopausal : % postmenopausal: 45 : 55 • Definition of postmenopausal: > 1 year of amenorrhoea or an age of > 50 years in women who have had a hysterectomy • CA125 (U/ml): mean 329.2 (SD 648); range 12–2821 • HE4 (pmol/l): NR	RMI 1 threshold comparison

Study details	Selection criteria	Participant details	Test(s)
Al Musalhi *et al.* (2016)[103] Country: Oman Funding: other (unfunded) Recruitment start–end: March 2014–April 2015	Inclusion criteria: women attending a gynaecology department for investigation of an ovarian mass Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to a gynaecology department	*Benign* • Number tested: 165 • Age (years): median 33; range 13–80 • % premenopausal : % postmenopausal: 85 : 15 • Definition of postmenopausal: previous hysterectomy and aged ≥ 50 years, or amenorrhoea for ≥ 1 year • CA125 (U/ml): median 23; range 1–978 • HE4 (pmol/l): median 43; range 18–2677 *Malignant* • Number tested: 48 • Age (years): median 55; range 21–83 • % premenopausal : % postmenopausal: 44 : 56 • Definition of postmenopausal: previous hysterectomy and aged ≥ 50 years, or amenorrhoea for ≥ 1 year • CA125 (U/ml): median 261; range 7–14507 • HE4 (pmol/l): median 207; range 27–5932	ROMA score using Abbott Diagnostics' tumour marker assay and RMI 1
Alcázar *et al.* (2013)[52] Country: Spain Funding: NR Recruitment start–end: January 2011–June 2012	Inclusion criteria: women with an adnexal mass, referred to one of two Spanish university centres (Clinica Universidad de Navarra, Pamplona or Spain and Institut Dexeus, Barcelona) Exclusion criteria: pregnancy, spontaneous resolution of the mass by the time of a 2- to 3-month follow-up scan, surgery not performed because of physician's and/or patient's decision at follow-up, or surgery performed in another centre Study setting: secondary care (Department of Obstetrics and Gynaecology) Point in care pathway at which index test is given: following referral to secondary care	*All* • Number tested: 340 • Age (years): mean 42.1 (SD 13.2); range 13–79 • % premenopausal : % postmenopausal: 77.1 : 22.9 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Asif *et al.* (2004)[77] Country: Pakistan Funding: NR Recruitment start–end: January 2001–January 2002	Inclusion criteria: consecutive women admitted to the Department of Gynaecology and Obstetrics (Military Hospital and Combined Military Hospital Rawalpindi) for elective surgical exploration and resection of proven ovarian mass Exclusion criteria: NR Study setting: unclear Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 55 • Age (years): mean 45 (SD 11) • % premenopausal : % postmenopausal: 40 : 60 • Definition of postmenopausal: ≥ 1 year of amenorrhoea • CA125 (U/ml): mean 1107, SD NR; range 15–1107 • HE4 (pmol/l): NR *Benign* • Number tested: 45 • Age (years): mean 37 (SD 14) • % premenopausal : % postmenopausal: 75 : 25 • Definition of postmenopausal: ≥ 1 year of amenorrhoea • CA125 (U/ml): mean 26.5; range 2–210 • Median HE4 (pmol/l): NR	RMI 1 threshold comparison
Baker *et al.* (2013)[66] Country: UK Funding: NR Recruitment start–end: NR	Inclusion criteria: premenopausal women with ovarian masses Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to a district general hospital	*All* • Number tested: 48 • Age (years): NR • % premenopausal : % postmenopausal: 100 : 0 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

Study details	Selection criteria	Participant details	Test(s)
Chan *et al.* (2013)[82] Country: Hong Kong, Taiwan, Republic of Korea, Japan, Thailand and the Philippines Funding: industry (Abbott Diagnostics) Recruitment start–end: NR 2009–NR 2010	Inclusion criteria: consecutive women (aged > 18 years) diagnosed with an adnexal mass by ultrasound scan, CT, PET or MRI scan Exclusion criteria: previous history of ovarian cancer, primary peritoneal or any known malignancy; or previous bilateral oophorectomy Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to one of six obstetrics and gynaecology departments	*All* • Number tested: 414 • Age (years): NR • % premenopausal : % postmenopausal: 73.9 : 26.1 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Abbott Diagnostics' tumour marker assay
Clemente and Benitez (2015)[90] Country: the Philippines Funding: NR Recruitment start–end: October 2010–December 2013	Inclusion criteria: women with an adnexal mass who underwent surgery Exclusion criteria: NR Study setting: unclear Point in care pathway at which index test is given: following referral to a tertiary care hospital (unclear whether or not referral was to a specialist gynaecological oncology department)	*All* • Number tested: 62 • Age (years): median NR; range 22–79 • % premenopausal : % postmenopausal: 77.4 : 22.6 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Abbott Diagnostics' tumour marker assay

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Coleman *et al.* (2016)[70] Country: USA Funding: industry (Vermillion Inc.) Recruitment start–end: August 2010–December 2011	Inclusion criteria: women ≥ 18 years with a documented pelvic mass who were scheduled for surgical intervention within 3 months of imaging, and who agreed to phlebotomy Exclusion criteria: diagnosis of malignancy in the previous 5 years (except of non-melanoma skin cancers) or enrolment by a gynaecologic oncologist Study setting: secondary care Point in care pathway at which index test is given: following referral to secondary care	*All* • Number tested: 493 • Age (years): median 48; range 18–87 • % premenopausal : % postmenopausal: 56.0 : 44.0 • Definition of postmenopausal: absence of menses for ≥ 12 months or aged ≥ 50 years • CA125 (U/ml): NR • HE4 (pmol/l): NR	Overa (MIA2G)
Davies *et al.* (1993)[79] Country: UK Funding: NR Recruitment start–end: NR	Inclusion criteria: retrospective review of women admitted consecutively to a gynaecology department for surgical investigation of an adnexal mass Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 37 • Age (years): NR • % premenopausal : % postmenopausal: 83.8 : 16.2 • Definition of postmenopausal: ≥ 1 year of amenorrhoea or aged > 50 years for women who had previously undergone a hysterectomy • CA125 (U/ml): median 173; range 5–1405 • HE4 (pmol/l): NR *Benign* • Number tested: 87 • Age (years): NR • % premenopausal : % postmenopausal: 53 : 47 • Definition of postmenopausal: ≥ 1 year of amenorrhoea or aged > 50 years for women who had previously undergone a hysterectomy • CA125 (U/ml): median 18; range 5–760 • HE4 (pmol/l): NR	RMI 1 threshold comparison

Study details	Selection criteria	Participant details	Test(s)
Di Legge *et al.* (2012)[61] Country: Sweden, Belgium, Italy, Poland, UK, Czech Republic, China and Canada Funding: government (Swedish Research Council and the Research Foundation of Flanders) Recruitment start–end: NR	Inclusion criteria: women with an adnexal mass recruited from 11 oncology referral centes, five general hospitals and three referral centres for ultrasonography Exclusion criteria: surgical removal of the mass > 120 days after ultrasound, pregnancy or inability to tollerate transvaginal ultrasonography Study setting: mixed Point in care pathway at which index test is given: following referral to secondary or tertiary care	*Tumour size ≤ 4 cm* • Number tested: 396 • Age (years): median 42; range 15–87 • % premenopausal : % postmenopausal: 71 : 29 • Definition of postmenopausal: NR • CA125 (U/ml): median 21; range 3–9814 • HE4 (pmol/l): NR *Tumour size 4–9.9 cm* • Number tested: 1457 • Age (years): median 43; range 9–89 • % premenopausal : % postmenopausal: 65 : 35 • Definition of postmenopausal: NR • CA125 (U/ml): median 23; range 2–38161 • Median HE4 (pmol/l): NR *Tumour size ≥ 10 cm* • Number tested: 592 • Age (years): median 53; range 15–94 • % premenopausal : % postmenopausal: 45 : 55 • Definition of postmenopausal: NR • CA125 (U/ml): median 58; range 2–40140 • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules; and RMI 1

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Fathallah *et al.* (2011)[63] Country: France Funding: NR Recruitment start–end: January 2002–December 2005	Inclusion criteria: women who had undergone surgery and histological analysis, following observation of at least one persistent ovarian cyst on two consecutive ultrasound examinations Exclusion criteria: NR Study setting: mixed Point in care pathway at which index test is given: following referral to secondary or tertiary care	*All* • Number tested: 109 masses • Age (years): mean 45.5; range 21–76 • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: NR • Age (years): mean 51; range 22–75 • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Benign* • Number tested: NR • Age (years): mean 41.5; range 21–76 • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules
IOTA5 2017 Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed

Study details	Selection criteria	Participant details	Test(s)
Jacobs *et al.* (1990)[78] Country: UK Funding: charity (Gynaecology Cancer Research Fund, Cancer Research Campaign) Recruitment start–end: NR	Inclusion criteria: women admitted consecutively for surgical investigation of an adnexal mass Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 42 • Age (years): mean 59 (SD 11.8) • % premenopausal : % postmenopausal: 19.5 : 80.5 • Definition of postmenopausal: ≥ 1 year of amenorrhoea or aged > 50 years for women who had previously undergone a hysterectomy • CA125 (U/ml): median 122 (SD NR), 95% CI 6.1 to 3394.8 • HE4 (pmol/l): NR *Benign* • Number tested: 101 • Age (years): mean 48.8 (SD 14.3) • % premenopausal : % postmenopausal: 52 : 48 • Definition of postmenopausal: ≥ 1 year of amenorrhoea or aged > 50 years for women who had previously undergone a hysterectomy • CA125 (U/ml): median 17.5 (SD NR), 95% CI 4.3 to 70.2 • HE4 (pmol/l): NR	RMI 1 threshold comparison
Janas *et al.* (2015)[97] Country: Poland Funding: government (MNISW) Recruitment start–end: NR 2011–NR 2014	Inclusion criteria: women referred for surgery for an adnexal mass Exclusion criteria: NR Study setting: mixed Point in care pathway at which index test is given: following referral to gynaecology or gynaecological oncology clinic	*All* • Number tested: 259 • Age (years): NR • % premenopausal : % postmenopausal: 51 : 49 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Roche Diagnostics' tumour marker assay
Joyeux *et al.* (2016)[43] Country: France Funding: NR	Inclusion criteria: women aged 14–100 years, received or referred with an adnexal mass (detected on ultrasound) requiring surgery	*Malignant* • Number tested: • Age (years): mean 57.5 (SD 3.7) • % premenopausal : % postmenopausal: NR	ADNEX model

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Recruitment start–end: January 2013–December 2015	Exclusion criteria: the absence of TVS, pregnancy, an echographic aspect of functional ovarian cyst or the lack of a CA125 level Study setting: secondary care (Department of Obstetrics and Gynaecology or Gynaecological Surgery) Point in care pathway at which index test is given: following referral to secondary care (Department of Obstetrics and Gynaecology)	• Definition of postmenopausal: NR • CA125 (U/ml): mean 653.6 (SD 321) • HE4 (pmol/l): NR *Benign* • Number tested: • Age (years): mean 50.3 (SD 16) • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): mean 21.4 (SD 34.9) • HE4 (pmol/l): NR	
Karlsen *et al.* (2012)[83] Country: Denmark Funding: industry (Abbott Diagnostics provided assay reagents) Recruitment start–end: September 2004–January 2010	Inclusion criteria: women admitted to the gynaecology clinic for surgery because of a pelvic mass or pelvic pains potentially caused by a malignant disease or endometriosis Exclusion criteria: preoperative-known relapse of a previous cancer or active cancer other than ovarian cancer Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to a gynaecology clinic	*All* • Number tested: 1218 • Age (years): median 51; range 16–90 • % premenopausal : % postmenopausal: 49 : 51 • Definition of postmenopausal: previous hysterectomy and aged ≥ 50 years, or amenorrhoea for ≥ 1 year • CA125 (U/ml): NR • Median HE4 (pmol/l): NR *Benign* • Number tested: NR • Age (years): NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): median 28.7; range 3–3586 • HE4 (pmol/l): median 53.4; range 19–1426 *Malignant* • Number tested: NR • Age (years): NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): median 647; range 10–10,000 • HE4 (pmol/l): median 436; range 16–15,000	ROMA score using Abbott Diagnostics' tumour marker assay; and RMI 1

Study details	Selection criteria	Participant details	Test(s)
Knafel *et al.* (2015)[49] Country: Poland Funding: NR Recruitment start–end: January 2011–October 2012	Inclusion criteria: women, aged ≥ 18 years with an adnexal tumour requiring surgery Exclusion criteria: pregnancy, lack of a histopathology result or surgery performed > 90 days after diagnosis Study setting: unclear Point in care pathway at which index test is given: following referral to a university hospital, Department of Oncology and Gynaecology	*All* • Number tested: 226 • Age (years): mean 47 (SD NR) • % premenopausal : % postmenopausal: 63.3 : 36.7 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules
Langhe *et al.* (2013)[94] Country: NR Funding: NR Recruitment start–end: NR	Inclusion criteria: women scheduled for surgery for invasive, borderline and benign ovarian disease Exclusion criteria: NR Study setting: unclear Point in care pathway at which index test is given: following referral to hospital	*All* • Number tested: 223 • Age (years): median 56 (SD NR) • % premenopausal : % postmenopausal: 35 : 65 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Fujirebio Diagnostics' tumour marker assay
Li *et al.* (2016)[96] Country: China Funding: government (Specialised Research Fund for the Doctoral Programme of Higher Education of China; Science and Technology Department, Guangdong province; Natural Science Foundation, Guangdong province; and the Science and Technology Department of Guangzhou City) Recruitment start–end: September 2012–April 2014	Inclusion criteria: women with gynaecological diseases, diagnosed by ultrasound, CT, PET-CT or MRI Exclusion criteria: previous or concomitant history of malignant disease; bilateral oophorectomy Study setting: unclear Point in care pathway at which index test is given: following referral to a university hospital	*All* • Number tested: 917 • Age (years): NR • % premenopausal : % postmenopausal: 81.2 : 18.8 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: 190 • Age (years): median 50 (SD NR); range 18–82 • % premenopausal : % postmenopausal: 56.8 : 43.2 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Abbott Diagnostics' tumour marker assay

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
		Benign • Number tested: 727 • Age (years): NR • % premenopausal : % postmenopausal: 87.6 : 12.4 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	
Lou *et al.* (2010)[73] Country: China Funding: NR Recruitment start–end: June 2008–December 2008	Inclusion criteria: women with an adnexal mass Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to an obstetrics and gynaecology department	*All* • Number tested: 223 • Age (years): NR • % premenopausal : % postmenopausal: 74 : 26 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: 61 • Age (years): NR • % premenopausal : % postmenopausal: 47.5 : 52.5 • Definition of postmenopausal: NR • CA125 (U/ml): mean 145.9 (SD NR) • HE4 (pmol/l): NR *Benign* • Number tested: 162 • Age (years): NR • % premenopausal : % postmenopausal: 84.0 : 16.0 • Definition of postmenopausal: NR • CA125 (U/ml): mean 16.1 (SD NR) • HE4 (pmol/l): NR	RMI 1 threshold comparison

Study details	Selection criteria	Participant details	Test(s)
Manjunath *et al.* (2001)[75] Country: India Funding: NR Recruitment start–end: January 1997–August 1999	Inclusion criteria: retrospective study of women admitted for surgical exploration of pelvic masses Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 93 • Age (years): NR • % premenopausal : % postmenopausal: 51 : 48 • Definition of postmenopausal: ≥ 1 year of amenorrhoea or aged > 50 years for women who had previously undergone a hysterectomy • CA125 (U/ml): mean 1215 (SD 3315.8); range 1–24,607 • HE4 (pmol/l): NR *Benign* • Number tested: 55 • Age (year): NR • % premenopausal : % • postmenopausal: 65 : 34 • Definition of postmenopausal: ≥ 1 year of amenorrhoea or aged > 50 years for women who had previously undergone a hysterectomy • CA125 (U/ml): mean 27.7 (SD 41.9); range 2–250 • HE4 (pmol/l): NR	RMI 1 threshold comparison
Meys *et al.* (2016)[44] Country: the Netherlands Funding: government (Academic fund of the University of Maastricht) Recruitment start–end: July 2011–July 2015	Inclusion criteria: consecutive women with adnexal pathology Exclusion criteria: no pathology result obtained, pathology result known before the ultrasound scan, pathology > 120 days after ultrasound or previous oophorectomy	*Metastases* • Number tested: 14 • Age (years): median 64.6 (SD NR); range 20–87.1 • % premenopausal : % postmenopausal: 21.4 : 78.6 • Definition of postmenopausal: previous hysterectomy, aged ≥ 50 years or amenorrhrea for ≥ 1 year • CA125 (U/ml): median 78.6 (SD NR); IQR 27.5–260.8 • HE4 (pmol/l): NR	ADNEX model; IOTA group's simple ultrasound rules and RMI 1

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
	Study setting: secondary Point in care pathway at which index test is given: following referral to an obstetrics and gynaecology department	*Stages II–IV* • Number tested: 56 • Age (years): median 67.7 (SD NR); range 32.3–87 • % premenopausal : % postmenopausal: 12.5 : 87.5 • Definition of postmenopausal: previous hysterectomy, aged ≥ 50 (years) or amenorrhoea for ≥ 1 year • CA125 (U/ml): median 456 (SD NR); IQR 170.8–1175 • HE4 (pmol/l): NR *Stage I* • Number tested: 18 • Age (years): median 63.1 (SD NR); range 50.3–68.5 • % premenopausal : % postmenopausal: 33.3 : 66.7 • Definition of postmenopausal: previous hysterectomy, aged ≥ 50 years or amenorrhoea for ≥ 1 year • CA125 (U/ml): median 109.5 (SD NR); IQR 16.8–361.5 • HE4 (pmol/l): NR *Borderline* • Number tested: 27 • Age (years): median 50.6 (SD NR); range 36.9–65.8 • % premenopausal : % postmenopausal: 55.6 : 44.4 • Definition of postmenopausal: previous hysterectomy, aged ≥ 50 years or amenorrhoea for ≥ 1 year • CA125 (U/ml): median 61.9 (SD NR); IQR 27.5–295 • HE4 (pmol/l): NR *Benign* • Number tested: 211 • Age (years): median 53.2 (SD NR); range 16.1–87.2 • % premenopausal : % postmenopausal: 45.9 : 54 • Definition of postmenopausal: previous hysterectomy, aged ≥ 50 years or amenorrhoea for ≥ 1 year • CA125 (U/ml): median 26 (SD NR); IQR 16.5–27 • HE4 (pmol/l): NR	

Study details	Selection criteria	Participant details	Test(s)
Moffatt *et al.* (2016)[45] Country: UK Funding: NR Recruitment start–end: January 2014–September 2015	Inclusion criteria: women with excised adnexal masses that had been sent for histological analysis Exclusion criteria: ectopic pregnancy, no ultrasound available or no CA125 level Study setting: unclear Point in care pathway index test is given: following referral to secondary care	*All* • Number tested: 81 • Age (years): NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ADNEX model
Moore *et al.* 2011[101] Country: USA Funding: mixed (Fujirebio Diagnostics and the National Cancer Institute) Recruitment start–end: October 2009–August 2010	Inclusion criteria: women (aged ≥ 18 years) presenting to a generalist (general gynaecologist, internist, family practitioner, gastroenterologist or general surgeon) with an ovarian cyst or adnexal mass and subsequently scheduled to undergo surgery Exclusion criteria: NR Study setting: mixed Point in care pathway at which index test is given: following referral to a general or specialist hospital	*All* • Number tested: 472 • Age (years): mean 50.3 (SD NR); range 18–89 • % premenopausal : % postmenopausal: 54 : 46 • Definition of postmenopausal: aged ≥ 55 years or FSH levels of > 22 IU/l • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: NR • Age (years): NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): median 266.8; range 31.2–13,250 • HE4 (pmol/l): median 366.8; range 11.9–1073.9 *Benign* • Number tested: NR • Age (years): NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): median 58.1; range 27.1–403.2 • HE4 (pmol/l): median 19.9; range 3.6–1085.1	ROMA score using Abbott Diagnostics' tumour marker assay

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Morgante *et al.* (1999)[80] Country: Italy Funding: NR Recruitment start–end: January 1995–December 1997	Inclusion criteria: women aged > 30 years admitted consecutively for surgical excision of ovarian masses Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 31 • Age (years): NR • % premenopausal : % postmenopausal: 29 : 68 • Definition of postmenopausal: > 1 year of amenorrhoea or aged > 50 years with a hysterectomy • CA125 (U/ml): mean 354 (SD NR); IQR 102–290 • HE4 (pmol/l): NR *Benign* • Number tested: 93 • Age (years): NR • % premenopausal : % postmenopausal: 65 : 35 • Definition of postmenopausal: > 1 year of amenorrhoea or aged > 50 years with a hysterectomy • CA125 (U/ml): mean 29.6 (SD NR); IQR 10–22 • HE4 (pmol/l): NR	RMI 1 threshold comparison
Murala *et al.* (2014)[60] Country: UK Funding: NR Recruitment start–end: September 2012–September 2013	Inclusion criteria: women referred to Poole District General Hospital or the Royal Bournmouth District General Hospital, with suspected adnexal pathology Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care	*All* • Number tested: 51 • Age (years): NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's Simple Rules and RMI 1

Study details	Selection criteria	Participant details	Test(s)
Novotny *et al.* (2012)[86] Country: the Czech Republic Funding: government (Ministry of Health, the Czech Republic) Recruitment start–end: NR	Inclusion criteria: women with abnormalities of the pelvis Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to a gynaecology and obstetrics department	*Malignant* • Number tested: 21 • Age (years): median 63 (SD NR); range 47–82 • % premenopausal : % postmenopausal: 0 : 100 • Definition of postmenopausal: aged ≥ 55 years or FSH level of > 22 IU/l • CA125 (U/ml): median 295 (SD NR); range 32.8–44,850 • HE4 (pmol/l): median 312; range 17.1–1842 *Benign* • Number tested: 256 • Age (years): median 64 (SD NR); range 48–93 • % premenopausal : % postmenopausal: 0 : 100 • Definition of postmenopausal: aged ≥ 55 years or FSH level of > 22 IU/l • CA125 (U/ml): median 16.2 (SD NR); range 3.6–2331 • HE4 (pmol/l): median 39.5; range 26.7–3590	ROMA score using Abbott Diagnostics' tumour marker assay
Piovano *et al.* (2016)[58] Country: Italy Funding: NR Recruitment start–end: February 2013–January 2015	Inclusion criteria: consecutive women (aged ≥ 18 years), with an adnexal mass, who were candidates for surgery Exclusion criteria: NR Study setting: unclear Point in care pathway at which index test is given: following referral to hospital	*All* • Number tested: 391 • Age (years): median 47 (SD NR); range 18–86 • % premenopausal : % postmenopausal: 56.5 : 43.5 • Definition of postmenopausal: amenorrhoea for at least 12 months or aged > 50 years and hysterectomy before menopause • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Presl *et al.* (2012)[81] Country: the Czech Republic Funding: government Recruitment start–end: June 2010–January 2011	Inclusion criteria: women with abnormalities in the pelvis Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to a university hospital's Department of Obstetrics and Gynaecology	*All* • Number tested: 552 • Age (years): NR • % premenopausal : % postmenopausal: 53.6 : 46.4 • Definition of postmenopausal: FSH level of ≥ 22 IU/l • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Abbott Diagnostics' tumour marker assay
Ruiz de Gauna *et al.* (2015)[64] Country: Spain Funding: NR Recruitment start–end: June 2012–December 2013	Inclusion criteria: women diagnosed with a persistent adnexal mass evaluated in one of two Spanish centres Exclusion criteria: pregnant women, masses with spontaneous resolution or masses removed surgically in another centre from recruitment Study setting: mixed Point in care pathway at which index test is given: following referral to secondary care	*All* • Number tested: 247 • Age (years): mean 43.6 (SD 14.1); range 14–83 • % premenopausal : % postmenopausal: 72.1 : 27.9 • Definition of postmenopausal: CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

Study details	Selection criteria	Participant details	Test(s)
Sayasneh 2013[62] Country: UK Funding: government (NHS, NIHR and Imperial College London) Recruitment start–end: September 2010–September 2012	Inclusion criteria: women with at least one adnexal mass, who underwent TVS examination at one of the participating centres Exclusion criteria: surgical removal of the mass > 120 days after ultrasound, refusal to undergo TVS, pregnancy, examined by a consultant with a specialist interest in gynaecological malignancy, or cytology rather than histology used to establish diagnosis Study setting: mixed Point in care pathway at which index test is given: following referral to secondary or tertiary care	*All* • Number tested: 255 • Age (years): mean 46 (SD NR), 95% CI 34 to 57 • % premenopausal : % postmenopausal: 64.7 : 35.3 • Definition of postmenopausal: ≥ 50 years who had undergone hysterectomy • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules; and RMI 1
Sayasneh *et al.* (2016)[46] Country: UK and Italy Funding: charity (NIHR; FWO grants; and a KU Leuven grant) Recruitment start–end: September 2010–February 2015	Inclusion criteria: women presenting ≥ 1 adnexal mass who underwent transvaginal ultrasonography Exclusion criteria: pregnancy, women examined by a consultant, refusal of TVS, cytology rather than histology as an outcome and failure to undergo surgery within 120 days of the ultrasound examination Study setting: tertiary care (cancer centres) Point in care pathway at which index test is given: referral to tertiary care	*All* • Number tested: 610 • Age (years): median 47 • % premenopausal : % postmenopausal: 58 : 42 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ADNEX model

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Shulman *et al.* (2016)[104] Country: USA Funding: NR Recruitment start–end: NR	Inclusion criteria: two published registries of women undergoing surgery for an adnexal mass Exclusion criteria: NR Study setting: secondary care Point in care pathway at which index test is given: following referral to secondary care	*All* • Number tested: 993 • Age (years): mean 50.3 (SD NR); range 18–92 • % premenopausal : % postmenopausal: 51 : 49 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	Overa (MIA2G) and ROMA score using Roche Diagnostics' tumour marker assay
Silvestre *et al.* (2015)[55] Country: Brazil Funding: NR Recruitment start–end: September 2008–December 2010	Inclusion criteria: women who were consecutively scheduled for surgery to remove adnexal masses Exclusion criteria: NR Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care (Department of Obstetrics and Gynaecology)	*Malignant* • Number tested: 32 • Age (years): median 52 (SD NR); range 20–78 • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Benign* • Number tested: 43 • Age (years): median 42 (SD NR); range 18–82 • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

Study details	Selection criteria	Participant details	Test(s)
Szubert *et al.* (2016)[42] Country: Poland and Spain Funding: NR Recruitment start–end: December 2012–April 2015	Inclusion criteria: women requiring surgery for an ovarian tumour, who had complete data required for the ADNEX calculation and who were evaluated between 1 and 5 days before surgery Exclusion criteria: NR Study setting: secondary care (Department of Obstetrics and Gynaecology or Gynaecological Surgery) Point in care pathway at which index test is given: following referral to secondary care (Division of Gynaecological Surgery or Department of Obstetrics and Gynaecology)	*Poland* • Number tested: 204 • Age (years): median 46 (SD NR); range 15–84 • % premenopausal : % postmenopausal: 62.4 : 32.4 • Definition of postmenopausal: NR • CA125 (U/ml): median 40 (SD NR); range 4–4909 • HE4 (pmol/l): NR *Spain* • Number tested: 123 • Age (years): median 47 (SD NR); range 12–81 • % premenopausal : % postmenopausal: 58.5 : 41.5 • Definition of postmenopausal: NR • CA125 (U/ml): median 40 (SD NR); range 1–3137 • HE4 (pmol/l): NR	ADNEX model
Tantipalakorn *et al.* (2014)[51] Country: Thailand Funding: government (Faculty of Medicine Research, Fund of Chiang Mai University and the National Research University Project, Thailand) Recruitment start–end: April 2007–March 2012	Inclusion criteria: women scheduled for surgery because of the detection of an adnexal mass (by pelvic examination, previous ultrasonography or both) Exclusion criteria: known diagnoses of adnexal masses, ovarian cancers scheduled for second-look operation or endometrioma diagnosed by previous laparoscopy, etc.; or patients undergoing surgery > 24 hours after ultrasound examination Study setting: secondary care Point in care pathway at which index test is given: following referral to secondary care	*All* • Number tested: 319 masses • Age (years): mean 42.4 (SD 16.2); range 13–82 • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Testa *et al.* (2014)[50] Country: Sweden, Belgium, Italy, Poland, Spain and the Czech Republic Funding: government (Swedish Research Council and the UK's NIHR) Recruitment start–end: October 2009–May 2012	Inclusion criteria: women with at least one adnexal mass (ovarian, paraovarian or tubal), who underwent TVS examination by a principal investigator at one of the participating centres Exclusion criteria: surgical removal of the mass > 120 days after ultrasound, pregnancy at ultrasound, unresolved data inconsistencies or incomplete final histology Study setting: mixed Point in care pathway at which index test is given: following referral to secondary or tertiary care	*All* • Number tested: 2403 • Age (years): NR • % premenopausal : % postmenopausal: 56.3 : 43.7 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: 980 • Age (years): median 57 (SD NR); IQR 46–66 • % premenopausal : % postmenopausal: 38.6 : 61.4 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Benign* • Number tested: 1423 • Age (years): median 44 (SD NR); IQR 33–56 • % premenopausal : % postmenopausal: 68.6 : 31.4 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules and RMI 1
Timmerman *et al.* (2010)[65] Country: Sweden, Belgium, Italy, Poland, UK, the Czech Republic, China and Canada Funding: government (Swedish Research Council and Research Council KU Leuven) Recruitment start–end: NR 2005–NR 2007	Inclusion criteria: women with at least one adnexal mass, who underwent TVS examination by a principal investigator at one of the participating centres Exclusion criteria: surgical removal of the mass > 120 days after ultrasound, refusal to undergo TVS, or pregnancy Study setting: mixed Point in care pathway at which index test is given: following referral to secondary or tertiary care	*All* • Number tested: 1938 • Age (years): mean 46 (SD NR); range 11–94 • % premenopausal : % postmenopausal: 62 : 38 • Definition of postmenopausal: aged ≥ 50 years or previous hysterectomy • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's Simple Rules and RMI 1

Study details	Selection criteria	Participant details	Test(s)
Tingulstad *et al.* (1996)[76] Country: Norway Funding: NR Recruitment start–end: February 1992–February 1994	Inclusion criteria: women with a pelvic mass, who were scheduled for laparotomy and who were at least 30 years old Exclusion criteria: NR Study setting: secondary care (Department of Gynaecology and Obstetrics) Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 56 • Age (years): NR • % premenopausal : % postmenopausal: 20 : 80 • Definition of postmenopausal: > 1 year of amenorrhoea or aged > 50 years with a hysterectomy • CA125 (U/ml): median 180 (SD NR); range 7–18400 • HE4 (pmol/l): NR *Benign* • Number tested: 117 • Age (years): NR • % premenopausal : % postmenopausal: 61 : 39 • Definition of postmenopausal: > 1 year of amenorrhoea or aged > 50 years with a hysterectomy • CA125 (U/ml): median 12 (SD NR); range 5–538 • HE4 (pmol/l): NR	RMI 1 threshold comparison
Tinnangwattana *et al.* (2015)[47] Country: Thailand Funding: government (Office of the Higher Education Commission) Recruitment start–end: March 2014–December 2014	Inclusion criteria: women scheduled for surgery because of an adnexal mass either detected by pelvic examination or previous ultrasound examination Exclusion criteria: known diagnoses or surgery > 24 hours after ultrasound examination Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to secondary care (Department of Obstetrics and Gynaecology)	*All* • Number tested: 100 • Age (years): mean 44.21 (SD 12.9); range 19–75 • % premenopausal : % postmenopausal: 73 : 27 • Definition of postmenopausal: postmenopausal period • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Ulusoy *et al.* (2007)[74] Country: Turkey Funding: NR Recruitment start–end: September 2002–November 2004	Inclusion criteria: consecutive women undergoing surgery for an adnexal mass Exclusion criteria: known ovarian malignancy or pregnancy Study setting: mixed Point in care pathway at which index test is given: following referral to secondary or tertiary care (Department of Gynaecology and Obstetrics and Gynaecological Oncology Clinic)	*Malignant* • Number tested: 106 • Age (years): mean 47 (SD NR) • % premenopausal : % postmenopausal: 53.8 : 46.2 • Definition of postmenopausal: > 1 year of amenorrhoea or an age of > 50 years in women who have had a hysterectomy • CA125 (U/ml): median 152.75 (SD NR); range 1–5000 • HE4 (pmol/l): NR *Benign* • Number tested: 190 • Age (years): mean 42 (SD NR) • % premenopausal : % postmenopausal: 67.9 : 31.2 • Definition of postmenopausal: > 1 year of amenorrhoea or an age of > 50 years in women who have had a hysterectomy • CA125 (U/ml): median 31.42 (SD NR); range 3–1153 • HE4 (pmol/l): NR	RMI 1 threshold comparison
Van Calster *et al.* (2014)[17] Country: Belgium, Sweden, Italy, the Czech Republic, Poland and the UK Funding: government (FWO) Recruitment start–end: NR 1999–NR 2012	Inclusion criteria: consecutive women with ≥ 1 adnexal mass (judged not to be a physiological cyst), examined with TVS and selected for surgical intervention Exclusion criteria: refusal of TVS, pregnancy or surgical removal of the mass > 120 days after ultrasound Study setting: mixed (oncology centres, general hospitals and gynaecology units) Point in care pathway at which index test is given: following referral to secondary or tertiary care	*All* • Number tested: 2403 • Age (years): NR • % premenopausal : % postmenopausal: 56.3 : 43.7 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR • Comments: for IOTA phase 3 validation data set *Malignant* • Number tested: 980 • Age (years): median 57 (SD NR); IQR 46–66 • % premenopausal : % postmenopausal: 38.6 : 61.4 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ADNEX model

Study details	Selection criteria	Participant details	Test(s)
		Benign • Number tested: 1423 • Age (years): median 44, (SD NR); IQR 33–56 • % premenopausal : % postmenopausal: 68.6 : 31.4 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	
Van Gorp *et al.* (2012)[98] Country: Belgium Funding: government (Belgian Federation against Cancer and FWO) Recruitment start–end: August 2005–March 2009	Inclusion criteria: women with a pelvic mass, suspected to be of ovarian origin, who were scheduled to undergo surgery Exclusion criteria: prior bilateral oophorectomy Study setting: unclear Point in care pathway at which index test is given: following referral to a university hospital for resection of a pelvic mass	*Malignant* • Number tested: 150 • Age (years): mean 57.7 (SD NR), 95% CI 55.7 to 59.8 • % premenopausal : % postmenopausal: 26 : 74 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Benign* • Number tested: 224 • Age (years): mean 46.2 (SD NR), 95% CI 44.1 to 48.3 • % premenopausal : % postmenopausal: 62.1 : 37.9 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Fujirebio Diagnostics' tumour marker assay and RMI 1
Weinberger and Minar (2013)[53] Country: NR Funding: other (NR) Recruitment start–end: NR 2010–NR 2012	Inclusion criteria: women with suspiscious adnexal mass Exclusion criteria: NR Study setting: unclear Point in care pathway at which index test is given: following referral (unclear whether secondary or tertiary care)	*All* • Number tested: 347 • Age: NR • % premenopausal : % postmenopausal: NR • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	IOTA group's simple ultrasound rules

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Winarto *et al.* (2014)[99] Country: Indonesia Funding: NR Recruitment start–end: November 2010–May 2011	Inclusion criteria: women diagnosed with an ovarian tumour, by physical examination and TVS Exclusion criteria: unresectable tumour, non-epithelial histopathological results, history of oophorectomy, ovarian cancer treatment or pregnancy Study setting: unclear Point in care pathway at which index test is given: following referral to hospital	*Malignant* • Number tested: • Age (years): mean 44 (SD NR) • % premenopausal : % postmenopausal: 62.2 : 37.8 • Definition of postmenopausal: NR • CA125 (U/ml): median 357.5 (SD NR); range 13.1–9872.3 • HE4 (pmol/l): median 495.5; range 436.3–15,000 *Benign* • Number tested: • Age (years): mean 41 (SD NR) • % premenopausal : % postmenopausal: 72 : 27.9 • Definition of postmenopausal: NR • CA125 (U/ml): median 82.5 (SD NR); range 8.1–2441.4 • HE4 (pmol/l): median 52.3; range 29.5–26.1	ROMA score using Abbott Diagnostics' tumour marker assay and RMI 1
Xu *et al.* (2016)[95] Country: China Funding: government (Guangdong Natural Science Foundation, Guangdong Province Science and Technology Project Plan and Social Development Foundation, and the Medical Science and Technology Research Foundation of Guangdong Province) Recruitment start–end: July 2013–November 2014	Inclusion criteria: retrospective study of women with an ovarian mass Exclusion criteria: missing tumour marker data or women with non-epithelial ovarian cancer Study setting: unclear Point in care pathway at which index test is given: following referral to secondary care	*Malignant* • Number tested: 239 • Age (years): mean 57 (SD NR) • % premenopausal : % postmenopausal: 54 : 46 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Borderline* • Number tested: 45 • Age (years): mean 40 (SD NR) • % premenopausal : % postmenopausal: 80 : 20 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Roche Diagnostics' tumour marker assay

Study details	Selection criteria	Participant details	Test(s)
		Benign • Number tested: 311 • Age (years): mean 42 (SD NR) • % premenopausal : % postmenopausal: 85 : 15 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	
Yamamoto *et al.* (2009)[72] Country: Japan Funding: NR Recruitment start–end: January 2002–April 2005	Inclusion criteria: women with a pelvic mass scheduled for laparotomy and laparoscopy at the Department of Obstetrics and Gynaecology, Kochi Medical School Exclusion criteria: NR Study setting: secondary care (Department of Gynaecology and Obstetrics) Point in care pathway at which index test is given: following referral to secondary care (Department of Obstetrics and Gynaecology)	*Malignant* • Number tested: 40 • Age (years): NR • % premenopausal : % postmenopausal: 37.5 : 62.5 • Definition of postmenopausal: > 1 year of amenorrhoea or an age of > 50 years in women who have had a hysterectomy • CA125 (U/ml): median 124 (SD NR); range 11.4–4340 • HE4 (pmol/l): NR *Benign* • Number tested: 213 • Age (years): NR • % premenopausal : % postmenopausal: 81.7 : 18.3 • Definition of postmenopausal: > 1 year of amenorrhoea or an age of > 50 years in women who have had a hysterectomy • CA125 (U/ml): median 35.2 (SD NR); range 5–616 • HE4 (pmol/l): NR	RMI 1 threshold comparison

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Yanaranop *et al.* (2016)[89] Country: Thailand Funding: NR Recruitment start–end: January 2012–December 2012	Inclusion criteria: women, aged ≥ 18 years, undergoing elective surgery for clinically diagnosed pelvic or adnexal mass Exclusion criteria: pregnancy, previous history of ovarian cancer, any known malignancy, previous history of adnexal surgery, incomplete ultrasound or biomarker results, or cancelled surgery Study setting: secondary care (general gynaecology) Point in care pathway at which index test is given: following referral to an obstetrics and gynaecology department	*All* • Number tested: 260 • Age (years): mean 48.2 (SD 14.2) • % premenopausal : % postmenopausal: 56.9 : 43.1 • Definition of postmenopausal: aged > 55 years, aged < 45 years with amenorrhoea for > 1 year or a FSH level of < 25 IU/l • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: 74 • Age (years): NR • % premenopausal : % postmenopausal: 37.8 : 62.2 • Definition of postmenopausal: aged > 55 years, aged < 45 years with amenorrhoea for > 1 year or a FSH level of < 25 IU/l • CA125 (U/ml): median 274.1 (SD NR) • HE4 (pmol/l): median 165.1; range NR *Benign* • Number tested: 186 • Age (years): NR • % premenopausal : % postmenopausal: 64.5 : 35.5 • Definition of postmenopausal: aged > 55 years, aged < 45 years with amenorrhoea for > 1 year or a FSH level of < 25 IU/l • CA125 (U/ml): median 32.9 (SD NR), IQR NR • HE4 (pmol/l): median 57.3; range NR	ROMA score using Roche Diagnostics' tumour marker assay and RMI 1

Study details	Selection criteria	Participant details	Test(s)
Zhang *et al.* (2015)[68] Country: USA Funding: NR Recruitment start–end: NR	Inclusion criteria: women with a documented pelvic mass scheduled for surgery Exclusion criteria: NR Study setting: unclear Point in care pathway at which index test is given: following referral (unclear whether secondary or tertiary care)	*All* • Number tested: 305 • Age (years): NR • % premenopausal : % postmenopausal: 50.5 : 49.5 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Malignant* • Number tested: 264 • Age (years): NR • % premenopausal : % postmenopausal: 36 : 64 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Benign* • Number tested: 348 • Age (years): NR • % premenopausal : % postmenopausal: 81.9 : 18.1 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	Overa (MIA2G)

continued

TABLE 33 Study details and baseline participant characteristics (*continued*)

Study details	Selection criteria	Participant details	Test(s)
Zhang *et al.* (2015)[102] Country: China Funding: government (National High Technology Research and Development Programme, China Postdoctoral Science Special Foundation, National Science and Technology Infrastructure, and the National Science Foundation of China) Recruitment start–end: October 2012–February 2013	Inclusion criteria: women with a pelvic mass, suspected to be of ovarian origin, who were to undego surgery Exclusion criteria: aged < 18 years; missing clinical examination results; blood sample of < 0.5 ml, stored or transported at > 0 °C, lipaemic or haemolytic appearance; pregnancy; family history of ovarian cancer; or receiving chemotherapy, radiotherapy and other treatments Study setting: unclear Point in care pathway at which index test is given: following referral to one of nine centres	*Malignant* • Number tested: 264 • Age (years): NR • % premenopausal:% postmenopausal: 36: 64 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR *Benign* • Number tested: 348 • Age (years): NR • % premenopausal : % postmenopausal: 81.9 : 18.1 • Definition of postmenopausal: NR • CA125 (U/ml): NR • HE4 (pmol/l): NR	ROMA score using Roche Diagnostics' tumour marker assay

FWO, *Fonds voor Wetenschappelijk Onderzoek-Vlaanderen* (Research Foundation – Flanders); IQR, interquartile range; MNISW, Ministerstwa Nauki i Szkolnictwa Wyższego (Ministry of Science and Higher Education); NR, not reported; SD, standard deviation.

Index test details

Risk of Ovarian Malignancy Algorithm score using Abbott Diagnostics' tumour marker assay

	ROMA score using Abbott Diagnostics' tumour marker assay details		
Study (year of publication)	Analyser, manufacturer of CA125 and HE4 assays	Sample collection, storage	Time from test to surgery
Al Musalhi *et al.* (2016)[103]	ARCHITECT I2000, Abbott Diagnostics	Samples collected using serum separator tubes and centrifuged immediately Serum samples were stored at –20 °C	NR
Chan *et al.* (2013)[82]	ARCHITECT, Abbott Diagnostics	Blood samples were collected after pelvic mass was confirmed and surgery scheduled, to minimise the time between testing and surgery. Samples were centrifuged and serum separated within 4 hours of collection Serum samples were stored at –20 °C	NR
Chan *et al.* (2013)[82]	ARCHITECT, Abbott Diagnostics	Blood samples were collected after pelvic mass was confirmed and surgery scheduled, to minimise the time between testing and surgery. Samples were centrifuged and serum separated within 4 hours of collection Serum samples were stored at –20 °C	NR
Clemente *et al.* (2015)[90]	NR, Abbott Diagnostics	NR	NR
Karlsen *et al.* (2012)[83]	ARCHITECT I 2000sr, Abbott Diagnostics	Blood samples were collected within 2 weeks prior to surgery Samples were centrifuged within 6 hours of collection. After centrifugation, serum samples were stored at –80 °C until analysis	NR
Li *et al.* (2016)[96]	ARCHITECT, Abbott Diagnostics	Blood samples were collected on the day of surgery, before anaesthesia. Samples were centrifuged and serum separated Serum samples were stored at –80 °C	< 1 day
Moore *et al.* (2011)[101]	ARCHITECT i2000, Abbott Diagnostics	Blood samples were collected within 30 days prior to surgery and before induction of anaesthesia Samples were collected into a serum separator tube and centrifuged after clotting Serum samples were stored at –20 °C	≤ 30 days
Novotny *et al.* (2012)[86]	ARCHITECT 1000i, Abbott Diagnostics	Blood samples were collected prior to surgery or treatment and centrifuged Serum samples were stored at –80 °C	NR

Study (year of publication)	ROMA score using Abbott Diagnostics' tumour marker assay details		
	Analyser, manufacturer of CA125 and HE4 assays	Sample collection, storage	Time from test to surgery
Presl *et al.* (2012)[81]	ARCHITECT 1000, Abbott Diagnostics	Blood samples were centrifuged immediately or within 24 hours of collection Serum samples were stored at –80 °C	NR
Winarto *et al.* (2014)[99]	NR, Abbott Diagnostics	NA	Blood samples collected 1 day before surgery, time from ultrasound to surgery unclear

NA, not applicable; NR, not reported.

Risk of Ovarian Malignancy Algorithm score using Roche Diagnostics' tumour marker assays

Study (year of publication)	ROMA score using Roche Diagnostics' tumour marker assay details		
	Analyser, manufacturer of CA125 and HE4 assays	Sample collection, storage	Time from test to surgery
Janas *et al.* (2015)[97]	NR, Roche Diagnostics	NR	NR
Shulman *et al.* (2016)[104]	NR	NR	NR
Xu *et al.* (2016)[95]	Cobas E170, Roche Diagnostics	Blood samples were collected before surgery and centrifuged within 3 hours Serum samples were stored at –80 °C	NR
Yanaranop *et al.* (2016)[89]	Cobas 6000, Roche Diagnostics	Samples were collected within 48 hours prior to surgery and centrifuged immediately Serum samples were stored at –20 °C	Within 6 weeks before surgery
Zhang *et al.* (2015)[102]	Cobas 601, Roche Diagnostics	Blood samples were collected into a tube containing a clot activator and centrifuged Serum samples were stored at –80 °C	NR

NR, not reported.

Risk of Ovarian Malignancy Algorithm score using Fujirebio Diagnostics' tumour marker assays

	ROMA score using Fujirebio Diagnostics' tumour marker assay details		
Study (year of publication)	Analyser, manufacturer of CA125 and HE4 assay	Sample collection, storage	Time from test to surgery
Langhe *et al.* (2013)[94]	NR, Fujirebio Diagnostics	NR	Collected before surgery
Van Gorp *et al.* (2012)[98]	NR, Fujirebio Diagnostics	Blood samples were collected in clotting tubes, immediately before surgery After centrifugation, serum samples were stored at –80 °C until analysis	Time from ultrasound to surgery NR

NR, not reported.

Assessment of Different NEoplasias in the adneXa

	ADNEX model test details			
Study (year of publication)	Analyser, manufacturer of CA125 assay	Ultrasound details	Sample collection storage	Time from test to surgery
IOTA5 (2017)[a]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Joyeux *et al.* (2016)[43]	NR	TVS could be complemented by another imaging technique (abdominal ultrasound, CT scan, MRI)	NR	NR
Meys *et al.* (2016)[44]	NR	TVS or transrectal ultrasound with transabdominal ultrasound for larger masses	NA	≤ 120 days
Meys *et al.* (2016)[44]	NR	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	NA	≤ 120 days
Moffatt *et al.* (2016)[45]	NR	NR	NR	NR
Sayasneh *et al.* (2016)[46]	NR	TVS examinations performed by EFSUMB level 2 ultrasound examiners (non-consultant gynaecology specialist, gynaecology trainees doctors and gynaecology sonographers) The ultrasound examiners were blind to the results of the reference test TVS was performed using the standardised approach previously published by the IOTA group[14] Transabdominal ultrasonography was undertaken for a large mass	NR	NR

Study (year of publication)	ADNEX model test details			
	Analyser, manufacturer of CA125 assay	Ultrasound details	Sample collection storage	Time from test to surgery
Szubert *et al.* (2016)[42]	Unclear Immunoenzymatic test (ST AIA-PACK OVCATosoH Bioscience) and Cobas-Core CA-125-II (Roche Diagnostics) Roche Diagnostics and Tosoh Bioscience	Poland: Aloka Alpha 10 (3.75–7.5 MHz) endovaginal probe and Aloka 3500 (7.5 MHz) endovaginal probe (Hitachi Aloka) Spain: TVS or transrectal ultrasound Voluson E8 (RIC5–9 MHz) endovaginal probe (GE Healthcare) A transabdominal probe was used for large tumours Tumours were ultrasonographically assessed according to the 2000 IOTA criteria 14	Poland: assessed 1–5 days before surgery Spain: assessed 1–5 days before ultrasound No further details reported	1–5 days
Van Calster *et al.* (2014)[17]	Immunoradiometric assay kits for CA-125 II Roche Diagnostics, Centocor, Cis-Bio, Abbott Laboratories, Bayer Diagnostics, bioMérieux, DiaSorin, Siemens and Beckman Coulter	Standardised TVS examination (additional transabdominal sonography for women with large masses) All ultrasound examinations were performed by one of three experienced practitioners [with 8–20 years' experience in gynaecological sonography and EFSMUB level 2 (Poland) and level 3 (Spain)]	NR	NR

NR, not reported.
a Frances Nixon, personal communication.

International Ovarian Tumour Analysis group's simple ultrasound rules

Study (year of publication)	Ultrasound details	Time from test to surgery
Abdalla *et al.* (2013)[48]	TVS with transabdominal ultrasound for tumours larger than 5 cm and extended beyond the pelvis minor. Morphology, echostructure and vascularisation were assessed by Doppler examination. Ultrasound examinations were performed by the attending physician (various levels of experience) prior to referral to the hospital	≤ 90 days
Alcazar *et al.* (2013)[52]	Transvaginal colour Doppler ultrasound (5- to 9-MHz transducers), Voluson E8 or 730 machine (GE Healthcare, Chicago, IL, USA). Transabdominal scanning was also performed in large masses. Ultrasound was performed by a trainee or junior staff under the supervision of an expert	Surgery was performed within 3 weeks after ultrasound examination
Baker *et al.* (2013)[66]	Retrospective review of ultrasound scan reports	NR
DiLegge *et al.* (2012)[61]	High-frequency transvaginal probe with transabdominal ultasonography for large masses that could not be entirely visualised using a transvaginal probe	≤ 120

continued

Study (year of publication)	Ultrasound details	Time from test to surgery
Fathallah *et al.* (2011)[63]	Endovaginal	NR
IOTA5 (2017)[a]	Confidential information has been removed	Confidential information has been removed
Knafel *et al.* (2015)[49]	Transvaginal (5–9 MHz) ultrasound with transabdominal (2–5 MHz) ultrasound for larger tumours. Examinations were performed by both EFSUMB level 1 and level 2 examiners. All examiners received 1 half-day of practical training in the IOTA group's simple ultrasound rules before the study	≤ 90 days
Meys *et al.* (2016)[44]	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	≤ 120 days
Murala *et al.* (2014)[60]	Scan images were analysed by non-expert gynaecology trainees and masses were classified as benign, malignant or inconclusive	NR
Piovano *et al.* (2016)[58]	Greyscale and Doppler TVS, performed by a trainee who had undergone IOTA group's Simple Rules training and was supervised by an experienced examiner. A transabdominal probe was used for large masses that could not be entirely visualised transvaginally. All inconclusive masses were re-evaluated by a consultant expert (EFSUMB level)	≤ 30 days
Ruiz de Gauna *et al.* (2015)[64]	Transvaginal colour Doppler ultrasound (5- to 9-MHz transducers), Voluson E8 machine. Transabdominal scanning was also performed in large masses. In Centre A, ultrasound scanning was performed by an expert, and in centre B, ultrasound scanning was performed by a trainee. In both centres, masses classified as inconclusive by the IOTA group's simple ultrasound rules were given a classification of benign, malignant or uncertain, based on the subjective assessment of an expert examiner; patients with a final classification of malignant or uncertain were referred to specialist gynaecological oncology services	Surgery was performed within 3 weeks after ultrasound examination
Sayasneh *et al.* (2013)[62]	Standardised ultrasound conducted at one of the participating centres. Transabdominal ultasonography was used for large masses that could not be entirely visualised using a transvaginal probe	≤ 120 days
Silvestre *et al.* (2015)[55]	The descriptions of the masses were interpreted based on the IOTA group's simple ultrasound rules[15] to characterise whether the features were malignant or benign. Vascular power Doppler score is included in the IOTA group's simple ultrasound rules as one variable: a score of 1 is given when no blood flow is found in the tumour, a score of 2 when only minimal flow is detected, a score of 3 when moderate flow is present and a score of 4 when the tumour presents marked blood flow	7 days
Tantipalakorn *et al.* (2014)[51]	Transabdominal (3.5- to 5-MHz curvilinear transducer) or transvaginal (real-time 5–7.5 MHz) or both, connected to Aloka model SSD alpha-10 (Tokyo, Japan). IOTA group's simple ultrasound rules[15] were applied to determine whether there were malignant (M) features or benign (B) features. If one or more M-rules apply in the absence of a B-rule, the mass is classified as malignant. If one or more B-rules apply in the absence of an M-rule, the mass is classified as benign. If both M-rules and B-rules apply, the mass cannot be classified or inconclusive. Likewise, if no rule applies, the mass cannot be classified or inconclusive	All participants underwent ultrasound examination within 24 hours of operation

Study (year of publication)	Ultrasound details	Time from test to surgery
Testa *et al.* (2014)[50]	Standardised TVS by examiners experienced in gynaecological ultrasound (Level III Education, Practical Standards Committee, EFSUMB), grey scale and Doppler imaging; when there was more than one adnexal mass, the mass with the most complex morphology was assessed and analysed	≤ 120 days
Timmerman *et al.* (2010)[65]	Standardised ultrasound conducted by a principal investigator at one of the participating centres. All principal investigators were fully trained gynaecologists or radiologists with a special interest in gynaecological ultrasound and at least 5 years' experience. Transvaginal probe frequencies ranged from 5 to 12 MHz and transabdominal ultasonography was used for large masses that could not be entirely visualised using a transvaginal probe. Doppler ultrasound images were used to obtain morphological and blood-flow variables	≤ 120 days
Tinnangwattana *et al.* (2015)[47]	All examinations were done with either transabdominal or transvaginal approach as suitable, using real-time 5- to 7.5-MHz transvaginal or 2.5- to 5-MHz transabdominal curvilinear transducer connected to a machine (Hitachi Aloka model ProSound37)	≤ 24 hours
Weinberger and Minar (2013)[53]	Retsopective analysis by an experienced sonographer	NR

NR, not reported.
a Frances Nixon, personal communication.

Overa (MIA2G)

Study (year of publication)	Overa (MIA2G) test details		
	Analyser, manufacturer of assays	Sample collection, storage	Time from test to surgery
Coleman *et al.* (2016)[70]	Roche Diagnostics' Cobas 6000 clinical analyser (c501 and e601 modules) Roche Diagnostics' Cobas assays for apo A-1, TRF (immunoturbidimetric assays), CA125-II, HE4 and FSH (electrochemiluminescent detection)	A preoperative blood sample of 80 ml was processed within 1–6 hours of collection, and serum was frozen at the collection site Frozen and stored at –65 to –85 °C. No sample had undergone > 2 or < 2 freeze–thaw cycles	Median 1 week (range 0–11)
Shulman *et al.* (2016)[104]	NR	NR	NR
Zhang *et al.* (2015)[68]	NR	NR	NR

NR, not reported.

Risk of Malignancy Index 1

Study (year of publication)	RMI test details			
	Analyser, manufacturer of CA125 assay	Ultrasound details	Sample collection, storage	Time from test to surgery
Abdalla *et al.* (2013)[48]	NR	TVS with transabdominal ultrasound for tumours larger than 5 cm and extended beyond the pelvis minor Morphology, echostructure and vascularisation were assesed by Doppler examination Ultrasound examinations were performed by the attending physician (various levels of experience) prior to referral to the hospital	NR	≤ 90 days
Aktürk *et al.* (2011)[71]	Electrochemiluminescence immunoassay, Roche Diagnostics	Siemens transvaginal 7.5-MHz transducer, with transabdominal ultrasound if the mass was too large for complete visualisation transvaginally	Serum samples were collected preoperatively NR	NR
Al Musalhi *et al.* (2016)[103]	ARCHITECT I2000, Abbott Diagnostics	Pelvic ultrasonography by specialist gynaecologists	Samples were collected using serum separator tubes and centrifuged immediately Serum samples were stored at –20 °C	NR
Asif *et al.* (2004)[77]	IMMULITE-Automated Analyser DPC-U5 A (CA125), Siemens Healthineers, Erlangen, Germany	Score based on presence of multilocular cystic lesion, solid areas, bilateral lesions, ascites and abdominal metastasis: 0 = no positive factor; 1 = single positive factor; 3 = two–five positive factors	Venous blood was collected in a plain tube, avoiding haemodialysis. Serum was isolated by centrifugation Serum was stored at –20 °C	NR
Davies *et al.* (1993)[79]	Radioimmunoassay, CIS bioindustries	One point score was assigned for ultrasound investigation for each of the following: multilocular cyst, evidence of solid areas, evidence of metastases, presence of ascites and bilateral lesions	Peripheral venous blood samples were drawn from each patient before surgery Blood was allowed to clot at room temperature then centrifuged at 3000 r.p.m. for 10 minutes and serum was separated and stored at –20 °C	NR

Study (year of publication)	RMI test details			
	Analyser, manufacturer of CA125 assay	Ultrasound details	Sample collection, storage	Time from test to surgery
Di Legge *et al.* (2012)[61]	Centocor or Cis-Bio or Abbott Diagnostics' Axsym system or Immuno-I-analyser or Vidas Centocor or Cis-Bio or Abbott Diagnostics' or Bayer or Vidas	High-frequency transvaginal probe with transabdominal ultasonography for large masses that could not be entirely visualised using a transvaginal probe	NR	≤ 120 days
IOTA5 (2017)[a]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Jacobs *et al.* (1990)[78]	Radioimmunoassay, Abbott Laboratories CA125	One point score was assigned for ultrasound investigation for each of the following: multilocular cyst, evidence of solid areas, evidence of metastases, presence of ascites and bilateral lesions	Peripheral venous blood samples were drawn from each patient before surgery Blood was allowed to clot at room temperature then centrifuged at 3000 r.p.m. for 10 minutes and serum was separated and stored at –20 °C	NR
Karlsen *et al.* (2012)[83]	ARCHITECT I 2000sr, Abbott Diagnostics	No details reported	Blood samples were collected within 2 weeks prior to surgery Samples were centrifuged within 6 hours of collection. After centrifugation, serum samples were stored at –80 °C until analysis	Time from ultrasound to surgery, NR
Lou *et al.* (2010)[73]	NR	NR	NR	NR
Manjunath *et al.* (2001)[75]	Microparticle EIA, Abbott Diagnostics' AXSYM System	The ultrasound was performed vaginally by a 5-MHz transducer (Ultramark 4 PLUS, Advanced Technology Laboratories) and extended to the transabdominal approach with full bladder if the mass was huge. A score (1 point each) was assigned for the following morphological features seen on ultrasound, suggestive of malignancy: the presence of a multilocular cystic lesion, solid areas, bilateral lesions, ascites and intra-abdominal metastases	Serum samples were collected preoperatively No details of storage were reported	NR

Study (year of publication)	RMI test details			
	Analyser, manufacturer of CA125 assay	Ultrasound details	Sample collection, storage	Time from test to surgery
Meys *et al.* (2016)[44]	NR	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	NR	≤ 120 days
Morgante *et al.* (1999)[80]	NA, Centocor	Siemens Somoline SL2 with transabdominal probe (3.5 MHz) and transvaginal probe (5–7.5 MHz). One point score was assigned for ultrasound investigation for each characteristics: presense of multilocular cystic lesions, solid areas, bilateral lesions, ascites and intra-abdominal metastates	Peripheral venous blood samples were drawn from each patient before surgery Serum stored at –15 °C	NR
Sayasneh *et al.* (2013)[62]	Abbott Diagnostics' ARCHITECT CA125 II immunoassay kit, ADVIA Centaur XP Immunoassay System, UniCel DxI Immunoassay System Beckman; Abbott Diagnostics; Siemens	Transabdominal ultasonography was used for large masses that could not be entirely visualised using a transvaginal probe	NR	NR
Testa *et al.* (2014)[50]	NR	Standardised TVS by examiners experienced in gynaecological ultrasound (EFSUMB level 3) Greyscale and Doppler imaging When there was more than one adnexal mass, the mass with the most complex morphology was assessed and analysed	NR	≤ 120 days
Timmerman *et al.* (2010)[65]	NR	Standardised ultrasound conducted by a principal investigator at one of the participating centres. All principal investigators were fully trained gynaecologists or radiologists with a special interest in gynaecological ultrasound and at least 5 years' experience. Transvaginal probe frequencies ranged from 5 to 12 MHz and transabdominal ultasonography was used for large masses that could not be entirely visualised using a transvaginal probe. Doppler ultrasound images were used to obtain morphologiocal and blood flow variables	NR	NR

Study (year of publication)	RMI test details			
	Analyser, manufacturer of CA125 assay	Ultrasound details	Sample collection, storage	Time from test to surgery
Tingulstad *et al.* (1996)[76]	NR, Abbott Diagnostics	TVS with transabdominal ultrasound as needed	NR	NR
Ulusoy *et al.* (2007)[74]	Roche Diagnostics–Hitachi Modular E170 Immunologic Analyser System, NR	Ultrasound examinations performed with Toshiba Sonolayer SSA-270A and/or a Siemens Sonoline G50 (abdominal convex transducers and/or endovaginal probes). Gynaecological oncologists evaluated all patients	NR	NR
Van Gorp *et al.* (2012)[98]	NR, Fujirebio Diagnostics	Standardised TVS performed by an experienced examiner or a trainee supervised by an experienced examiner. Transabdominal ultrasound was added for large masses that could not be visualised completely using a transvaginal probe	Blood samples were collected in clotting tubes immediately before surgery After centrifugation, serum samples were stored at –80 °C until analysis	Time from ultrasound to surgery NR
Winarto *et al.* (2014)[99]	NR, Abbott Diagnostics	TVS	NR	Samples collected 1 day before surgery, time from ultrasound to surgery unclear
Yamamoto *et al.* (2009)[72]	ECLusys CA125 II assay, Roche Diagnostics	Transvaginally with a 6.0 MHz transducer (an abdominal scan was also conducted when indicated)	Blood samples were taken pre-operatively No further details reported	NR
Yanaranop *et al.* (2016)[89]	Cobas 6000, Roche Diagnostics	Pelvic (transabdominal or transvaginal) ultrasound using a Voluson E8 Examiner blinded to clinical information and serum biomarkers Morphological features noted (wall structure and thickness, echogenicity, multiocularity, solid areas, bilaterality, ascites and intra-abdominal metastases)	Samples were collected within 48 hours prior to surgery and centrifuged immediately Serum samples were stored at –20 °C	Within 6 weeks to surgery

NA, not applicable; NR, not reported; r.p.m., revolutions per minute.
a Frances Nixon, personal communication.

TABLE 34 Study-level data for the histological details of malignant tumour diagnoses

Study (year of publication)	Test	Histology details for malignancy (*n*)
Abdalla *et al.* (2013)[48]	IOTA group's simple ultrasound rules; and RMI 1	Malignancy included serous cystadenocarcinoma ($n = 7$), metastatic tumours from the gastrointestinal tract ($n = 3$), borderline tumours ($n = 3$), mucinous cystadenocarcinoma ($n = 1$), fallopian tube carcinoma ($n = 1$), mixed carcinoma ($n = 1$) and undifferentiated carcinoma ($n = 1$)
Aktürk *et al.* (2011)[71]	RMI 1 threshold comparison	Malignancy included primary ovarian cancer ($n = 19$) and metastases ($n = 1$)
Al Musalhi *et al.* (2016)[103]	ROMA score using Abbott Diagnostics' tumour marker assay and RMI 1	Malignancy included serous adenocarcinoma ($n = 20$), mucinous adenocarcinoma ($n = 1$), endometrial adenocarcinoma ($n = 3$), undifferentiated ($n = 1$), borderline epithelial ($n = 7$), granulosa ($n = 5$), yolk sac cancer ($n = 1$), teratoma ($n = 2$), secondaries ($n = 7$) and lymphoma ($n = 1$)
Alcázar *et al.* (2013)[52]	IOTA group's simple ultrasound rules	Malignancy included invasive epithelial ovarian cancer ($n = 29$), borderline ovarian cancer ($n = 16$) and other malignancies ($n = 7$)
Asif *et al.* (2004)[77]	RMI 1 threshold comparison	No histology details
Baker *et al.* (2013)[66]	IOTA group's simple ultrasound rules	No histology details
Chan *et al.* (2013)[82]	ROMA score using Abbott Diagnostics' tumour marker assay	Malignancy included epithelial ovarian carcinoma only ($n = 65$), serous adenocarcinoma ($n = 30$), mucinous adenocarcinoma ($n = 14$), endometrial adenocarcinoma ($n = 7$), clear cell ($n = 8$), mixed ($n = 5$) and poorly differentiated ($n = 1$) Subgroups looked at stages I–IV and borderline
Clemente *et al.* (2015)[90]	ROMA score using Abbott Diagnostics' tumour marker assay	No histology details
Coleman *et al.* (2016)[70]	Overa (MIA2G)	Malignancy included epithelial ovarian cancer ($n = 60$), non-epithelial ovarian cancer ($n = 5$), borderline ovarian cancer ($n = 17$), metastases to the ovaries ($n = 6$) and other non-ovarian malignancies ($n = 4$)
Davies *et al.* (1993)[79]	RMI 1 threshold comparison	Malignancy included epithelial ovarian cancer ($n = 28$), borderline ovarian cancer ($n = 7$) and other malignancies ($n = 2$)
Di Legge *et al.* (2012)[61]	IOTA group's simple ultrasound rules; and RMI 1	Malignancy included primary invasive ($n = 476$), borderline ($n = 128$) and metastatic tumours ($n = 78$)
Fathallah *et al.* (2011)[63]	IOTA group's simple ultrasound rules	Malignancy included primary ovarian cancer ($n = 8$) and borderline ovarian cancer ($n = 6$)
IOTA5 (2017)[a]	ADNEX model, IOTA group's simple ultrasound rules and RMI 1	Confidential information has been removed
Jacobs *et al.* (1990)[78]	RMI 1 threshold comparison	Malignancy included primary invasive epithelial ovarian malignancies ($n = 36$), dysgerminoma ($n = 1$), metastatic bowel adenocarcinoma ($n = 1$) and borderline malignancy ($n = 4$)
Janas *et al.* (2015)[97]	ROMA score using Roche Diagnostics' tumour marker assay	Malignancy included primary ovarian cancer ($n = 44$), metastases to the ovary ($n = 14$) and borderline tumours ($n = 8$) Subgroups looked at ovarian cancer alone

continued

TABLE 34 Study-level data for the histological details of malignant tumour diagnoses (*continued*)

Study (year of publication)	Test	Histology details for malignancy (*n*)
Joyeux *et al.* (2016)[43]	ADNEX model	Malignancy included primary ovarian cancer ($n = 25$) and borderline ovarian cancer ($n = 5$)
Karlsen *et al.* (2012)[83]	ROMA score using Abbott Diagnostics' tumour marker assay; and RMI 1	No histology details
Knafel *et al.* (2015)[49]	IOTA group's simple ultrasound rules	Malignancy included ovarian carcinoma ($n = 60$: serous, clear cell, endometrioid, mucinous, undifferentiated, carcinosarcoma), borderline ($n = 7$), sex cord–stromal tumours ($n = 2$), germ cell tumours ($n = 5$) and metastases ($n = 8$) Note that 15 out of 82 malignancies were not classed as ovarian
Langhe *et al.* (2013)[94]	ROMA score using Fujirebio Diagnostics' tumour marker assay	Malignant included 53 borderline tumours. No further details reported
Li *et al.* (2016)[96]	ROMA score using Abbott Diagnostics' tumour marker assay	Ovarian malignancy included serous ($n = 80$), mucinous ($n = 42$), endometrioid ($n = 40$), clear cell ($n = 21$) and undifferentiated ($n = 7$)
Lou *et al.* (2010)[73]	RMI 1 threshold comparison	Ovarian malignancy included epithelial ovarian cancer ($n = 50$), non-epithelial ovarian cancer ($n = 8$) and metastatic carcinoma ($n = 3$)
Manjunath *et al.* (2001)[75]	RMI 1 threshold comparison	Malignancy included primary ovarian malignancies ($n = 88$), germ cell tumours ($n = 3$) and metastases ($n = 2$)
Meys *et al.* (2016)[44]	ADNEX model, IOTA group's simple ultrasound rules and RMI 1	Malignancy included epithelial ovarian cancer ($n = 70$), borderline ($n = 27$), granulosa cell carcinoma ($n = 3$), yolk sac tumour ($n = 1$), metastatic tumour ($n = 10$) and non-primary ovarian carcinoma ($n = 4$)
Moffatt *et al.* (2016)[45]	ADNEX model	No histology details
Moore *et al.* (2011)[101]	ROMA score using Abbott Diagnostics' tumour marker assay	Epithelial ovarian cancer ($n = 43$) and low malignant potential tumours ($n = 14$) – non-epithelial tumours and other gynaecological cancers, other cancers and metastatic cancers excluded Subgroups looked at stages I–IV
Morgante *et al.* (1999)[80]	RMI 1 threshold comparison	Malignancy ($n = 31$) included serous cystadenocarcinoma ($n = 14$), mucinous cystadenocarcinoma ($n = 6$), borderline ($n = 2$), clear cell carcinoma ($n = 2$) undifferentiated carcinoma ($n = 2$), granulosa cell carcinoma ($n = 1$) Kruckenberg ($n = 1$), immature teratoma ($n = 1$), endometrioid adenocarcinoma ($n = 1$) and metastatic carcinoma ($n = 1$)
Murala *et al.* (2014)[60]	IOTA group's simple ultrasound rules; and RMI 1	No histology details
Novotny *et al.* (2012)[86]	ROMA score using Abbott Diagnostics' tumour marker assay	No histology details
Piovano *et al.* (2016)[58]	IOTA group's simple ultrasound rules	Ovarian malignancy included epithelial ovarian cancer ($n = 45$), non-epithelial ovarian cancer ($n = 8$), borderline ($n = 22$) and metastatic carcinoma ($n = 9$)
Presl *et al.* (2012)[81]	ROMA score using Abbott Diagnostics' tumour marker assay	No histology details

TABLE 34 Study-level data for the histological details of malignant tumour diagnoses (*continued*)

Study (year of publication)	Test	Histology details for malignancy (*n*)
Ruiz de Gauna *et al.* (2015)[64]	IOTA group's simple ultrasound rules	Malignancy included primary ovarian cancer, borderline ovarian cancer and metastases; no numbers reported
Sayasneh *et al.* (2013)[62]	IOTA group's simple ultrasound rules and RMI 1	Malignancy included borderline ($n = 18$), serous cyst/adenocarcinoma ($n = 26$), mucinous cyst/adenocarcinoma ($n = 7$), endometrioid carcinoma ($n = 6$), clear cell carcinoma ($n = 5$), granulosa cell tumour ($n = 1$), transitional cell tumour ($n = 1$), signet ring cell adenocarcinoma ($n = 1$), peritoneal serous adenocarcinoma ($n = 1$), gastrointestinal adenocarcinomas ($n = 5$), malignant mixed Mullerian tumour ($n = 1$), large cell neuroendocrine carcinoma ($n = 1$) and endocrine tumour ($n = 1$)
Sayasneh *et al.* (2016)[46]	ADNEX model	Malignancy included primary ovarian cancer ($n = 116$), borderline ovarian cancer ($n = 42$) and metastatic ovarian cancer ($n = 24$)
Shulman *et al.* (2016)[104]	Overa (MIA2G) and ROMA score using Roche Diagnostics' tumour marker assay	Malignancy included epithelial ovarian cancer ($n = 150$), non-epithelial ovarian cancer ($n = 16$), borderline ovarian cancer ($n = 42$), metastases ($n = 23$) and non-ovarian malignancies ($n = 14$)
Silvestre *et al.* (2015)[55]	IOTA group's simple ultrasound rules	Malignancy included primary ovarian malignancies ($n = 15$), borderline ovarian malignancy ($n = 5$), metastases ($n = 5$) and other malignancies ($n = 7$)
Szubert *et al.* (2016)[42]	ADNEX model	Spain. Malignancy included primary ovarian cancer ($n = 26$), borderline ovarian cancer ($n = 3$) and metastatic ovarian cancer ($n = 5$) Poland. Malignancy included primary ovarian cancer ($n = 53$), borderline ovarian cancer ($n = 12$) and metastatic ovarian cancer ($n = 5$)
Tantipalakorn *et al.* (2014)[51]	IOTA group's simple ultrasound rules	Malignancy included primary ovarian cancer ($n = 62$), borderline ovarian cancer ($n = 12$), germ cell tumours ($n = 9$), sex cord–stromal tumour ($n = 6$), metastatic adenocarcinoma ($n = 10$) and other malignant tumours ($n = 8$)
Testa *et al.* (2014)[50]	IOTA group's simple ultrasound rules and RMI 1	Malignancy included primary invasive ovarian cancer ($n = 633$), borderline ovarian tumours ($n = 153$), metastatic ovarian cancer ($n = 126$) and rare primary invasive (e.g. dysgerminom granulosa cell tumour, yolk sac tumour or malignant treatoma) ($n = 68$)
Timmerman *et al.* (2010)[65]	IOTA group's Simple Rules and RMI 1	Malignancy included borderline ($n = 111$), primary invasive, stages I–IV and rare ($n = 373$) and metastatic ovarian cancer ($n = 58$)
Tingulstad *et al.* (1996)[76]	RMI 1 threshold comparison	Malignancy included ovarian cancer ($n = 51$), neurosarcoma ($n = 1$), leiomyosarcoma ($n = 1$), lymphoma ($n = 1$), Kruckenberg tumour ($n = 1$) and rectal cancer ($n = 1$)
Tinnangwattana *et al.* (2015)[47]	IOTA group's simple ultrasound rules	Malignancy included primary ovarian malignancies ($n = 13$), borderline ($n = 8$), metastases ($n = 5$) and other malignancies ($n = 3$)
Ulusoy *et al.* (2007)[74]	RMI 1 threshold comparison	Malignancy included primary ovarian cancers ($n = 84$), borderline ovarian cancers ($n = 15$) and metastases ($n = 7$)

continued

TABLE 34 Study-level data for the histological details of malignant tumour diagnoses (*continued*)

Study (year of publication)	Test	Histology details for malignancy (*n*)
Van Calster *et al.* (2014)[17]	ADNEX model	Malignancy included primary ovarian cancer ($n = 701$), borderline ovarian cancer ($n = 153$) and metastases ($n = 126$). IOTA phase 3 – validation data set
Van Gorp *et al.* (2012)[98]	ROMA score using Fujirebio Diagnostics' tumour marker assay and RMI 1	Malignancy included epithelial ovarian cancer stages I–IV ($n = 120$), non-epithelial ovarian cancer ($n = 4$) and metastatic ovarian cancer ($n = 25$)
Weinberger and Minar (2013)[53]	IOTA group's simple ultrasound rules	Malignancy included all invasive ovarian cancers and borderline tumours. No further details reported
Winarto *et al.* (2014)[99]	ROMA score using Abbott Diagnostics' tumour marker assay and RMI 1	Malignancy included serous cystadenocarcinoma ($n = 19$), endometrioid ($n = 14$), mucinous ($n = 8$), clear cell ($n = 7$), carcinosarcoma ($n = 2$) and borderline ($n = 17$)
Xu *et al.* (2016)[95]	ROMA score using Roche Diagnostics' tumour marker assay	Malignancy was described as epithelial ovarian cancer ($n = 210$), endometrioid ($n = 80$), serous ($n = 59$), papillary serous ($n = 15$), mucinous ($n = 6$), seromucinous ($n = 2$), clear cell ($n = 12$) and adenocarcinoma ($n = 36$)
Yamamoto *et al.* (2009)[72]	RMI 1 threshold comparison	Malignancy included primary ovarian cancer ($n = 29$), borderline ovarian cancer ($n = 8$) and tubal cancer ($n = 3$)
Yanaranop *et al.* (2016)[89]	ROMA score using Roche Diagnostics' tumour marker assay; and RMI 1	Malignancy included epithelial ovarian carcinoma ($n = 66$) and non-epithelial ovarian cancer ($n = 8$) Subgroups looked at epithelial and stages I–IV
Zhang *et al.* (2015)[68]	Overa (MIA2G)	Malignancy ($n = 72$) included stage I/II ($n = 19$) and LMP ($n = 13$)
Zhang *et al.* (2015)[102]	ROMA score using Roche Diagnostics' tumour marker assay	Malignancy was described as epithelial ovarian cancer ($n = 264$), serous ($n = 170$), mucinous ($n = 20$), endometrioid ($n = 25$), other kinds ($n = 13$) and unknown ($n = 36$) Subgroups looked at stages I–IV

LMP, low malignant potential.
a Frances Nixon, personal communication.

Appendix 5 Additional results

TABLE 35 Additional accuracy data for the ROMA score (accuracy using thresholds other than those recommended by the manufacturers)

Test	Study (year of publication)	Subgroup	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
ROMA score using Abbott Diagnostics' tumour marker assay	***Target condition: ovarian malignancies (undefined – not clear whether or not borderline tumours were included)***									
	[a]Clemente and Benitez (2015)[90]	All women	NR	4	5	14	39	62	44.4 (13.7 to 78.8)	73.6 (59.7 to 84.7)
		Premenopausal women	NR	2	3	11	32	48	40.0 (5.3 to 85.3)	74.4 (58.8 to 86.5)
		Postmenopausal women	NR	2	2	3	7	14	50.0 (6.8 to 93.2)	70.0 (34.8 to 93.9)
	[a]Novotny *et al.* (2012)[86]	Postmenopausal women	37.7	18	3	13	243	277	85.7 (63.7 to 97.0)	94.9 (91.5 to 97.3)
	Target condition: epithelial ovarian malignancies including borderline									
	Moore *et al.* (2011)[101]	All women	13.1/27.7	59	8	96	287	450	88.1 (77.8 to 94.7)	74.9 (70.3 to 79.2)
		Premenopausal women	13.1	13	3	60	173	249	81.3 (54.4 to 96.0)	74.2 (68.1 to 79.7)
		Postmenopausal women	27.7	46	5	36	114	201	90.2 (78.6 to 96.7)	76.0 (68.4 to 82.6)
	Target condition: epithelial ovarian malignancies (stage III/IV) – borderline and stage I/II tumours excluded									
	Moore *et al.* (2011)[101]	All women	13.1/27.7	34	0	96	287	417	100 (89.7 to 100)	74.9 (70.3 to 79.2)
		Premenopausal women	13.10	5	0	60	173	238	100 (47.8 to 100)	74.2 (68.1 to 79.7)
		Postmenopausal women	27.7	29	0	36	114	179	100 (88.1 to 100)	76.0 (68.4 to 82.6)
	Target condition: epithelial ovarian malignancies (stage I/II) – borderline and stage III/IV tumours excluded									
	Moore *et al.* (2011)[101]	All women	13.1/27.7	9	3	96	287	395	75.0 (42.8 to 94.5)	74.9 (70.3 to 79.2)
		Premenopausal women	13.1	3	0	60	173	236	100 (29.2 to 100)	74.2 (68.1 to 79.7)
		Postmenopausal women	27.7	6	3	36	114	159	66.7 (29.9 to 92.5)	25.3 (21.4 to 29.6)

Test	Study (year of publication)	Subgroup	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
ROMA score using Roche Diagnostics' tumour marker assay	***Target condition: epithelial ovarian malignancies excluding borderline***									
	Xu *et al.* (2016)[95]	Premenopausal women	13.40	58	49	23	241	371	54.2 (44.3 to 63.9)	91.3 (87.2 to 94.4)
		Postmenopausal women	18.70	69	34	6	41	150	67.0 (57.0 to 75.9)	87.2 (74.3 to 95.2)
	Target condition: epithelial ovarian malignancies (stage III/IV) – borderline and stage I/II tumours excluded									
	[a]Zhang *et al.* (2015)[102]	All women	11.4–29.9	143	16	72	276	507	89.9 (84.2 to 94.1)	79.3 (74.7 to 83.4)
		Premenopausal women	11.4	40	10	58	227	335	80.0 (66.3 to 90.0)	79.6 (74.5 to 84.2)
		Postmenopausal women	29.9	103	6	14	49	172	94.5 (88.4 to 98.0)	77.8 (65.5 to 87.3)
	Target condition: epithelial ovarian malignancies (stage I/II) – borderline and stage III/IV tumours excluded									
	Zhang *et al.* (2015)[102]	All women	11.4–29.9	49	15	72	276	412	76.6 (64.3 to 86.2)	79.3 (74.7 to 83.4)
		Premenopausal women	11.4	21	9	58	227	315	70.0 (50.6 to 85.3)	79.6 (74.5 to 84.2)
		Postmenopausal women	29.9	28	6	14	49	97	82.4 (65.5 to 93.2)	77.8 (65.5 to 87.3)

NR, not reported.
a 2 × 2 data were calculated (other studies reported 2 × 2 data).

TABLE 36 Accuracy of the ADNEX model at thresholds other than 10%

Study (year of publication)	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
All malignant tumours including borderline									
Sayasneh *et al.* (2016)[46]	1	182	0	377	51	610	Calculated	100 (97.4 to 100)	11.9 (9.1 to 15.5)
	3	182	0	297	131	610	Calculated	100 (97.4 to 100)	30.6 (26.3 to 35.3)
	5	180	2	200	228	610	Calculated	99 (94.9 to 99.8)	53.2 (48.2 to 58.1)
	15	172	10	106	322	610	Calculated	94.4 (90 to 97)	75.2 (70.7 to 79.2)
	20	165	17	89	339	610	Calculated	90.6 (85.2 to 94.1)	79.3 (75.1 to 83)
	30	157	25	69	359	610	Calculated	86.3 (80.4 to 90.6)	83.9 (80.1 to 87.2)
Van Calster *et al.* (2014)[17]	3	969	11	760	663	2403	Calculated	98.9 (98 to 99.4)	46.6 (44 to 49.2)
	5	964	16	578	845	2403	Calculated	98.4 (97.4 to 99.1)	59.4 (56.8 to 62)
	15	913	67	324	1099	2403	Calculated	93.2 (92.5 to 95.6)	77.2 (74.9 to 79.3)
Ovarian malignancies including borderline									
Joyeux *et al.* (2016)[43]	3	30	0	134	120	284	Calculated	100 (88.4 to 100)	47.2 (41 to 53.6)
	5	29	1	78	176	284	Calculated	96.6 (82.8 to 99.9)	69.2 (63.2 to 74.9)
	15	26	4	38	216	284	Calculated	86.6 (69.3 to 96.2)	85 (80 to 89.2)

TABLE 37 Accuracy of the IOTA group's simple ultrasound rules, through which inconclusive results were not classified

Study (year of publication)	Threshold	Subgroup	Index test variations	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline											
Alcazar *et al.* (2013)[52]	Malignant (inconclusives were excluded)	All women		29	4	6	231	270	Reported	87.9 (72.4 to 95.2)	97.5 (94.6 to 98.8)
Ruiz de Gauna *et al.* (2015)[64]		Centre A		27	0	4	62	93	Reported	100 (87.5 to 100)	93.9 (85.4 to 97.6)
		Centre B		11	2	4	92	109	Reported	84.6 (57.8 to 95.7)	95.8 (89.8 to 98.4)
Silvestre *et al.* (2015)[55]		All women		32	0	8	26	66	Reported	100 (89.1 to 100)	76.5 (58.8 to 89.3)
Tantipalakorn *et al.* (2014)[51]		All women		88	19	10	202	319 (masses)	Reported	82.2 (75 to 89.5)	95.3 (92.4 to 98.1)
Tinnangwattana *et al.* (2015)[47]		All women		25	3	11	55	94	Reported	89.3 (77.8 to 100)	83.3 (74.3 to 92.3)
Piovano *et al.* (2016)[58]	Malignant (inconclusives were classified by expert SA) and ROMA score of > 11.4%/29.9%	All women	+ ROMA	76	8	61	246	391	Calculated	90.5 (82.1 to 95.8)	80.1 (75.2 to 84.4)
		Postmenopausal women	+ ROMA	58	5	25	82	170	Calculated	92.0 (85.0 to 99.0)	77.0 (69.0 to 85.0)
		Premenopausal women	+ ROMA	18	3	36	164	221	Calculated	86.0 (71.0 to 100)	82.0 (77.0 to 87.0)
	Malignant (inconclusives were classified by expert SA) and HE4 level of ≥ 70/140 pmol/l	All women	+ HE4	73	11	42	265	391	Calculated	86.9 (77.8 to 93.3)	86.3 (82.0 to 90.0)
		Postmenopausal women	+ HE4	55	8	18	89	170	Calculated	87.0 (79.0 to 96.0)	83.0 (76.0 to 90.0)
		Premenopausal women	+ HE4	18	3	24	176	221	Calculated	86.0 (71.0 to 100)	88.0 (83.0 to 92.0)

continued

TABLE 37 Accuracy of the IOTA group's simple ultrasound rules, through which inconclusive results were not classified (*continued*)

Study (year of publication)	Threshold	Subgroup	Index test variations	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	2 × 2 data	Sensitivity, % (95% CI)	Specificity, % (95% CI)
	Malignant on IOTA group's simple ultrasound rules (inconclusives were classified by expert SA) and CA125 level of ≥ 35 U/ml	All women	+ CA125	76	8	98	209	391	Calculated	90.5 (82.1 to 95.8)	68.1 (62.5 to 73.3)
		Postmenopausal women	+ CA125	58	5	26	81	170	Calculated	92.0 (85.0 to 99.0)	76.0 (68.0 to 84.0)
		Premenopausal women	+ CA125	18	3	72	128	221	Calculated	86.0 (71.0 to 100)	64.0 (57.0 to 61.0)
Ruiz de Gauna *et al.* (2015)[64]	Malignant (inconclusives were classified by expert SA; final ratings of unclassifiable were treated as malignant)	Centre A		31	0	9	74	114	Reported	100 (88.8 to 100)	89.2 (80.4 to 94.9)
		Centre B		13	2	13	105	133	Reported	86.7 (59.5 to 98.3)	89 (81.9 to 94.0)
Target condition: ovarian malignancies including borderline											
Fathallah *et al.* (2011)[63]	Malignant (inconclusives were excluded)	All women		8	3	3	95	109	Reported	73.0 (45.0 to 100)	97.0 (94.0 to 100)
Weinberger and Minar (2013)[53]	Malignant (handling of inconclusives was unclear)	All women		118	7	16	206	347	Calculated	94.0 (88.8 to 97.7)	93.0 (88.6 to 95.8)
Target condition: ovarian malignancies excluding borderline											
Weinberger *et al.* (2013)[53]	Malignant (handling of inconclusives was unclear)	All women		99	2	16	222	323	Calculated	98.0 (93.0 to 99.8)	93.0 (89.3 to 96.1)
Target condition: ovarian malignancies (undefined – not clear whether borderline tumours were included)											
Baker *et al.* (2013)[66]	Malignant (inconclusives were excluded)	Premenopausal women		2	0	5	21	28	Calculated	100 (15.8 to 100)	80.8 (60.6 to 93.4)
Target condition: ovarian borderline tumours											
Weinberger *et al.* (2013)[53]	Malignant (handling of inconclusives was unclear)	All		19	5	16	222	262	Calculated	79.2 (57.8 to 92.9)	93.3 (89.3 to 96.1)

SA, subjective assessment.

TABLE 38 Accuracy of the RMI 1 at decision thresholds other than 200 and 250

Study (year of publication)	CA125 assay	Ultrasound details	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
All malignant tumours including borderline										
Davies *et al.* (1993)[79]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer	25	36	2	39	48	124	94.7 (82.3 to 99.4)	55.2 (44.1 to 65.9)
Jacobs *et al.* (1990)[78]	RIA (Abbott Diagnostics)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer		41	0	37	61	139	100 (91.4 to 100)	62.2 (51.9 to 71.8)
Morgante *et al.* (1999)[80]	RIA (Centocor)	A Siemens Sonoline SL2 was used with a 3.5-MHz transabdominal sectorial probe and a 5.0- to 7.5-MHz transvaginal probe		30	1	37	56	124	96.8 (83.3 to 99.9)	60.2 (49.5 to 70.2)
Tingulstad *et al.* (1996)[76]	IMx™ (Abbott Diagnostics)	Transvaginal with transabdominal as needed		51	5	37	80	173	91.1 (80.4 to 97.0)	68.4 (59.1 to 76.7)
Ulusoy *et al.* (2007)[74]	NR	Toshiba Sonolayer SSA-270A and/or a Siemens Sonoline G50 with 3.75-MHz and 5-MHz abdominal convex transducers and/or 5-MHz and 9-MHz endovaginal probes		100	6	136	54	296	94.3 (88.1 to 97.9)	28.4 (22.1 to 35.4)
Summary estimates									**94.9 (91.5 to 97.2)**	**51.1 (47.0 to 55.2)**
Davies *et al.* (1993)[79]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer	50	36	2	28	59	124	94.7 (82.3 to 99.4)	67.8 (56.9 to 77.4)
Jacobs *et al.* (1990)[78]	RIA (Abbott Diagnostics)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer		39	2	23	75	139	95.1 (83.5 to 99.4)	76.5 (66.9 to 84.5)
Lou *et al.* (2010)[73]	NR	NR		49	12	36	126	223	80.3 (68.2 to 89.4)	77.8 (70.6 to 83.9)
Morgante *et al.* (1999)[80]	RIA (Centocor)	A Siemens Sonoline SL2 was used with a 3.5-MHz transabdominal sectorial probe and a 5.0- to 7.5-MHz transvaginal probe		29	2	23	70	124	93.5 (78.6 to 99.2)	75.3 (65.2 to 83.6)
Tingulstad *et al.* (1996)[76]	IMx™ (Abbott Diagnostics)	Transvaginal with transabdominal as needed		49	7	22	95	173	87.5 (75.9 to 94.8)	81.2 (72.9 to 87.8)
Ulusoy *et al.* (2007)[74]	NR	Toshiba Sonolayer SSA-270A and/or a Siemens Sonoline G50 with 3.75-MHz and 5-MHz abdominal convex transducers and/or 5-MHz and 9-MHz endovaginal probes		96	10	106	84	296	90.6 (83.3 to 95.4)	44.2 (37 to 51.6)

continued

TABLE 38 Accuracy of the RMI 1 at decision thresholds other than 200 and 250 (*continued*)

Study (year of publication)	CA125 assay	Ultrasound details	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Summary estimates									**89.5 (85.7 to 92.6)**	**68.1 (64.7 to 71.5)**
Davies *et al.* (1993)[79]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer	75 or 80	33	4	18	69	124	89.2 (74.6 to 97.0)	79.3 (69.3 to 87.3)
Jacobs *et al.* (1990)[78]	RIA (Abbott Diagnostics)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer		38	3	15	83	139	92.7 (80.2 to 98.5)	84.7 (76 to 91.2)
Morgante *et al.* (1999)[80]	RIA (Centocor)	A Siemens Sonoline SL2 was used with a 3.5-MHz transabdominal sectorial probe and a 5.0- to 7.5-MHz transvaginal probe		25	6	19	74	124	80.6 (62.5 to 92.5)	79.6 (69.9 to 87.2)
Tingulstad *et al.* (1996)[76]	IMx™ (Abbott Diagnostics)	Transvaginal with transabdominal as needed		44	12	14	107	173	78.6 (65.6 to 88.4)	88.4 (81.3 to 93.5)
Summary estimates									**84.8 (78.5 to 89.9)**	**83.5 (79.4 to 87.0)**
Davies *et al.* (1993)[79]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer	100	32	5	13	74	124	86.5 (71.2 to 95.5)	85.1 (75.8 to 91.8)
Jacobs *et al.* (1990)[78]	RIA (Abbott Diagnostics)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer		35	6	12	86	139	85.4 (70.8 to 94.4)	87.8 (79.6 to 93.5)
Lou *et al.* (2010)[73]	NR	NR		45	16	18	144	223	73.8 (60.9 to 84.2)	88.9 (83.0 to 93.3)
Morgante *et al.* (1999)[80]	RIA (Centocor)	A Siemens Sonoline SL2 was used with a 3.5-MHz transabdominal sectorial probe and a 5.0- to 7.5-MHz transvaginal probe		24	7	9	84	124	77.4 (58.9 to 90.4)	90.3 (82.4 to 95.5)
Tingulstad *et al.* (1996)[76]	IMx™ (Abbott Diagnostics)	Transvaginal with transabdominal as needed		44	12	13	108	173	78.6 (65.6 to 88.4)	89.3 (82.3 to 94.2)

Study (year of publication)	CA125 assay	Ultrasound details	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Summary estimates									**79.6 (73.8 to 84.7)**	**88.4 (85.5 to 90.9)**
Morgante *et al.* (1999)[80]	RIA (Centocor)	A Siemens Sonoline SL2 was used with a 3.5-MHz transabdominal sectorial probe and a 5.0- to 7.5-MHz transvaginal probe	120 or 125	23	8	7	86	124	74.2 (55.4 to 88.1)	92.5 (85.1 to 96.9)
Tingulstad *et al.* (1996)[76]	IMx™ (Abbott Diagnostics)	Transvaginal with transabdominal as needed		44	12	12	109	173	78.6 (65.6 to 88.4)	90.1 (83.3 to 94.8)
Ulusoy *et al.* (2007)[74]	NR	Toshiba Sonolayer SSA-270A and/or a Siemens Sonoline G50 with 3.75-MHz and 5-MHz abdominal convex transducers and/or 5-MHz and 9-MHz endovaginal probes		86	20	59	131	296	81.1 (72.4 to 88.1)	68.9 (61.8 to 75.4)
Summary estimates									**79.3 (72.9 to 84.8)**	**80.7 (76.5 to 84.4)**
Davies *et al.* (1993)[79]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer	150	30	7	13	74	124	81.1 (64.8 to 92.0)	85.1 (75.8 to 91.8)
Jacobs *et al.* (1990)[78]	RIA (Abbott Diagnostics)	Diasonics DS 1 sector scanner with a 3.5-MHz transducer		35	6	6	92	139	85.4 (70.8 to 94.4)	93.9 (87.2 to 97.7)
Lou *et al.* (2010)[73]	NR	NR		37	24	8	154	223	60.7 (47.3 to 72.9)	95.1 (90.5 to 97.8)
Morgante *et al.* (1999)[80]	RIA (Centocor)	A Siemens Sonoline SL2 was used with a 3.5-MHz transabdominal sectorial probe and a 5.0- to 7.5-MHz transvaginal probe		20	11	6	87	124	64.5 (45.4 to 80.8)	93.5 (86.5 to 97.6)
Tingulstad *et al.* (1996)[76]	IMx™ (Abbott Diagnostics)	Transvaginal with transabdominal as needed		43	13	7	110	173	76.8 (63.6 to 87.0)	94.0 (88.1 to 97.6)
Ulusoy *et al.* (2007)[74]	NR	Toshiba Sonolayer SSA-270A and/or a Siemens Sonoline G50 with 3.75-MHz and 5-MHz abdominal convex transducers and/or 5-MHz and 9-MHz endovaginal probes		81	25	42	148	296	76.4 (67.2 to 84.1)	77.9 (71.3 to 83.6)

continued

TABLE 38 Accuracy of the RMI 1 at decision thresholds other than 200 and 250 (*continued*)

Study (year of publication)	CA125 assay	Ultrasound details	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Summary estimates									**74.1 (69.0 to 78.7)**	**89.0 (86.6 to 91.2)**
Lou *et al.* (2010)[73]	NR	NR	300	33	28	2	160	223	54.1 (40.8 to 66.9)	98.8 (95.6 to 99.9)
Ulusoy *et al.* (2007)[74]	NR	Toshiba Sonolayer SSA-270 A and/or a Siemens Sonoline G50 with 3.75-MHz and 5-MHz abdominal convex transducers and/or 5-MHz and 9-MHz endovaginal probes	500	57	49	12	178	296	53.8 (43.8 to 63.5)	93.7 (89.2 to 96.7)
All malignant tumours excluding borderline										
Aktürk *et al.* (2011)[71]	Electrochemiluminescence immunoassay (Roche Diagnostics)	Siemens transvaginal 7.5-MHz transducer	50	17	3	33	47	100	85 (62.1 to 96.8)	58.8 (47.2 to 69.6)
			100	15	5	13	67	100	75 (50.9 to 91.3)	83.8 (73.8 to 91.1)
			150	15	5	12	68	100	75 (50.9 to 91.3)	85.0 (75.3 to 92.0)
			300	9	11	2	78	100	45 (23.1 to 68.5)	97.5 (91.3 to 99.7)
			350	9	11	2	78	100	45 (23.1 to 68.5)	97.5 (91.3 to 99.7)
			400	6	14	2	78	100	30 (11.9 to 54.3)	97.5 (91.3 to 99.7)

Study (year of publication)	CA125 assay	Ultrasound details	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Manjunath *et al.* (2001)[75]	Microparticle EIA (Abbott Diagnostics)	The ultrasound was performed vaginally by a 5-MHz transducer (Ultramark 4 PLUS, Advanced Technology Laboratories, Signal Hill, CA, USA) and extended to the transabdominal approach with full bladder if the mass was large	25	85	8	27	28	148	91.4 (83.8 to 96.2)	50.9 (37.1 to 64.6)
			50	75	18	21	34	148	80.6 (71.1 to 88.1)	61.8 (47.7 to 74.6)
			80	74	19	18	37	148	79.6 (69.9 to 87.2)	67.3 (53.3 to 79.3)
			100	74	19	14	41	148	79.6 (69.9 to 87.2)	74.5 (61.0 to 85.3)
			125	73	20	11	44	148	78.5 (68.8 to 86.3)	80.0 (67.0 to 89.6)
			150	72	21	9	46	148	77.4 (67.6 to 85.4)	83.6 (71.2 to 92.2)
			300	60	33	3	52	148	64.5 (53.9 to 74.2)	94.5 (84.9 to 98.9)
			350	58	35	3	52	148	62.4 (51.7 to 72.2)	94.5 (84.9 to 98.9)
			400	57	36	3	52	148	61.3 (50.6 to 71.2)	94.5 (84.9 to 98.9)

continued

TABLE 38 Accuracy of the RMI 1 at decision thresholds other than 200 and 250 (*continued*)

Study (year of publication)	CA125 assay	Ultrasound details	Threshold	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
All malignant tumours (undefined – not clear whether or not borderline tumours were included)										
Asif *et al.* (2004)[77]	IA (IMMULITE)	NR	25	54	1	15	30	100	98.2 (90.3 to 100)	66.7 (51.0 to 80.0)
			50	53	2	10	35	100	96.4 (87.5 to 99.6)	77.8 (62.9 to 88.8)
			75	52	3	8	37	100	94.5 (84.9 to 98.9)	82.2 (67.9 to 92.0)
			100	49	6	7	38	100	89.1 (77.8 to 95.9)	84.4 (70.5 to 93.5)
			125	48	7	5	40	100	87.3 (75.5 to 94.7)	88.9 (75.9 to 96.3)
			150	47	8	4	41	100	85.5 (73.3 to 93.5)	91.1 (78.8 to 97.5)
			175	47	8	4	41	100	85.5 (73.3 to 93.5)	91.1 (78.8 to 97.5)
			190	47	8	4	41	100	85.5 (73.3 to 93.5)	91.1 (78.8 to 97.5)
			300	40	15	0	45	100	72.7 (59.0 to 83.9)	100 (92.1 to 100)
Ovarian malignancies including borderline										
Yamamoto *et al.* (2009)[72]	Elecsys CA125 II	Transvaginal (6.0-MHz transducer), with transabdominal as indicated	100	39	1	65	148	253	97.5 (86.8 to 99.9)	69.5 (62.8 to 75.6)
			150	34	6	36	177	253	85.0 (70.2 to 94.3)	83.1 (77.4 to 87.9)
			300	27	13	18	195	253	67.5 (50.9 to 81.4)	91.5 (87.0 to 94.9)

IA, immnunoassay; NR, not reported; RIA, radioimmunoassay.

TABLE 39 Comparative accuracy of the ROMA score using Fujirebio Diagnostics' tumour marker assay vs. the RMI 1

Study (year of publication)	Subgroup	ROMA threshold (%)	TP	FN	FP	TN	Total	Sensitivity, % (95% CI)	Specificity, % (95% CI)	RMI 1	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline																	
Van Gorp *et al.* (2012)[98]	All women	12.5/14.4	127	23	52	172	374	84.7 (77.9 to 90)	76.8 (70.7 to 82.2)	200	114	36	17	207	374	76.0 (68.4 to 82.6)	92.4 (88.1 to 95.5)
	Premenopausal women	12.5	26	13	17	122	178	66.7 (49.8 to 80.9)	87.8 (81.1 to 92.7)	200	25	14	6	133	178	64.1 (47.2 to 78.8)	95.7 (90.8 to 98.4)
	Postmenopausal women	14.4	101	10	35	50	196	91 (84.1 to 95.6)	58.8 (47.6 to 69.4)	200	89	22	11	74	196	80.2 (71.5 to 87.1)	87.1 (78 to 93.4)

TABLE 40 Accuracy of the ROMA score using Fujirebio Diagnostics' tumour marker assay at the manufacturer's recommended thresholds

Study (year of publication)	Subgroup	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: all malignant tumours including borderline									
[a]Langhe *et al.* (2013)[94]	All women	12.5/14.4	129	47	31	170	377	73.3 (66.1 to 79.7)	84.6 (78.8 to 89.3)
Van Gorp *et al.* (2012)[98]	All women	12.5/14.4	127	23	52	172	374	84.7 (77.9 to 90.0)	76.8 (70.7 to 82.1)
Summary estimates								**78.5 (73.7 to 82.9)**	**80.5 (76.4 to 84.1)**
[a]Langhe *et al.* (2013)[94]	Premenopausal women	12.5	23	22	6	81	132	51.1 (35.8 to 66.3)	93.1 (85.6 to 97.4)
Van Gorp *et al.* (2012)[98]	Premenopausal women	12.5	26	13	17	122	178	66.7 (49.8 to 80.9)	87.8 (81.1 to 92.7)
Summary estimates								**58.3 (47.1 to 69.0)**	**89.8 (85.1 to 93.4)**
[a]Langhe *et al.* (2013)[94]	Postmenopausal women	14.4	105	26	25	89	245	80.2 (72.3 to 86.6)	78.1 (69.4 to 85.3)
Van Gorp *et al.* (2012)[98]	Postmenopausal women	14.4	101	10	35	50	196	91 (84.1 to 95.6)	58.8 (47.6 to 69.4)
Summary estimates								**85.1 (80.0 to 89.4)**	**69.8 (63.0 to 76.1)**

a 2 × 2 data were calculated.

TABLE 41 Accuracy of the ROMA score using Abbott Diagnostics' ARCHITECT tumour marker assay at the manufacturer's recommended thresholds (unclear whether or not borderline tumours were included in the analysis)

Study (year of publication)	Subgroup	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: ovarian malignancies (undefined – not clear whether or not borderline tumours were included)									
Li *et al.* (2016)[96]	All women	7.4/25.3	166	24	141	586	917	87.4 (81.8 to 91.7)	80.1 (77.0 to 83.0)
[a]Presl *et al.* (2012)[81]	All women	7.3/26.3	25	5	72	450	552	83.3 (65.3 to 94.4)	86.2 (82.9 to 89.0)
Summary estimates								**86.8 (81.6 to 91.0)**	**82.7 (80.5 to 84.8)**
Li *et al.* (2016)[96]	Premenopausal women	7.4	96	12	136	501	745	88.9 (81.4 to 94.1)	78.6 (75.3 to 81.8)
[a]Presl *et al.* (2012)[81]	Premenopausal women	7.30	5	4	44	243	296	55.6 (21.2 to 86.3)	84.7 (80.0 to 88.6)
Summary estimates								**86.3 (78.7 to 92.0)**	**80.5 (77.8 to 83.0)**
Li *et al.* (2016)[96]	Postmenopausal women	25.3	70	12	5	85	172	85.4 (75.8 to 92.2)	94.4 (87.5 to 98.2)
Novotny *et al.* (2012)[86,a]	Postmenopausal women	26.3	20	1	31	225	277	95.2 (76.2 to 99.9)	87.9 (83.3 to 91.6)
[a]Presl *et al.* (2012)[81]	Postmenopausal women	26.3	20	1	28	207	256	95.2 (76.2 to 99.9)	88.1 (83.2 to 91.9)
Summary estimates								**88.7 (81.8 to 93.7)**	**89.0 (86.2 to 91.4)**

a 2 × 2 data were calculated (other studies reported 2 × 2 data).

TABLE 42 Accuracy of the ROMA score using Roche Diagnostics' Elecsys tumour marker assay at the manufacturer's recommended thresholds (unclear whether or not borderline tumours were included in the analysis)

Study (year of publication)	Subgroup	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: ovarian malignancies (undefined – not clear whether or not borderline tumours were included)									
[a]Zhang *et al.* (2015)[102]	All women	11.4/29.9	224	40	73	275	612	84.8 (79.9 to 88.9)	79.0 (74.4 to 83.2)
	Premenopausal women	11.4	70	25	59	226	380	73.7 (63.6 to 82.2)	79.3 (74.1 to 83.9)
	Postmenopausal women	29.9	154	15	14	49	232	91.1 (85.8 to 94.9)	77.8 (65.5 to 87.3)

a 2 × 2 data were calculated.

TABLE 43 Accuracy of the ADNEX model (unclear whether or not borderline tumours were included in the analysis)

Study (year of publication)	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Moffatt *et al.* (2016)[45]	10	4	2	29	46	81	66.7 (22.3 to 95.7)	61.3 (64.4 to 81.6)

TABLE 44 Additional accuracy data for the RMI 1 score (unclear whether or not borderline tumours were included in the analysis)

		Threshold											
		200						250					
Study (year of publication)	Total number	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
All malignant tumours (undefined – not clear whether or not borderline tumours were included)													
Asif *et al.* (2004)[77]	100	47	8	3	42	85 (NR)	93 (NR)	40	15	2	43	72 (NR)	95 (NR)

NR, not reported.

TABLE 45 Accuracy of the ROMA score using Roche Diagnostics' tumour marker assays at the manufacturer's recommended thresholds (using unclear inclusion of borderline tumours)

Study (year of publication)	Subgroup	Threshold (%)	TP, *n*	FN, *n*	FP, *n*	TN, *n*	Total, *n*	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Target condition: ovarian malignancies (undefined – not clear whether or not borderline tumours were included)									
[a]Zhang *et al.* (2015)[102]	All women	11.4/29.9	224	40	73	275	612	84.8 (79.9 to 88.9)	79.0 (74.4 to 83.2)
	Premenopausal women	11.4	70	25	59	226	380	73.7 (63.6 to 82.2)	79.3 (74.1 to 83.9)
	Postmenopausal women	29.9	154	15	14	49	232	91.1 (85.8 to 94.9)	77.8 (65.5 to 87.3)

a 2 × 2 data were calculated (other studies reported 2 × 2 data).

Appendix 6 Cost calculations for risk scores

Risk assessment tool cost calculations

Test	Cost (£)				
	Test cost per kit	Sum of HE4 test-related costs (capital, other, personnel as per below)	Ultrasound	CA125	Total cost for risk score
ROMA score using Abbott Diagnostics' ARCHITECT[a]	21.33	6.64	76.75	25.58	130.31
ROMA score using Roche Diagnostics' Elecsys[a]	15.95	7.81	76.75	25.58	126.09
Vermillion Overa (MIA2G)[a]	99.00	–	76.75	–	175.80
IOTA group's simple ultrasound rules	–	–	76.75	–	76.75
IOTA group's ADNEX model	–	–	76.75	25.58	102.34
RMI 1	–	–	76.75	25.58	102.34

a Manufacturers stated that final costs may be subject to volume-based discounts.

Risk assessment tool components: cost breakdown

Risk assessment tool component	Cost per test kit/ultrasound (£)
Serum CA125	25.58
TVS	76.75
Abbott Diagnostics' ARCHITECT HE4	21.33
Roche Diagnostics' Elecsys HE4	15.95
Vermillion Overa (MIA2G)	99.00

Capital cost calculation items for Abbott Diagnostics and Roche Diagnostics' HE4 test	Capital cost items for HE4 tests	Per year (annuitised)	Cost per test (annuitised)
Costs of LUMIPULSE (average of G1200 and G600II)	£56,432.00	£6785.46	£1.92
Resale value	0	–	–
Lifetime of analyser equipment	10 years	–	–
Number of tests per year on one analyser (full capacity)	3542	–	–

Other cost items for Abbott Diagnostics and Roche Diagnostics' HE4 tests	Cost (£)		
	Item	Per year	Per test
Quality control for Abbott Diagnostics' HE4 test (1.5 times per year)	87.52	131.28	0.04
Quality control for Roche Diagnostics' HE4 test (12 times per year)	354.37	4252.44	1.20
Maintenance (per year, but not in the first year), taken from Fujirebio Diagnostics and assumed to be the same for Abbott Diagnostics and Roche Diagnostics in the absence of other information	3819.51	3437.56	0.97
Calibration (six times per year) Abbott Diagnostics and Roche Diagnostics, taken from Roche Diagnostics, assumed to be the same for Abbott Diagnostics in the absence of other information	566.98	3401.88	0.96
Shipment (per month), taken from Fujirebio Diagnostics and assumed to be the same for Roche Diagnostics in the absence of other information	0.26	3.12	0.001

Personnel cost items for Abbott Diagnostics and Roche Diagnostics' HE4 tests	Personnel cost	
	Item	Per test
Personnel time to prepare and perform test	0.05 hours	–
Personnel costs to prepare and perform test (per hour)	£55.16	£2.76

Appendix 7 Test accuracy estimates used for scenario and subgroup analyses

Comparison of different Risk of Malignancy 1 thresholds

Threshold	Sensitivity, % (SE)	Specificity, % (SE)	Source (systematic review, see *Appendix 5*)
250	64.4 (1.4)	91.8 (0.7)	Summary estimate derived from all studies, six published studies[73,74,76,78–80] and one unpublished study[a] that reported data for the RMI 1 (threshold of 250) and the target condition 'all malignant tumours'
200	68.1 (0.9)	90.1 (0.5)	Summary estimate derived from all studies, 12 published studies[44,48,50,62,73,74,76,78–80,98,103] and one unpublished study that reported data for the RMI 1 (threshold of 200) and the target condition 'all malignant tumours'
25	94.9 (1.5)	51.1 (2.1)	Summary estimate derived from all RMI 1 threshold comparison studies that reported data for the relevant threshold[74,76,78–80]
50	89.5 (1.8)	68.1 (1.7)	Summary estimate derived from all RMI 1 threshold comparison studies that reported data for the relevant threshold[73,74,76,78–80]
100	79.6 (2.8)	88.4 (1.4)	Summary estimate derived from all RMI 1 threshold comparison studies that reported data for the relevant threshold[73,76,78–80]
150	73.0 (3.1)	92.8 (1.1)	Summary estimate derived from all RMI 1 threshold comparison studies that reported data for the relevant threshold[74,76,78–80]
300	53.8 (5.0)	98.8 (1.1)	Estimate from one RMI 1 threshold comparison study that reported data for this threshold[73]

a Frances Nixon, personal communication.

Premenopausal subgroup

Test	Sensitivity, % (SE)	Specificity, % (SE)	Source (systematic review, see *Chapter 3*)
RMI 1 threshold of 250	64.4 (1.4)	91.8 (0.7)	No data available (sensitivity and specificity estimates for all participants used)
ROMA score using Abbott Diagnostics' ARCHITECT	52.4 (11.4)	90.1 (2.7)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[103] (see *Table 7*)
ROMA score using Roche Diagnostics' Elecsys	90.0 (11.3)	82.0 (3.6)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[97] (see *Table 10*)
Overa (MIA2G) from Vermillion	90.3 (5.5)	71.4 (2.9)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[70] (see *Table 17*)
IOTA group's simple ultrasound rules (inconclusive results treated as malignant)	94.5 (1.1)	79.3 (1.1)	Summary estimate derived from the four studies that reported subgroup data for the target condition 'all malignant tumours'[44,49,50,62] (see *Table 12*)
IOTA group's ADNEX model	97.0[a] (2.9)	71.0 (4.8)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[44] (see *Table 11*)
RMI 1 threshold of 200	53.3 (2.3)	93.5 (0.7)	Summary estimate derived from all five studies that reported subgroup data for the target condition 'all malignant tumours'[44,50,62,98,103]

a A weakly informative prior (alpha = 1; beta = 1) was used to handle 'zero events/counts issue' in the data obtained from Meys *et al.*[150] (indicating 100% sensitivity).

Postmenopausal subgroup

Test	Sensitivity, % (SE)	Specificity, % (SE)	Source (systematic review, see *Chapter 3*)
RMI 1 threshold 250	64.4 (1.4)	91.8 (0.7)	No data available (sensitivity and specificity estimates for all participants used)
ROMA score using Abbott Diagnostics' ARCHITECT	92.6 (6.0)	79.2 (9.0)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[103] (see *Table 8*)
ROMA score using Roche Diagnostics' Elecsys	78.6 (5.8)	76.1 (6.1)	Sensitivity and specificity estimates taken form the only study to report subgroup data for the target condition 'all malignant tumours'[97] (see *Table 11*)
Overa Vermillion (MIA2G)	91.8 (3.6)	65.4 (3.8)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[70] (see *Table 18*)
IOTA group's simple ultrasound rules (inconclusive results treated as malignant)	95.4 (0.8)	67.3 (1.9)	Summary estimate derived from the four studies that reported subgroup data for the target condition 'all malignant tumours'[44,49,50,62] (see *Table 13*)
IOTA group's ADNEX model	98.0 (2.3)	54.0 (4.8)	Sensitivity and specificity estimates taken from the only study to report subgroup data for the target condition 'all malignant tumours'[44] (see *Table 12*)
RMI threshold 200	79.4 (1.4)	79.2 (1.5)	Summary estimate derived from all five studies that reported subgroup data for RMI 1 (threshold of 200) and the target condition 'all malignant tumours'[44,50,62,98,103]

Appendix 8 Deterministic one-way sensitivity analyses

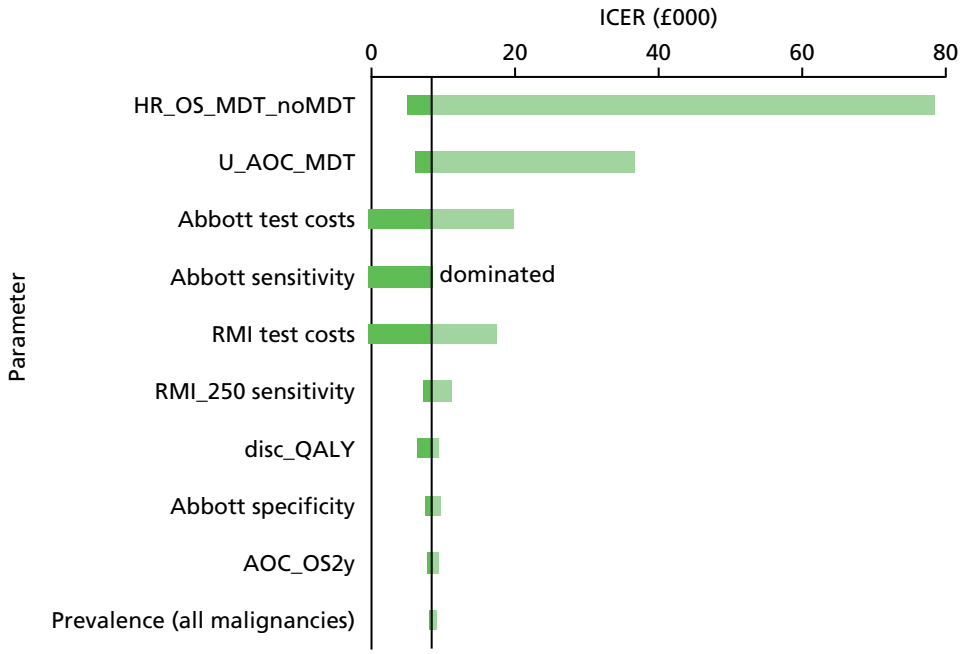

FIGURE 16 Top 10 influential parameters in the comparison of the RMI 1 (threshold of 250) with ROMA score using Abbott Diagnostics' ARCHITECT. AOC_OS2y, the overall survival estimate for patients with AOC at 2 years; disc_QALY, the discount rate used for QALYs/costs; HR_OS_MDT_noMDT, the overall survival HR for women referred to SMDT compared with those women not referred to SMDT; U_AOC_MDT, the utility associated with advanced/early ovarian cancer when the woman had been referred to SMDT.

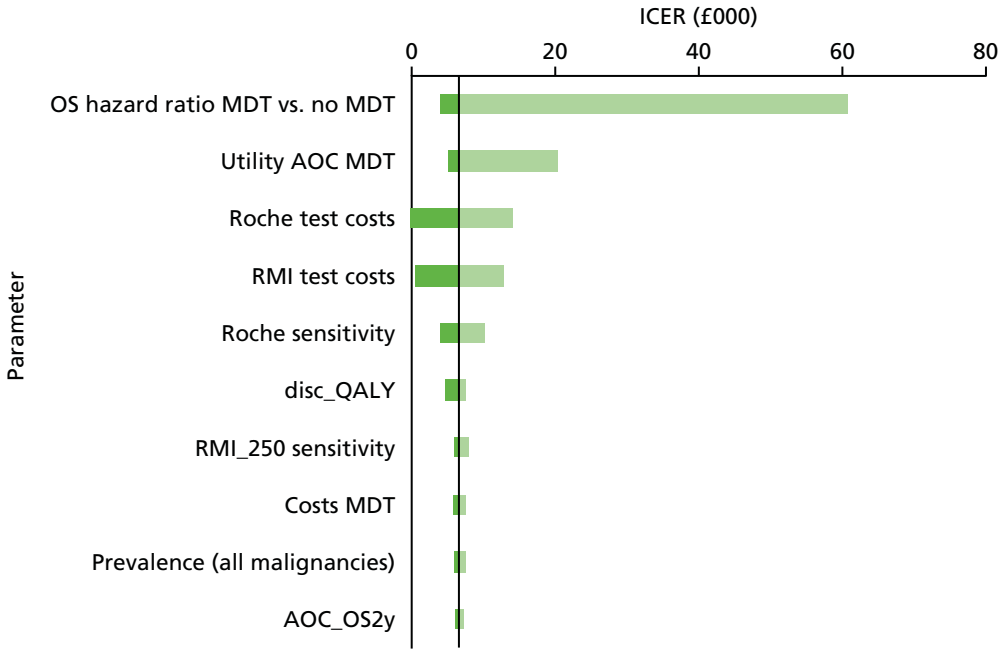

FIGURE 17 Top 10 influential parameters in the comparison of the RMI 1 (threshold of 250) with ROMA score using Roche Diagnostics' Elecsys. AOC_OS2y, the overall survival estimate for women with AOC at 2 years; disc_QALY, the discount rate used for QALYs/costs; OS, overall survival.

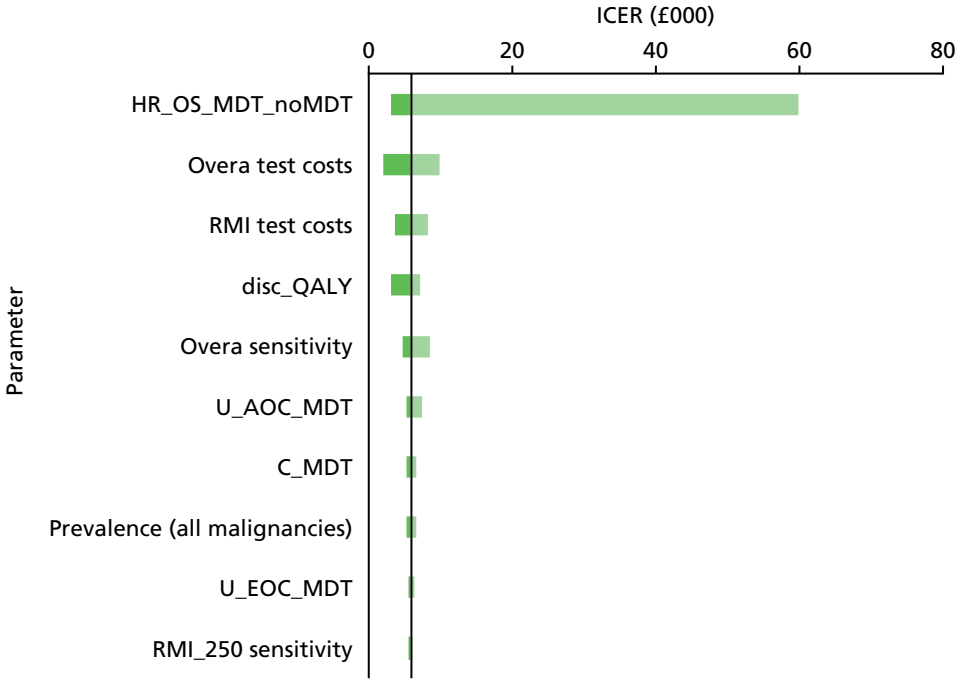

FIGURE 18 Top 10 influential parameters in the comparison of the RMI 1 (threshold of 250) with Overa [MIA2G (threshold of 5 units)]. C_MDT, costs associated with SMDT; disc_QALY, the discount rate used for QALYs/costs; HR_OS_MDT_noMDT, the overall survival HR for women referred to SMDT compared with those women not referred to SMDT; U_AOC_MDT/U_EOC_MDT, the utility associated with advanced/early ovarian cancer when the woman had been referred to SMDT.

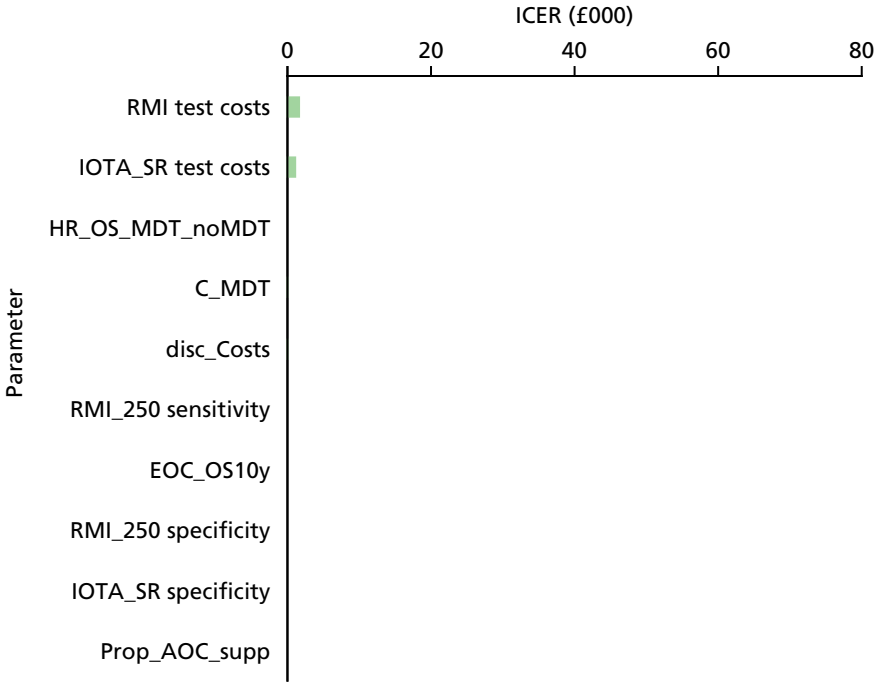

FIGURE 19 Top 10 influential parameters in the comparison of the RMI 1 (threshold of 250) with the IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant). C_MDT, costs associated with SMDT; disc_QALY, the discount rate used for QALYs/costs; EOC_OS10y, the overall survival estimate for women with early ovarian cancer at 10 years; HR_OS_MDT_noMDT, the overall survival HR for women referred to SMDT compared with those women not referred to SMDT; Prop_AOC_supp, proportion of women with AOC receiving only supportive care (and no surgery).

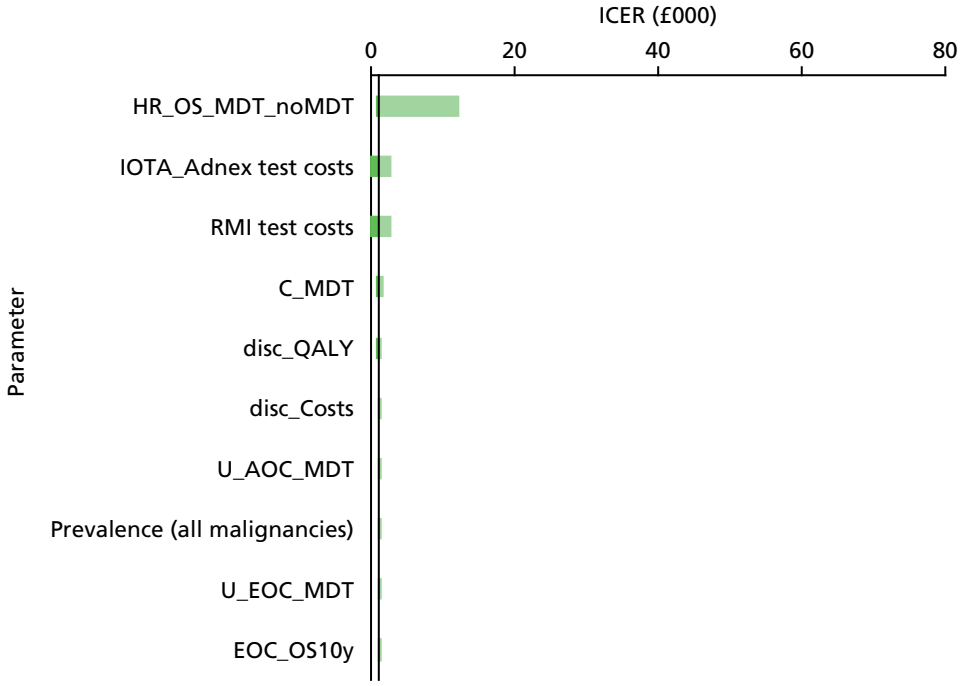

FIGURE 20 Top 10 influential parameters in the comparison of the RMI 1 (threshold of 250) with the IOTA group's ADNEX model (threshold of 10%). C_MDT, costs associated with SMDT; disc_QALY, the discount rate used for QALYs/costs; EOC_OS10y, the overall survival estimate for women with early ovarian cancer at 10 years; HR_OS_MDT_noMDT, the overall survival HR for women referred to SMDT compared with those women not referred to SMDT; U_AOC_MDT/U_EOC_MDT, the utility associated with advanced/early ovarian cancer when the woman had been referred to SMDT.

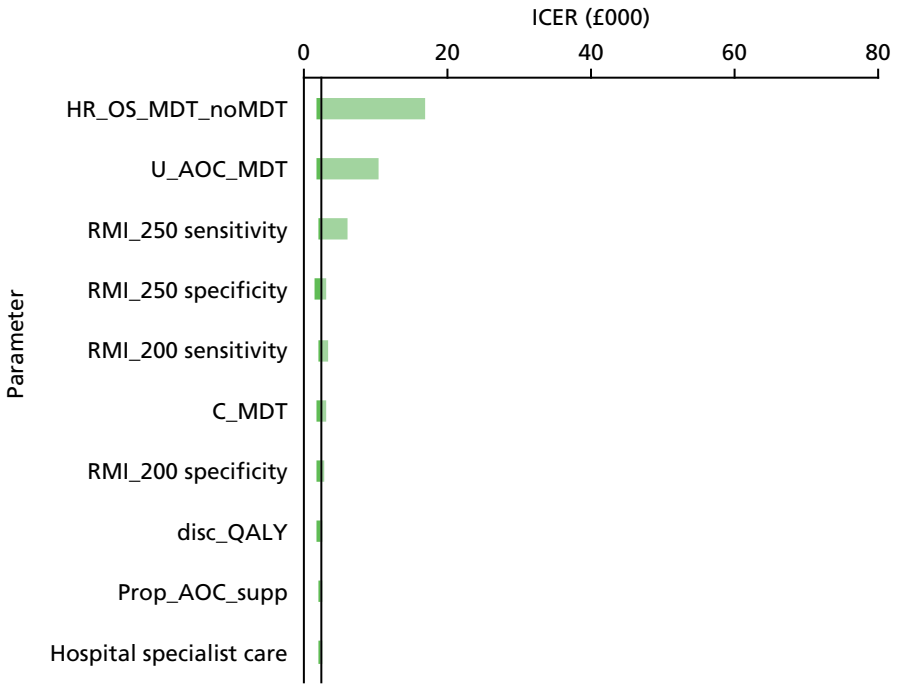

FIGURE 21 Top 10 influential parameters in the comparison of the RMI 1 (threshold of 250) with the RMI 1 (threshold of 200). C_MDT, costs associated with SMDT; disc_QALY, the discount rate used for QALYs/costs; HR_OS_MDT_noMDT, the overall survival HR for women referred to SMDT compared with those women not referred to SMDT; Prop_AOC_supp, proportion of women with AOC receiving only supportive care (and no surgery); U_AOC_MDT/U_EOC_MDT, the utility associated with advanced/early ovarian cancer when the woman had been referred to SMDT.

Appendix 9 Scenario analyses (deterministic)

Assuming a prevalence of 20% for all malignancies

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5540	14.026	17.156	–2	0.020	0.027	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5542	14.006	17.129	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5546	14.007	17.131	4	0.001	0.002	£2511	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5572	14.028	17.158	30	0.022	0.029	£1333	£16,137
ROMA score using Abbott Diagnostics' ARCHITECT	5580	14.010	17.136	37	0.004	0.007	£9008	Dominated
ROMA score using Roche Diagnostics' Elecsys	5586	14.012	17.138	43	0.006	0.009	£7169	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5648	14.022	17.151	106	0.017	0.022	£6385	Dominated

Assuming a prevalence of 30% for all malignancies

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	6432	12.709	15.626	–1	0.032	0.041	Dominant	Cheapest
RMI 1 (at a threshold of 250)	6433	12.677	15.585	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	6438	12.679	15.589	4	0.002	0.003	£2059	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	6464	12.712	15.630	31	0.035	0.045	£888	£10,504
ROMA score using Abbott Diagnostics' ARCHITECT	6473	12.683	15.595	40	0.006	0.010	£6408	Dominated
ROMA score using Roche Diagnostics' Elecsys	6478	12.687	15.600	45	0.010	0.015	£4466	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	6539	12.703	15.619	106	0.026	0.034	£4105	Dominated

Assuming 0% prevalence of non-ovarian malignancies

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5339	13.920	16.995	–4	0.024	0.031	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5343	13.896	16.964	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5347	13.898	16.967	4	0.002	0.002	£2427	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5371	13.922	16.997	28	0.026	0.033	£1083	£15,094
ROMA score using Abbott Diagnostics' ARCHITECT	5381	13.901	16.971	38	0.004	0.007	£8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5385	13.905	16.976	42	0.009	0.012	£4876	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5447	13.916	16.990	104	0.020	0.026	£5259	Dominated

Assuming an equal proportion of early-stage ovarian cancer versus advanced-stage ovarian cancer in the false-negative and true-positive groups (in the base case it was assumed that false negatives would predominantly/all be early stage)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
RMI 1 (at a threshold of 250)	5652	13.838	16.933	0	0.000	0.000		Cheapest
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5655	13.855	16.957	2	0.017	0.024	£147	£147
RMI 1 (at a threshold of 200)	5656	13.840	16.936	3	0.002	0.003	£1552	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5688	13.856	16.959	35	0.018	0.026	£1958	£27,656
ROMA score using Abbott Diagnostics' ARCHITECT	5689	13.844	16.942	36	0.006	0.009	£6057	Dominated
ROMA score using Roche Diagnostics' Elecsys	5694	13.847	16.945	42	0.008	0.012	£5052	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5761	13.853	16.954	109	0.015	0.021	£7451	Dominated

Assuming for International Ovarian Tumour Analysis group's simple ultrasound rules that subjective assessment would be used for inconclusive assessments (instead of assumed to be malignant)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5642	13.847	16.948	–17	0.016	0.022	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5659	13.831	16.926	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5663	13.832	16.928	4	0.002	0.002	£2427	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5689	13.855	16.957	30	0.024	0.031	£1249	£5922
ROMA score using Abbott Diagnostics' ARCHITECT	5697	13.835	16.933	38	0.004	0.007	£8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5703	13.837	16.936	44	0.007	0.010	£6625	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5765	13.849	16.950	106	0.018	0.024	£5949	Dominated

Assuming equal test costs for all risk scores

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
RMI 1 (a threshold 250)	5676	13.831	16.926	0	0.000	0.000		Cheapest
RMI 1 (a threshold 200)	5680	13.832	16.928	4	0.002	0.002	£2427	Extendedly dominated
ROMA score using Abbott Diagnostics' ARCHITECT	5686	13.835	16.933	10	0.004	0.007	£2227	Extendedly dominated
ROMA score using Roche Diagnostics' Elecsys	5696	13.837	16.936	20	0.007	0.010	£3025	Extendedly dominated
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5699	13.853	16.955	23	0.022	0.029	£1073	£1073
IOTA group's ADNEX model (a threshold 10%)	5706	13.855	16.957	30	0.024	0.031	£1249	£3057
Overa (MIA2G) from Vermillion (a threshold 5 units)	5708	13.849	16.950	33	0.018	0.024	£1832	Dominated

Assuming no ultrasound is performed in conjunction with Risk of Ovarian Malignancy Algorithm and Overa (MIA2G) risk scores, thus reducing the costs of these risk scores

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
ROMA score using Abbott Diagnostics' ARCHITECT	5621	13.835	16.933	–39	0.004	0.007	–£8759	Cheapest
ROMA score using Roche Diagnostics' Elecsys	5626	13.837	16.936	–33	0.007	0.010	–£5006	Extendedly dominated
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5657	13.853	16.955	–2	0.022	0.029	–£96	£2109
RMI 1 (at a threshold of 250)	5659	13.831	16.926	0	0.000	0.000	£0	Dominated
RMI 1 (at a threshold of 200)	5663	13.832	16.928	4	0.002	0.002	£2427	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5689	13.849	16.950	29	0.018	0.024	£1645	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5689	13.855	16.957	30	0.024	0.031	£1249	£15,094

Assuming additional costs for false positives (surgery costs with malignancy instead of without) and additional costs for false negatives (additional costs of benign surgery)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5759	13.853	16.955	–174	0.022	0.029	–£7986	Cheapest
IOTA group's ADNEX model (at a threshold of 10%)	5793	13.855	16.957	–140	0.024	0.031	–£5829	£16,372
ROMA score using Roche Diagnostics' Elecsys	5904	13.837	16.936	–29	0.007	0.010	–£4384	Dominated
ROMA score using Abbott Diagnostics' ARCHITECT	5905	13.835	16.933	–28	0.004	0.007	–£6261	Dominated
RMI 1 (at a threshold of 200)	5915	13.832	16.928	–18	0.002	0.002	–£11,809	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5921	13.849	16.950	–12	0.018	0.024	–£675	Dominated
RMI 1 (at a threshold of 250)	5933	13.831	16.926	0	0.000	0.000	£0	Dominated

Assuming additional costs for false positives (surgery costs with malignancy instead of without) and additional costs for false negatives (additional costs of benign surgery and specialist multidisciplinary team costs)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5760	13.853	16.955	–182	0.022	0.029	–£8322	Cheapest
IOTA group's ADNEX model (at a threshold of 10%)	5794	13.855	16.957	–147	0.024	0.031	–£6158	£16,128
ROMA score using Roche Diagnostics' Elecsys	5909	13.837	16.936	–32	0.007	0.010	–£4935	Dominated
ROMA score using Abbott Diagnostics' ARCHITECT	5911	13.835	16.933	–30	0.004	0.007	–£6851	Dominated
RMI 1 (at a threshold of 200)	5923	13.832	16.928	–19	0.002	0.002	–£12,400	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5923	13.849	16.950	–18	0.018	0.024	–£1033	Dominated
RMI 1 (at a threshold of 250)	5942	13.831	16.926	0	0.000	0.000	0	Dominated

Assuming a discount of 92% for carboplatin (CG122: discount in England of 91.8%; discount in Wales of 92.1%)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5363	13.853	16.955	–1	0.022	0.029	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5364	13.831	16.926	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5368	13.832	16.928	4	0.002	0.002	£2427	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5395	13.855	16.957	31	0.024	0.031	£1280	£15,136
ROMA score using Abbott Diagnostics' ARCHITECT	5402	13.835	16.933	38	0.004	0.007	£8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5408	13.837	16.936	44	0.007	0.010	£6629	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5470	13.849	16.950	106	0.018	0.024	£5977	Dominated

Assuming a discount of 95% for paclitaxel (CG122: discount in England of 91.0%; discount in Wales of 95.4%)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5102	13.853	16.955	–1	0.022	0.029	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5103	13.831	16.926	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5107	13.832	16.928	4	0.002	0.002	2427	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5134	13.855	16.957	32	0.024	0.031	1318	£15,186
ROMA score using Abbott Diagnostics' ARCHITECT	5141	13.835	16.933	38	0.004	0.007	8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5146	13.837	16.936	44	0.007	0.010	6635	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5210	13.849	16.950	107	0.018	0.024	6010	Dominated

Assuming an alternative hazard ratio of 0.808 for progression-free and overall survival for specialist multidisciplinary team referral versus no specialist multidisciplinary team referral

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5658	13.847	16.948	0	0.043	0.057	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5659	13.804	16.891	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5664	13.807	16.896	5	0.003	0.005	1664	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5690	13.851	16.953	31	0.047	0.062	660	7464
ROMA score using Abbott Diagnostics' ARCHITECT	5700	13.812	16.905	41	0.009	0.014	4812	Dominated
ROMA score using Roche Diagnostics' Elecsys	5707	13.816	16.911	48	0.013	0.020	3760	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5767	13.839	16.938	108	0.035	0.047	3078	Dominated

Assuming an alternative hazard ratio of 0.990 for progression-free and overall survival for specialist multidisciplinary team referral versus no specialist multidisciplinary team referral

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5656	13.858	16.961	–4	0.002	0.003	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5660	13.856	16.958	0	0.000	0.000		Dominated
RMI 1 (at a threshold of 200)	5663	13.856	16.959	3	0.000	0.000	£16,921	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5689	13.858	16.961	29	0.002	0.003	£12,374	£158,980
ROMA score using Abbott Diagnostics' ARCHITECT	5694	13.856	16.959	34	0.000	0.001	£78,602	Dominated
ROMA score using Roche Diagnostics' Elecsys	5699	13.856	16.959	39	0.001	0.001	£60,637	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5763	13.858	16.961	103	0.002	0.002	£60,005	Dominated

Assuming that the proportion of patients receiving supportive care (for advanced-stage ovarian cancer) is 10% (instead of 5%)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
RMI 1 (at a threshold of 250)	5780	13.831	16.926	0	0.000	0.000	0	Cheapest
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5781	13.853	16.955	1	0.022	0.029	37	£37
RMI 1 (at a threshold of 200)	5784	13.832	16.928	5	0.002	0.002	3005	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5813	13.855	16.957	33	0.024	0.031	1368	£15,066
ROMA score using Abbott Diagnostics' ARCHITECT	5820	13.835	16.933	40	0.004	0.007	9106	Dominated
ROMA score using Roche Diagnostics' Elecsys	5827	13.837	16.936	47	0.007	0.010	7134	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5889	13.849	16.950	109	0.018	0.024	6119	Dominated

Assuming an alternative transvaginal sonography cost of £142.46 (MA36Z) (instead of £76.75 based on CG122)[1]

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5723	13.853	16.955	–2	0.022	0.029	–96	Cheapest
RMI 1 (at a threshold of 250)	5725	13.831	16.926	0	0.000	0.000	0	Dominated
RMI 1 (at a threshold of 200)	5729	13.832	16.928	4	0.002	0.002	2427	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5755	13.855	16.957	30	0.024	0.031	1249	£15,094
ROMA score using Abbott Diagnostics' ARCHITECT	5763	13.835	16.933	38	0.004	0.007	8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5768	13.837	16.936	44	0.007	0.010	6625	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5831	13.849	16.950	106	0.018	0.024	5949	Dominated

Assuming an alternative transvaginal sonography cost of £142.46 (MA36Z) (instead of £76.75 based on CG122)[1] and increasing the transvaginal sonography for the International Ovarian Tumour Analysis group's risk scores by 20% (to reflect potential training costs)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
RMI 1 (at a threshold of 250)	5725	13.831	16.926	0	0.000	0.000	0	Cheapest
RMI 1 (at a threshold of 200)	5729	13.832	16.928	4	0.002	0.002	2427	Extendedly dominated
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5751	13.853	16.955	26	0.022	0.029	1206	£1206
ROMA score using Abbott Diagnostics' ARCHITECT	5763	13.835	16.933	38	0.004	0.007	8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5768	13.837	16.936	44	0.007	0.010	6625	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5783	13.855	16.957	58	0.024	0.031	2435	£15,094
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5831	13.849	16.950	106	0.018	0.024	5949	Dominated

Assuming additional costs of specialist multidisciplinary team referral of £2500 to reflect higher surgery costs

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
RMI 1 (at a threshold of 250)	6162	13.831	16.926	0	0.000	0.000	0	Cheapest
RMI 1 (at a threshold of 200)	6219	13.832	16.928	57	0.002	0.002	36,724	Extendedly dominated
ROMA score using Abbott Diagnostics' ARCHITECT	6333	13.835	16.933	171	0.004	0.007	38,526	Extendedly dominated
ROMA score using Roche Diagnostics' Elecsys	6533	13.837	16.936	371	0.007	0.010	56,351	Extendedly dominated
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	6627	13.853	16.955	464	0.022	0.029	21,275	£21,275
IOTA group's ADNEX model (at a threshold of 10%)	6807	13.855	16.957	645	0.024	0.031	26,929	£85,145
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	6915	13.849	16.950	753	0.018	0.024	42,337	Dominated

Assuming 90% of the non-malignancy surgery and complications costs for true negatives reflecting a scenario wherein 90% of the true negatives are operated on (instead of all)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
RMI 1 (at a threshold of 250)	5419	13.831	16.926	0	0.000	0.000	0	Cheapest
RMI 1 (at a threshold of 200)	5427	13.832	16.928	8	0.002	0.002	5311	Extendedly dominated
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5458	13.853	16.955	39	0.022	0.029	1791	£1791
ROMA score using Abbott Diagnostics' ARCHITECT	5467	13.835	16.933	48	0.004	0.007	10,837	Dominated
ROMA score using Roche Diagnostics' Elecsys	5496	13.837	16.936	77	0.007	0.010	11,685	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5508	13.855	16.957	89	0.024	0.031	3735	£23,755
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5593	13.849	16.950	174	0.018	0.024	9783	Dominated

Assuming Avastin for advanced-stage ovarian cancer

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	8412	13.887	17.010	–9	0.022	0.029	–406	Cheapest
RMI 1 (at a threshold of 250)	8421	13.865	16.980	0	0.000	0.000	0	Dominated
RMI 1 (at a threshold of 200)	8425	13.867	16.983	4	0.002	0.003	2284	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	8443	13.889	17.012	22	0.024	0.032	904	£14,728
ROMA score using Abbott Diagnostics' ARCHITECT	8459	13.870	16.988	38	0.005	0.008	7894	Dominated
ROMA score using Roche Diagnostics' Elecsys	8464	13.872	16.991	44	0.007	0.011	6151	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	8522	13.883	17.005	101	0.018	0.024	5526	Dominated

Assuming a disutility of 0.100 for false positives during the first year in the state-transition model

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5657	13.835	16.955	–2	0.010	0.029	–207	Cheapest
RMI 1 (at a threshold of 250)	5659	13.825	16.926	0	0.000	0.000	0	Dominated
RMI 1 (at a threshold of 200)	5663	13.825	16.928	4	0.000	0.002	13,465	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5689	13.832	16.957	30	0.007	0.031	4257	Dominated
ROMA score using Abbott Diagnostics' ARCHITECT	5697	13.826	16.933	38	0.002	0.007	24,820	Dominated
ROMA score using Roche Diagnostics' Elecsys	5703	13.822	16.936	44	–0.003	0.010	–15,114	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5765	13.823	16.950	106	–0.002	0.024	–66,380	Dominated

Assuming a disutility of 0.010 for false positives during the first year in the state-transition model

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER (£)	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5657	13.851	16.955	–2	0.021	0.029	–102	Cheapest
RMI 1 (at a threshold of 250)	5659	13.830	16.926	0	0.000	0.000	0	Dominated
RMI 1 (at a threshold of 200)	5663	13.832	16.928	4	0.001	0.002	2644	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5689	13.852	16.957	30	0.022	0.031	1344	£20,023
ROMA score using Abbott Diagnostics' ARCHITECT	5697	13.834	16.933	38	0.004	0.007	9126	Dominated
ROMA score using Roche Diagnostics' Elecsys	5703	13.836	16.936	44	0.006	0.010	7738	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5765	13.846	16.950	106	0.016	0.024	6676	Dominated

Comparison of different Risk of Malignancy Index 1 thresholds

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
RMI 1 (at a threshold of 300)	5647	13.826	16.919	–13	–0.004	–0.007	2865	Cheapest
RMI 1 (at a threshold of 250)	5659	13.831	16.926	0	0.000	0.000		Extendedly dominated
RMI 1 (at a threshold of 200)	5663	13.832	16.928	4	0.002	0.002	2427	Extendedly dominated
RMI 1 (at a threshold of 150)	5664	13.834	16.932	4	0.004	0.006	1172	Extendedly dominated
RMI 1 (at a threshold of 100)	5671	13.838	16.937	11	0.007	0.011	1619	Extendedly dominated
RMI 1 (at a threshold of 50)	5690	13.848	16.949	30	0.017	0.023	1783	£2006
RMI 1 (at a threshold of 25)	5706	13.853	16.956	46	0.023	0.030	2051	£2890

Using the most optimal Risk of Malignancy Index 1 threshold (i.e. Risk of Malignancy Index 1 threshold cost-effective at £20,000 and/or £30,000 per quality-adjusted life year gained in former scenario)

	Discounted			Compared with standard RMI 1				
	Costs (£)	QALYs	LYs	ΔCosts (£)	ΔQALYs	ΔLYs	ICER	Full incremental
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5657	13.853	16.955	–2	0.022	0.029	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5659	13.831	16.926	0	0.000	0.000		Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5689	13.855	16.957	30	0.024	0.031	£1249	£15,094
ROMA score using Abbott Diagnostics' ARCHITECT	5697	13.835	16.933	38	0.004	0.007	£8527	Dominated
ROMA score using Roche Diagnostics' Elecsys	5703	13.837	16.936	44	0.007	0.010	£6625	Dominated
RMI 1 (at a threshold of 25)	5706	13.853	16.956	46	0.023	0.030	£2051	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5765	13.849	16.950	106	0.018	0.024	£5949	Dominated

Appendix 10 Additional subgroup analyses (probabilistic)

TABLE 46 Probabilistic results for the base-case analysis: costs, QALYs and incremental analysis (subgroup aged 50 years)

			Compared with RMI 1 at a threshold of 250			Full incremental
Risk scores	Costs (£) (95% CI)	QALYs (95% CI)	ΔCosts (£)	ΔQALYs	ΔCosts (£)/ ΔQALYs	ΔCosts/ ΔQALYs
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5652 (4544 to 6922)	11.640 (11.306 to 11.911)	–3	0.020	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5654 (4542 to 6924)	11.621 (11.287 to 11.891)				Dominated
RMI 1 (at a threshold of 200)	5658 (4545 to 6929)	11.622 (11.287 to 11.893)	4	0.001	2561	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5684 (4574 to 6958)	11.642 (11.308 to 11.912)	29	0.021	1371	£17,212
ROMA score using Abbott Diagnostics' ARCHITECT	5691 (4584 to 6962)	11.625 (11.291 to 11.897)	37	0.005	7719	Dominated
ROMA score using Roche Diagnostics' Elecsys	5698 (4588 to 6979)	11.627 (11.291 to 11.899)	43	0.007	6657	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5760 (4638 to 7035)	11.637 (11.302 to 11.907)	106	0.016	6602	Dominated

TABLE 47 Probabilistic results for the base-case analysis: costs, QALYs and incremental analysis (subgroup of early-stage ovarian cancer)

Risk scores	Costs (£) (95% CI)	QALYs (95% CI)	Compared with RMI 1 at a threshold of 250			Full incremental
			ΔCosts (£)	ΔQALYs	ΔCosts (£)/ ΔQALYs	ΔCosts/ ΔQALYs
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5460 (4364 to 6710)	14.711 (14.363 to 15.018)	–10	0.029	Dominant	Cheapest
RMI 1 (at a threshold of 250)	5470 (4372 to 6724)	14.681 (14.333 to 14.987)				Dominated
RMI 1 (at a threshold of 200)	5471 (4373 to 6726)	14.685 (14.337 to 14.991)	2	0.004	480	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5492 (4387 to 6750)	14.713 (14.365 to 15.019)	22	0.031	715	£15,631
ROMA score using Abbott Diagnostics' ARCHITECT	5501 (4403 to 6754)	14.692 (14.343 to 14.999)	32	0.010	3052	Dominated
ROMA score using Roche Diagnostics' Elecsys	5506 (4404 to 6757)	14.696 (14.348 to 15.003)	36	0.014	2501	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5568 (4469 to 6825)	14.707 (14.359 to 15.013)	98	0.025	3897	Dominated

TABLE 48 Probabilistic results for the base-case analysis: costs, QALYs and incremental analysis (subgroup of AOC)

Risk scores	Costs (£) (95% CI)	QALYs (95% CI)	Compared with RMI 1 at a threshold of 250			Full incremental
			ΔCosts (£)	ΔQALYs	ΔCosts (£)/ ΔQALYs	ΔCosts/ ΔQALYs
RMI 1 (at a threshold of 250)	5719 (4582 to 6995)	13.556 (13.138 to 13.906)	0	0.000	0	Cheapest
RMI 1 (at a threshold of 200)	5712 (4579 to 6988)	13.544 (13.133 to 13.890)	4	0.002		Extendedly dominated
IOTA group's simple ultrasound rules (inconclusive, assumed to be malignant)	5716 (4583 to 6992)	13.545 (13.134 to 13.893)	7	0.012	571	£571
ROMA score using Abbott Diagnostics' ARCHITECT	5752 (4621 to 7024)	13.556 (13.138 to 13.906)	38	0.004	8837	Dominated
IOTA group's ADNEX model (at a threshold of 10%)	5750 (4623 to 7024)	13.548 (13.134 to 13.897)	40	0.013	3104	£39,171
ROMA score using Roche Diagnostics' Elecsys	5757 (4623 to 7028)	13.549 (13.135 to 13.898)	45	0.006	7495	Dominated
Overa (MIA2G) from Vermillion (at a threshold of 5 units)	5824 (4693 to 7105)	13.554 (13.137 to 13.904)	112	0.010	10,748	Dominated

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library