Tests in secondary care to identify people at high risk of ovarian cancer: A systematic review and cost effectiveness analysis

A Diagnostic Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Kleijnen Systematic Reviews Ltd in collaboration with Erasmus University Rotterdam and Maastricht University

Authors

Marie Westwood, Review Manager, Kleijnen Systematic Reviews Ltd, UK
Bram Ramaekers, Health Economist, Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical

Centre & CAPHRI, School for Public Health and Primary Care, Department of Health Services Research, Maastricht University, the Netherlands
Shona Lang, Reviewer, Kleijnen Systematic Reviews Ltd, UK
Sabine Grimm, Health Economist, Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre & CAPHRI, School for Public Health and Primary Care, Department of Health Services Research, Maastricht University, the Netherlands
Sohan Deshpande, Reviewer, Kleijnen Systematic Reviews Ltd, UK
Shelley de Kock, Information Specialist, Kleijnen Systematic Reviews Ltd, UK
Nigel Armstrong, Senior Health Economist, Kleijnen Systematic Reviews Ltd, UK
Manuela Joore, Professor of Health Technology Assessment and Decision Making, Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre & CAPHRI, School for Public Health and Primary Care, Department of Health Services Research, Maastricht University, the Netherlands
Jos Kleijnen, Professor of Systematic Reviews in Health Care, School for Public Health and Primary Care (CAPHRI), Maastricht University, the Netherlands

Correspondence to Marie Westwood
Kleijnen Systematic Reviews Ltd
Unit 6, Escrick Business Park
Riccall Road
Escrick
York YO19 6FD
Tel: 01904 727983
Email: marie@systematic-reviews.com

Date completed 18/05/2017

PROSPERO registration: CRD42016053326

Source of funding: This report was commissioned by the NIHR HTA Programme as project 16/30/02

Declared competing interests of the authors: None declared

All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare (1) no financial support for the submitted work from anyone other than their employer; (2) no financial relationships with commercial entities that might have an interest in the submitted work; (3) no spouses, partners, or children with relationships with commercial entities that might have an interest in the submitted work; and (4) no no-financial interests that may be relevant to the submitted work.

This report should be referenced as follows:

Academic in confidence (AiC) information is marked [redacted].
Commercial in confidence (CiC) information is marked [redacted].

ABSTRACT

Background

Current guidance, (NICE clinical guideline CG122) recommends the calculation of a risk malignancy index 1 (RMI 1) score, in secondary care, and measurement of serum CA125 for people with suspected ovarian cancer for whom this has not already been done in primary care. Patients with an RMI 1 score ≥250 are currently referred to a specialist gynaecological oncology multi-disciplinary team (SMDT).

Objectives

To assess the clinical and cost effectiveness of using alternative risk scores that include CA125, HE4 or ultrasound (ROMA, IOTA simple ultrasound rules, the IOTA ADNEX model, Overa (MIA2G), and RMI 1 at thresholds other than 250) to guide referral decisions for people with suspected ovarian cancer in secondary care.

Methods

Twenty-one databases were searched to November 2016. Search results were screened for relevance independently by two reviewers. Review methods followed published guidelines. 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, borderline cancer) for which data were available.

The cost effectiveness analysis considered long-term costs and quality adjusted life years (QALYs) associated with different risk scoring methods to decide the most appropriate care pathway for women with suspected ovarian cancer. Modelling comprised a decision tree and a Markov model. The decision tree was used to model the short-term outcomes. The Markov model was used to estimate long-term costs and QALYs associated with cancer treatment and progression.

Results

Fifty-one diagnostic cohort studies were included in the systematic review. No studies were identified that studied the effects on clinical management decisions or clinical outcomes.

The ROMA score, using either Abbott ARCHITECT or Roche Elecsys tumour marker assays, did not offer any significant performance advantage over the RMI 1, at a threshold of 200, where all participants were included in the analysis; no comparative data were identified for a threshold of 250. There were no studies evaluating the ROMA score using the Fujirebio Lumipulse G automated

CEIA system. The performance of the ROMA score did not differ significantly between pre- and post-menopausal 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.

[redacted] In order to achieve similar levels of sensitivity to those provided by the ADNEX model and IOTA 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 and 55.2%), respectively.

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

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

In the base-case analysis, the RMI 1 with a threshold of 250 was least effective (16.926 life years, 13.820 QALYs) and second cheapest (£5,669). The IOTA simple ultrasound rules (inconclusive assumed to be malignant), was cheapest (£5,667) and second most effective (16.954 life years, 13.841 QALYs) and thereby dominating the RMI 1 (at both the 200 and 250 thresholds). The IOTA ADNEX model (threshold 10%), cost £5,699, was most effective (16.957 life years, 13.843 QALYs) and compared with the IOTA simple ultrasound rules resulted in an ICER of £15,304 per QALY gained. The remaining risk scores (ROMA Abbott ARCHITECT, ROMA Roche Elecsys and Overa (MIA2G) Vermillion) were dominated. As a result, incremental analysis indicated that up to thresholds of £15,304 per QALY gained the IOTA simple ultrasound rules is cost effective whereas the IOTA ADNEX model (threshold 10%) is cost effective for higher thresholds.

Conclusions

There is evidence to suggest that using either the ADNEX model or IOTA simple ultrasound rules to assess the risk of malignancy in women with adnexal mass may offer increased sensitivity relative to

current practice (the RMI 1 at a decision threshold of 250 or 200), i.e. a higher proportion of those women who have a malignant tumour would be referred to an SMDT.

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

Funding

This report was commissioned by the NIHR HTA Programme.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for the non-specialist reader.

AiC	academic in confidence
ADNEX	Assessment of Different Neoplasms in the adnexa
AFP	Alpha-fetoprotein
BSGE	British Society of Gynaecological Endoscopy
CA125	Cancer Antigen 125
CADTH	Canadian Agency for Drugs and Technologies in Health
CCT	controlled clinical trial
CDSR	Cochrane Database of Systematic Reviews
CEAC	cost-effectiveness acceptability curve
CEF	cost-effectiveness frontier
CENTRAL	Cochrane Central Register of Controlled Trials
CI	confidence interval
CLEIA	chemiluminescent enzyme immunoassay
CMIA	chemiluminescent microparticle immunoassay
CRC	colorectal cancer
CT	computed tomography
CV	co-efficient of variation
DTA	diagnostic test accuracy
ECL	electrochemiluminescence
EED	Economic Evaluations Database
EFSUMB	European Federation of Societies for Ultrasound in Medicine and Biology
EIA	enzyme immunoassay
FIGO	International Federation of Gynaecology and Obstetrics
FN	false negative
FP	false positive
β-HCG	β Human Chorionic Gonadotropin
HCHS	Hospital and Community Health Services
HE4	Human Epididymis protein 4
HES	Hospital Episode Statistics
HRQoL	health-related quality of life
HSROC	hierarchical summary receiver operating characteristic
HTA	Health technology Assessment
ICER	incremental cost-effectiveness ratio
INAHTA	International Network of Agencies for Health Technology Assessment
IOTA	International Ovarian Tumour Analysis
LDH	lactate dehydrogenase
LY	life year
MDT	multi-disciplinary team
MRI	Magnetic Resonance Imaging
NA	not applicable
NCIN	National Cancer Intelligence Network

NICE	National Institute for Health and Care Excellence
NIH	National Institutes of Health
NIHR	National Institute for Health Research
NPV	negative predictive value
NR	not reported
ONS	Office for National Statistics
PET-CT	Positron Emission Tomography Computed Tomography
PPV	positive predictive value
PSA	probabilistic sensitivity analysis
PSSRU	Personal Social Service Research Unit
QALY	quality-adjusted life year
RCOG	Royal College of Obstetrics and Gynaecology
RCT	randomised controlled trial
ROC	receiver operating characteristic
ROMA	Risk of Ovarian Malignancy Algorithm
SIGN	Scottish Intercollegiate Guidelines Network
SMDT	specialist multidisciplinary team
SROC	summary receiver operating characteristic
TN	true negative
TP	true positive

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 healthcare interventions.
False negative	Incorrect negative test result – number of diseased persons with a negative test result.
False positive	Incorrect positive test result – 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 whose 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	Statistical techniques used to combine the results of two or more studies and obtain a combined estimate of effect.
Meta-regression	Statistical technique used to explore the relationship between study characteristics and study results.
Negative predictive value	The probability of non-disease among persons with a negative test result
Opportunity costs	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	Bias arising from the preferential publication of studies with statistically significant results.
Regression analysis	A statistical method for estimating relationships among variables
Quality of life	An individual's emotional, social and physical well-being and their ability to perform the ordinary tasks of living.
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.
Receiver Operating Characteristic curve	A graph which 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 calculation of estimates of accuracy.
Sensitivity	Proportion of people with the target disorder who have a positive test result.
Specificity	Proportion of people without the target disorder who have a negative test result.
True negative	Correct negative test result – number of non-diseases persons with a negative test result.

True positive	Correct positive test result – number of diseased persons with a positive test result.

SCIENTIFIC SUMMARY

Background

Current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management[1] recommends that serum cancer antigen 125 (CA125) should be measured in secondary care, in all people with suspected ovarian cancer for whom serum CA125 has not already been measured in primary care. CG122 specifically recommends the calculation of a risk malignancy index 1 (RMI 1) score, which includes CA125, ultrasound features and menopausal status, with referral to an SMDT for people with an RMI score ≥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 and cost effectiveness of using alternative risk scores that include CA125, HE4 or ultrasound ((ROMA, IOTA simple ultrasound rules, the IOTA ADNEX model, Overa (MIA2G), and RMI 1 at thresholds other than 250) to guide referral decisions for people with suspected ovarian cancer in secondary care. The following research questions were defined:

- What is the accuracy of alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current practice), where the target condition is histologically confirmed ovarian cancer?
- What are the effects of using alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI 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 score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current practice), when routinely used, in secondary care, to guide decisions about referral to an SMDT, for people with suspected ovarian cancer?

Methods

Assessment of clinical effectiveness

Twenty-one databases, including MEDLINE and EMBASE, research registers and conference proceedings were searched to November 2016. Search results were screened for relevance independently by two reviewers. Full text inclusion assessment, data extraction, and quality assessment were conducted by one reviewer and checked by a second. Study quality was assessed

using QUADAS-2 and PROBAST. 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, borderline cancer) for which data were available.

Assessment of cost effectiveness

The base case analysis included seven risk scores:

- RMI (threshold 250)
- ROMA Abbott ARCHITECT
- ROMA Roche Elecsys
- Overa (MIA2G) Vermillion (threshold 5 units)
- IOTA Simple Rules (inconclusive assumed to be malignant)
- IOTA ADNEX model (threshold 10%)
- RMI (threshold 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 an SMDT and treatment in tertiary care, or no tertiary referral. It was also taken into account that a small proportion of the patients with pelvic masses are diagnosed with colorectal cancer (consistent with CG 122).

We developed a decision tree and a Markov model. 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. receive a high-risk test result (either true or false positive) are referred to an SMDT, and patients who receive a low risk test result (either true or false negative) are not referred to an 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 IOTA simple ultrasound rules, seven for the IOTA 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 included patients were included in the analysis), 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 Abbott ARCHITECT or Roche Elecsys tumour marker assays, did not offer any clear performance advantage over the RMI 1. The only ROMA score study (n=213), using Abbott ARCHITECT assays, 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, where participants with borderline tumours and/or those with malignancies other than epithelial ovarian cancer were excluded from the analyses (two studies, n=1,172), 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 decision threshold of 200, 80.3% (95% CI: 77.5 to 82.9) and the summary sensitivity estimates were similar and high, 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 Elecsys tumour marker assays, and the RMI 1 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 slightly higher than that for the RMI 1, 83.8% (95% CI: 73.4 to 91.3%) versus 78.4 (95% CI: 67.3 to 87.1%) and the specificity estimate for the ROMA score appeared slightly lower than that for the RMI 1, 68.8% (95% CI: 61.6 to 75.4%) versus 79.6% (95% CI: 73.1 to 85.1%), but neither difference was statistically significant. The summary estimates of sensitivity and specificity for the ROMA score, using Roche Elecsys tumour marker assays at the manufacturer's recommended thresholds, derived from non-comparative accuracy studies where all participants 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 (two studies, n=1,252). In studies where the manufacturer's recommended cut-offs were used, the performance of the ROMA score did not differ significantly between pre- and post-menopausal 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 the Fujirebio Lumipulse G automated CEIA system.

The summary estimates of sensitivity, derived from direct comparison studies that included all study participants in their analyses (two studies, n=[redacted]), were significantly higher for both the ADNEX model, 96% (95% CI: 94.5 to 97.1%), and IOTA simple ultrasound rules, 92.8% (95% CI: 90.9 to 94.3%), than for the RMI 1 at a decision threshold of 200, 66% (95% CI: 62.9 to 69%) [redacted] Conversely, the summary estimates of specificity, for both the ADNEX model 67% (95% CI: 64.2 to 69.6%) and IOTA simple ultrasound rules 71.6% (95% CI: 68.9 to 74.1%), were significantly lower than that for the RMI 1 at a decision threshold of 200, 89% (95%CI: 87 to 90.7%) [redacted]. In order to achieve similar levels of sensitivity to those provided by the ADNEX model and IOTA 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 and 55.2%), respectively.

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

Studies evaluating the RMI 1 at different thresholds indicated no significant difference in the 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 least effective (16.926 life years, 13.820 QALYs) and second cheapest (£5,669). The IOTA simple ultrasound rules (inconclusive assumed to be malignant), was cheapest (£5,667) and second most effective (16.954 life years, 13.841 QALYs) and thereby dominating the RMI 1 (at both the 200 and 250 thresholds). The IOTA ADNEX model (threshold 10%), cost £5,699, was most effective (16.957 life years, 13.843 QALYs) and compared with the IOTA simple ultrasound rules resulted in an ICER of £15,304 per QALY gained. The remaining risk scores (ROMA Abbott ARCHITECT, ROMA Roche Elecsys and Overa (MIA2G) Vermillion) were dominated. As a result, incremental analysis indicated that up to thresholds of £15,304 per QALY gained the IOTA simple ultrasound rules is cost-effective whereas the IOTA ADNEX model (threshold 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 simple ultrasound rules and IOTA ADNEX model (threshold 10%) this was 39% and 60% respectively (£20,000 threshold) and 23% and 75% respectively (£30,000 threshold). The probabilities for the other risk scores were <1% for these thresholds.

Sensitivity and scenario analyses indicated the hazard ratio for SMDT versus no SMDT referral (for patient with ovarian cancer) was the most influential parameter in the model and that the results are reasonably robust. Most scenario analyses indicated that at thresholds of £20,000 and £30,000 per QALY gained, the IOTA ADNEX model (threshold 10%) remained the cost effective strategy. In two scenario analyses IOTA simple ultrasound rules (inconclusive assumed to be malignant) was considered 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 £2,890 per QALY gained or higher), the RMI 1 was still dominated.

For the pre-menopausal and post-menopausal subgroups, the IOTA ADNEX model (threshold 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 IOTA simple ultrasound rules to assess the risk of malignancy in women with adnexal mass may offer increased sensitivity relative to current practice (the RMI 1 at a decision threshold of 250 or 200), i.e. a higher proportion of those women who have a malignant tumour would be referred to an 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 that achievable using either the ADNEX model or IOTA 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.

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

Suggested research priorities

Further studies or further analyses of the IOTA data set are needed to understand the role of menopausal status, and other potentially relevant factor such as family history of ovarian cancer, in the performance of both the IOTA and 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 to 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 investigation of an adnexal mass. Further studies are required to explore the distribution of histological diagnoses amongst patients with false negative low risk classifications; a more complete exploration of the types of patients who are likely to be misclassified as low risk, using the various risk scoring options available, as well as an investigation of the downstream clinical consequences for these patients, are required.

PLAIN ENGLISH SUMMARY

Ovarian cancer is the sixth most common cancer in women in the UK and it is more common in older women, with approximately half of new cases occurring in women over 65. There is a better chance that ovarian cancer can be treated successfully if it is 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 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. infections or benign growths).

Because a diagnosis of ovarian cancer, where the cancer has not yet spread outside the ovaries, is usually confirmed using tissue samples taken during surgery, it is important to find tests which can help to predict which women are more likely to have ovarian cancer so that they can be referred to a specialist centre as quickly as possible; appropriate referral to a specialist centre is needed to ensure that the correct surgery is carried out in the correct place and the need for further surgeries is minimised.

Tumour markers, such as CA125 and HE4, are proteins that are produced by some ovarian cancer cells, which can be measured using a blood test. However, these markers are not produced by all ovarian cancers and their levels can also be raised in women with other conditions (e.g. fibroids, endometriosis, and infections); this is particularly a problem in younger, pre-menopausal women. The ability of tumour markers to determine which women are more likely to have ovarian cancer may be improved by combining them with other tests, such as ultrasound examination.

This assessment considered how best to combine information from tumour marker blood tests, ultrasound examinations and clinical examination (signs and symptoms reported by the patient and menopausal status), in secondary care settings, in order to decide when a woman is more likely to have ovarian cancer and should therefore be referred to a specialist gynaecological oncology multi-disciplinary team (SMDT) for further investigations (including biopsy or surgery) and treatment.

We included 51 studies, which looked at how well a variety of tools could predict ovarian cancer. Two of the tools assessed, one based on features seen by ultrasound (IOTA simple ultrasound rules) and one which combined ultrasound features, a tumour marker and clinical information (the ADNEX model) identified a higher proportion of those women with cancer than the method that is currently recommended (the Risk of Malignancy Index (RMI)). This means that, if RMI were replaced by either of these tools, more women with ovarian cancer would be referred to an SMDT, however, more women with benign (non-cancerous) lumps would also be referred. The available information, from

research studies that we identified, did not indicate that any other methods of assessing risk offered a clear advantage over current practice.

Health economic analyses indicated that the IOTA ADNEX model (threshold 10%), may be cost effective compared to alternative risk scores (including the RMI as used in current practice), to guide decisions about referral to an SMDT.

1. OBJECTIVE

The overall objective of this assessment was to summarise the evidence on the clinical and cost effectiveness of using alternative risk scores that include CA125, HE4 or ultrasound (detailed in section 2.3 below) to guide referral decisions for people with suspected ovarian cancer in secondary care. Current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management[1] recommends that serum cancer antigen 125 (CA125) should be measured in secondary care, in all people with suspected ovarian cancer for whom serum CA125 has not already been measured in primary care. CA125 levels are a component of secondary care investigation and are not used in isolation; NICE guideline CG122 specifically recommends the calculation of a risk malignancy index I (RMI I) score, which includes CA125, and referral to a specialist gynaecological oncology multi-disciplinary team (SMDT) for people with an RMI score ≥250. 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 is the accuracy of alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current practice), where the target condition is histologically confirmed ovarian cancer?

- What are the effects of using alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI 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 score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current practice), when routinely used, in secondary care, to guide decisions about referral to an SMDT, for people with suspected ovarian cancer?

2. BACKGROUND AND DEFINITION OF THE DECISION PROBLEM(S)

2.1 Population

The primary indication for this assessment is optimisation of the routine secondary care assessment of people with suspected ovarian cancer to decide whether a patient should be referred to an SMDT. The assessment was conducted in the context of an update to current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management.[1] The relevant population was people of any age, including pre- and post-menopausal women, who have been referred to secondary care for the investigation of suspected ovarian cancer. This assessment includes data for people of any age, but no cost effectiveness modelling was undertaken for the population under 18 years, due to a lack of data on the performance of risk scores in this age group. People with a previous history of ovarian cancer, who are being monitored for possible recurrence, and those referred directly from primary care to an SMDT, were outside the scope of this assessment.

2.2 Target condition

The target condition, for this assessment, is 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 cancer; non-epithelial ovarian cancers are more common in pre-menopausal women.[2] The target conditions covered by NICE clinical guideline (CG122)were. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma, and borderline ovarian cancer;[1] excluded target conditions are pseudomyxoma peritonei, relapsed ovarian, fallopian tube or peritoneal cancer, germ cell tumour of the ovary, sex cord-stromal tumours of the ovary. This assessment did not limit to any particular type of ovarian cancer.

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

Stage I	confined to the organ of origin (ovaries or fallopian tubes)
Stage II	invasion of surrounding organs or tissues (pelvic extension or primary peritoneal cancer [below pelvic brim])
Stage III	spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes
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 females in the UK (2013), accounting for 4% of all new cancer cases in females.[4, 5] In 2013, there were 7,284 new cases of ovarian cancer 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 4,100 ovarian cancer deaths.[6] Incidence of ovarian cancer is strongly related to age, with 2011-2013 data indicating that approximately half (53%) of new cases were diagnosed in women over 65 years of age.[4, 5] Ovarian cancer mortality is also strongly related to age at diagnosis.[6]

Data from the Office of National Statistics, published by Cancer Research UK, indicate that, although ovarian cancer incidence rates have increased, overall, since the 1970's, UK age-standardised incidence rates decreased by 6% in the decade between 2002-2004 and 2011-2013.[7] 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 showed that the one-year and five-year survival rates for women diagnosed at stage I were 97% and 90% versus 53% and 4% for women diagnosed at stage IV.[6] 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] 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 pre-menopausal women because of the high rate of false positives resulting from various non-malignant conditions (fibroids, endometriosis, adenomyosis, pelvic infection).[10] It is therefore important to consider the effects of menopausal status of 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 RMI).

2.3 Intervention technologies

Serum tumour markers are used in the secondary care investigation of people with suspected ovarian cancer; they are not considered to be 'stand-alone' diagnostic tests, but are used in

conjunction with other test, signs and symptoms to assess 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 low volume disease outside the pelvis (stage I to IIIb).

An optimised risk assessment that reduces the number of people with ovarian cancer who are not referred for further specialist care (i.e. those with a 'false-negative' risk assessment) has the potential to improve prognosis, be cost saving in terms of unnecessary further investigations, and to reduce associated anxiety. Prognosis may be adversely affected by failure to refer patients to SMDT and specialist surgery. In particular, it is likely that patients 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 median overall survival between a set of regional centres in the UK and the UK as a whole.[11]

The current standard assessment (RMI) has been reported as having poor sensitivity, approximately 63% at an operating threshold of 200.[12] If referral decisions are based on RMI score at this threshold, there remains the potential for significant numbers of people with ovarian cancer to remain un-referred 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, and 74% for the ROMA score).[9] Risk scores with higher sensitivity are needed to facilitate prompt referral of the appropriate patient group.

2.3.1 ROMA 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 which differs depending on whether the person is pre- or postmenopausal (equations (1) and (2) in Table 1). This PI value can then be used to calculate the ROMA score (equation (3) in Table 1).[13] The ROMA score is intended for use in women who present with 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.

Table 1: ROMA equations

Premenopausal people: PI = –12.0 + 2.38 × ln[HE4] + 0.0626 × ln[CA125] (1)
Postmenopausal people: PI = –8.09 + 1.04 × ln[HE4] + 0.732 × ln[CA125] (2)
ROMA (%) = exp(PI) / [1 + exp(PI)] × 100% (3)
PI: predictive index; [HE4]: serum concentration of HE4 in pmol/L; [CA125]: serum concertation of CA125 in U/ml; ln: natural logarithm

Cut-off values for ROMA are used to classify individuals as having low or high risk of having 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 2).

Table 2: Technical characteristics of serum HE4 assays available to the UK NHS

Name of assay	Company	Detection limit	Detection range	Assay time
ARCHITECT HE4	Abbott Diagnostics	15 pmol/L	20 – 1500 pmol/L	28 minutes*
Lumipulse G HE4	Fujirebio Diagnostics	3.5 pmol/L	20-1500 pmol/L	35 minutes **
Elecsys® HE4 - Human epididymal protein 4	Roche Diagnostics	15 pmol/L	15-1,500 pmol/L	18 min***

* Time is for analyser to complete sample analysis once initiated
** Using the LUMIPULSE G1200 instrument; time is for 1 sample, time for all 42 results is 55 minutes
*** Report time is dependent on whether other tests are carried out on the same sample, but typically take less than 30 minutes

ARCHITECT HE4 (Abbott Diagnostics)

The ARCHITECT HE4 assay is a chemiluminescent micro particle immunoassay (CMIA) for the quantitative determination of HE4 in human serum. The assay is designed for use on an immunoassay analyser, specifically the Abbott ARCHITECT i2000SR or ARCHITECT i1000SR analysers. Additional materials required to run the assay are: ARCHITECT HE4 Assay software file, ARCHITECT HE4 Calibrators, ARCHITECT HE4 Controls, ARCHITECT Multi-Assay 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 CA 125 II assay as an aid in estimating the risk of epithelial ovarian cancer in people presenting with a pelvic mass who will undergo surgical intervention. The company recommends that the HE4 assay results are used in the ROMA, using the following cut-off values for ROMA scores, based on obtaining a specificity of 75%:

- in pre-menopausal patients, a ROMA value ≥7.4% indicates high risk of finding epithelial ovarian cancer
- in pre-menopausal patients, a ROMA value <7.4% indicates low risk of finding epithelial ovarian cancer
- in post-menopausal patients, a ROMA value ≥25.3% indicates high risk of finding epithelial ovarian cancer
- in post-menopausal patients, a ROMA value <25.3% indicates low risk of finding epithelial ovarian cancer

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

Lumipulse G HE4 (Fujirebio Diagnostics)

The Lumipulse G HE4 is a Chemiluminescent Enzyme Immunoassay (CLEIA) for the quantitative measurement of HE4 in human serum. The assay is designed for use on the LUMIPULSE G System (either the LUMIPULSE G1200 or 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 Lumipulse system, soda lime for Lumipulse system and Lumipulse G dilution cartridges.

The assay is intended for use in conjunction with CA125 levels (measured using the Lumipulse G CA125II 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 recommend the HE4 results are used in the ROMA and suggest the following cut-off values for ROMA scores, based on obtaining a specificity of 75%:

- in pre-menopausal patients, a ROMA value ≥13.1% indicates a high risk of finding epithelial ovarian cancer
- in pre-menopausal patients, a ROMA value <13.1% indicates a low risk of finding epithelial ovarian cancer
- in post-menopausal patients, a ROMA value ≥27.7% indicates a high risk of finding epithelial ovarian cancer
- in post-menopausal patients, a ROMA value <27.7% indicates a low risk of finding epithelial ovarian cancer

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

A further HE4 assay is also available from Fujirebio: the HE4 EIA assay, 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 HE4 immunoassay (Roche Diagnostics)

The Elecsys HE4 is an immunoassay test that uses Roche's ElectroChemiLuminescence (ECL) detection technology to quantity HE4 levels. The assay uses anti-HE4 mouse monoclonal antibodies to capture HE4 in a serum sample and label it with a ruthenium complex. 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 HE4 CalSet, PreciControl HE4 and Diluent MultiAssay. 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 CA 125 II assay as an aid in estimating the risk of epithelial ovarian cancer in premenopausal and postmenopausal people with a pelvic mass. The company recommend that the HE4 and CA125 assay results are used in the ROMA.

The company suggest the following cut-off values for ROMA scores, based on obtaining a specificity of 75%:

- in pre-menopausal patients, a ROMA value ≥11.4% indicates high risk of finding epithelial ovarian cancer
- in pre-menopausal patients, a ROMA value <11.4% indicates low risk of finding epithelial ovarian cancer
- in post-menopausal patients, a ROMA value ≥29.9% indicates high risk of finding epithelial ovarian cancer
- in post-menopausal patients, a ROMA value <29.9% indicates low risk of finding epithelial ovarian cancer

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

2.3.2 Simple rules ultrasound classification system (International Ovarian Tumour Analysis [IOTA] 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 ultrasound, which may be supplemented with abdominal ultrasound for larger masses. There are five 'rules' describing features of malignant tumours (M-rules) and five rules that describe benign tumours (B-rules) (Table 3).[14, 15] Because 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. repeat examination where ultrasound has been conducted in primary care).

Table 3: 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 largest diameter 100mm or more • Very strong blood flow (colour score 4).	• Unilocular • Solid components present, with largest solid component having a largest diameter of less than 7mm • Acoustic shadows present • Smooth multilocular tumour with largest diameter less than 100mm • No blood flow (colour score 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 and B rule (or neither) apply the mass is unclassifiable and the IOTA group state that there are then a number 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 ultrasound features described in the Simple Rules model

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

The IOTA group organise one day courses that teach the techniques for classifying masses required by the system, with participants assessed with a multiple choice test. An on-line training tool which will be freely accessible to NHS practitioners is also being developed. In addition to this training, the group also recommend 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 classification using Simple Rules, and the model has not been validated in this group.

2.3.3 The Assessment of Different NEoplasias in the adnexa (ADNEX) model (IOTA group)

The ADNEX model was developed by the IOTA group to aid pre-operative discrimination between benign, borderline, stage I invasive, stage II to IV invasive and secondary metastatic ovarian tumours, in women with an ovarian (including para-ovarian and tubal) mass.[17] The model uses nine predictors, three clinical variables (age, serum CA125, 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 ADNEX risk score http://www.iotagroup.org/adnexmodel/. Guidance has also been published on the application of ADNEX in clinical practice and the selection of risk cut-offs for risk stratification and choice of clinical

management.[18] The IOTA group note that, as with other diagnostic prediction models (other IOTA models, ROMA, RMI), ADNEX cannot be applied to women with conservatively treated adnexal tumours.

2.3.4 Overa (MIA2G)

The Overa (MIA2G) assay is a CE marked qualitative serum test that combines the results of five immunoassays into a single numeric result (the Overa (MIA2G) Risk Score). The five biomarkers included in the test are:

- Follicle-stimulating Hormone (FSH)
- Human Epididymis Protein 4 (HE4)
- Apolipoprotein A-1 (Apo A-1)
- Transferrin (TRF) and
- Cancer Antigen 125 (CA 125)

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

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

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

2.3.5 The Risk of Malignancy Index (RMI)

The RMI, used at thresholds other than that currently recommended in NICE clinical guidelines (see section 2.4 below) was considered as an alternative intervention technology.

2.4 Comparator

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

RMI I score = U x M x CA125

U: ultrasound score – one 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: menopause score – M=1 (premenopausal) or M=3 (postmenopausal). The classification of 'post-menopausal' is a woman who has had no period for more than one year or a woman over 50 who has had a hysterectomy.

CA125: serum CA125 concentration – measured in IU/ml

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

The NICE guideline CG122 recommends that people with an RMI I score of 250 or greater should be referred to a specialist gynaecological oncology multidisciplinary team.[1] However, this guideline also includes a research recommendation stating that further research should be undertaken to determine the optimum RMI I 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 I threshold to use in secondary care, and that the value used will have implications for the management options considered, and the number of people who will be referred for specialist treatment.

The SIGN guideline on the management of epithelial ovarian cancer (SIGN 135) recommends referring people with an RMI I score greater than 200 to a gynaecological oncology multidisciplinary team.[19] In addition, the RCOG guideline on ovarian cysts in postmenopausal women recommends the use of 200 as a threshold to predict the likelihood of ovarian cancer, although noting that the threshold of 250 is also acceptable; in current literature a score of 200 is often used as a cut-off value.[10]

2.5 Reference standard

Histopathology is the reference standard for assessing the accuracy of tests to identify people at high risk of 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 confined disease or low volume disease outside the pelvis) it is expected that pre-surgery biopsy would not routinely occur. Where 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.

2.6 Care pathway

2.6.1 Primary care assessment and criteria for referral to secondary care

The 2011 NICE clinical guideline CG122 (Ovarian cancer: recognition and initial management)[1] includes recommendations about recognising signs and symptoms and testing in primary care:

- Refer the woman urgently if physical examination identifies ascites and/or a pelvic or abdominal mass (which is not obviously uterine fibroids)
- Carry out tests in primary care if a woman (especially if 50 or over) reports having any of the following symptoms on a persistent or frequent basis – particularly more than 12 times per month:
    - persistent abdominal distension (women often refer to this as 'bloating')
    - feeling full (early satiety) and/or loss of appetite
    - pelvic or abdominal pain
    - increased urinary urgency and/or frequency
- Consider carrying out tests in primary care if a woman reports unexplained weight loss, fatigue or changes in bowel habit
- Advise any woman who is not suspected of having ovarian cancer to return to her GP if her symptoms become more frequent and/or persistent
- Carry out appropriate tests for ovarian cancer in any woman of 50 or over who has experienced symptoms within the last 12 months that suggest irritable bowel syndrome (IBS), because IBS rarely presents for the first time in women of this age
- Measure CA125 in primary care in women with symptoms that suggest ovarian cancer
- If CA125 is 35 IU/mL or greater, arrange an urgent ultrasound scan of the abdomen or pelvis
- If the ultrasound suggests ovarian cancer, refer the woman urgently for further investigation
- For any woman who has normal CA125 (less than 35 IU/mL), or CA125 of 35 IU/mL or greater but a normal ultrasound:
    - assess her carefully for other clinical causes of her symptoms and investigate if appropriate
    - if no other clinical cause is apparent, advise her to return to her GP if her symptoms become more frequent or persistent
    - repeat CA125

More recent guidance about cancer diagnosis, NICE guidance NG12 (Suspected cancer recognition and referral), published in 2015,[20] reproduces the recommendations from CG122 with no update.

The more recent (2013) guidance, from the Scottish Intercollegiate Guidelines Network[19] provides the following recommendations:

- in women presenting in general practice with one or more symptoms of abdominal distension or bloating with or without abdominal pain, feeling full quickly, difficulty eating, or urinary symptoms, of less than 12 months duration and occurring more than 12 times per month a diagnosis of ovarian cancer should be considered
- serum CA125 level should be measured and urgent pelvic ultrasound carried out in women with persistent abdominal distention or feeling full and/or loss of appetite or pelvic or abdominal pain or increased urinary urgency and/or frequency (particularly if symptoms occur more than 12 times per month and especially if the woman is over 50)
- if symptoms persist or worsen despite a normal CA125 level and a negative ultrasound scan, refer to secondary care.

2.6.2 Establishing a diagnosis in secondary care

The 2011 NICE clinical guideline CG122 (Ovarian cancer: recognition and initial management)[1] includes a number of specific recommendations about testing following referral to secondary care:

- Measure serum CA125 in secondary care in all women with suspected ovarian cancer, if this has not already been done in primary care
- In women under 40 with suspected ovarian cancer, measure levels of alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (beta-hCG) as well as CA125, to identify women who may not have epithelial ovarian cancer
- Calculate the risk of malignancy index I (RMI I) score (after performing ultrasound) and refer all women with an RMI I score of 250 or greater to a specialist multidisciplinary team
- Perform an ultrasound of the abdomen and pelvis as the first imaging test in secondary care for women with suspected ovarian cancer, if this has not already been done in primary care
- If the ultrasound, serum CA125 and clinical status suggest ovarian cancer, perform a computed tomography (CT) scan of the pelvis and abdomen to establish the extent of disease. Include the thorax if clinically indicated
- Do not use MRI routinely for assessing women with suspected ovarian cancer
- If offering cytotoxic chemotherapy to women with suspected ovarian cancer, first obtain a confirmed tissue diagnosis by histology (or by cytology if histology is not appropriate) in all but exceptional cases

- Offer cytotoxic chemotherapy for suspected ovarian cancer without a tissue diagnosis (histology or cytology) only:
    - in exceptional cases, after discussion at the multidisciplinary team **and**
    - after discussing with the woman the possible benefits and risks of starting chemotherapy without a tissue diagnosis
- If surgery has not been performed, use histology rather than cytology to obtain a tissue diagnosis. To obtain tissue for histology:
    - use percutaneous image-guided biopsy if this is feasible
    - consider laparoscopic biopsy if percutaneous image-guided biopsy is not feasible or has not produced an adequate sample.
    - Use cytology if histology is not appropriate

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

SIGN guideline (135)[19] includes similar recommendations about RMI score and further imaging investigations:

- RMI 1 score with a threshold of 200 should be used to predict the likelihood of ovarian cancer
- Women with an RMI 1 score >200 should be referred to a gynaecology-oncology multidisciplinary team
- CT of the abdomen and pelvis should be performed in secondary care for all patients suspected of having ovarian cancer who have an RMI score >200
- MRI is not recommended for routine staging of ovarian cancer
- PET-CT is not recommended in the diagnosis and staging of ovarian cancer

The Royal College of Obstetricians and Gynaecologists (RCOG) and the British Society for Gynaecological Endoscopy (BSGE) have produced a joint guideline about the management of suspected ovarian masses in pre-menopausal 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 determination of the presence or absence of local lymphadenopathy. Specific statements about pre-operative assessment are:

- A serum CA125 assay does not need to be undertaken in all premenopausal women when an ultrasonographic diagnosis of a simple ovarian cyst has been made
- If a serum CA125 assay is raised and less than 200IU/mL, further investigation may be appropriate to exclude/treat the common differential diagnoses
- When serum CA125 levels are raised, serial monitoring of CA125 may be helpful as rapidly rising levels are more likely to be associated with malignancy than high levels that remain static
- If serum CA125 level is higher than 200 IU/mL, discussion with a gynaecological oncologist is recommended
- Lactate dehydrogenase (LDH), AFP and beta-HCG should be measured in all women under age 40 with a complex ovarian mass because of the possibility of germ cell tumours
- A pelvic ultrasound is the single most effective way of evaluating an ovarian mass with transvaginal ultrasound being preferable due to its increased sensitivity over transabdominal ultrasound
- At present, the routine use of CT and magnetic resonance imaging (MRI) does not improve the sensitivity or specificity for ovarian malignancy obtained by transvaginal ultrasound
- An estimation of the risk of malignancy is essential in the assessment of an ovarian mass
    - A systematic review of diagnostic studies concluded that RMI I is the most effective for women with suspected ovarian cancer

The Royal College of Radiologists iRefer radiological investigation guidelines tool recommends that CT of the abdomen and pelvis has a role in identifying people 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. Positron emission tomography CT (PET-CT) is indicated as a specialised investigation for difficult management situations.[21]

2.6.3 Management of early (stage I) ovarian cancer

NICE guideline CG122 includes the following recommendations about the overall management of women with suspected early (stage I) ovarian cancer[1] and NICE Technology Appraisal Guidance (TA55) provides recommendations about first-line chemotherapy regimens[22]:

- perform retroperitoneal lymph node assessment as part of optimal surgical staging
- do not include systematic retroperitoneal lymphadenectomy (block dissection of lymph nodes from the pelvic side walls to the level of the renal veins) as part of standard surgical treatment

- do not offer adjuvant chemotherapy to women who have had optimal surgical staging and have low-risk stage I disease (grade 1 or 2, stage Ia or Ib)
- offer women with high risk stage I disease (grade 3 or stage Ic) adjuvant chemotherapy consisting of six cycles of carboplatin
- discuss the possible benefits and side effects of adjuvant chemotherapy with women who have had sub-optimal surgical staging and appear to have stage I disease
- it is recommended that paclitaxel in combination with a platinum-based compound or platinum-based therapy alone (cisplatin or carboplatin) be offered as alternatives for first-line chemotherapy (usually following surgery)
- the choice of treatment for first-line chemotherapy should be made after discussion between the responsible clinician and the patient about the risks and benefits of the options available

2.6.4 Management of advanced (stage II to IV) ovarian cancer

NICE guideline CG122 includes the following recommendations about the management of women with advanced (stage II to IV) ovarian cancer[1] and NICE Technology Appraisal Guidance (TA55 and TA284) provides recommendations about first-line chemotherapy regimens[22, 23]:

- if performing surgery for women with ovarian cancer, whether before chemotherapy or after neoadjuvant chemotherapy, the objective should be complete resection of all macroscopic disease
- do not offer intraperitoneal chemotherapy to women with ovarian cancer except as part of a clinical trial
- it is recommended that paclitaxel in combination with a platinum-based compound or platinum-based therapy alone (cisplatin or carboplatin) be offered as alternatives for first-line chemotherapy (usually following surgery)
- the choice of treatment for first-line chemotherapy should be made after discussion between the responsible clinician and the patient about the risks and benefits of the options available
- bevacizumab in combination with paclitaxel and carboplatin is not recommended for first-line treatment of advanced ovarian cancer (FIGO stages IIIB, IIIC and IV epithelial ovarian, fallopian tube or primary peritoneal cancer)

Further recommendations about chemotherapy regimens for women with recurrent ovarian cancer can be found in NICE Technology Appraisal Guidance TA389, TA381 and TA285.[24-26]

2.7 Summary of the decision problem

Current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management[1] recommends that serum cancer antigen 125 (CA125) should be measured in secondary care, in all women with suspected ovarian cancer for whom serum CA125 has not already been measured in primary care. CA125 levels can inform clinical decision making in secondary care and are not used in isolation; NICE guideline CG122 specifically recommends the calculation of a risk malignancy index I (RMI I) score, which includes CA125. CG122 does not currently include any recommendations on HE4 or risk scores or testing algorithms (other than RMI). An update to the section of CG122 that deals with establishing a diagnosis in secondary care is planned in order to assess the potential role of alternative risk scores in assessing people with suspected ovarian cancer for possible referral to an 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, HE4 or ultrasound (detailed in section 2.3 above) to guide referral decisions for people with suspected ovarian cancer in secondary care. The assessment focuses on direct comparisons between the interventions described and RMI 1, using the referral threshold of ≥250, (current practice as indicated in CG122).[1] However, we have also included assessments of the accuracy of individual risk scores. Data were collected on the accuracy and comparative accuracy of different risk scores, alternative cut-offs and risk scores used in combination, in order to determine the best way to incorporate tumour markers and ultrasound findings in the diagnostic process. We have also included prediction modelling studies, which report the development and validation of multivariable prediction models intended to be used to guide individual patient care.

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 people with suspected ovarian cancer in secondary care, compared to RMI as recommended in NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management.[1] Systematic review methods followed the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care[27] and the NICE diagnostics assessment programme manual.[28]

3.1 Systematic review methods

3.1.1 Search strategy

Search strategies were based on the specified risk scores (ROMA, IOTA simple ultrasound rules, ADNEX, Overa (MIA2G) and RMI) and the target condition (ovarian cancer), as recommended in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care and the Cochrane Handbook for Diagnostic Test Accuracy Reviews. [27, 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 text mining analysis of high-frequency subject indexing terms using Endnote reference management software. Strategy development involved an iterative approach testing candidate text and indexing terms across a sample of bibliographic databases, 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 CADTH 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.[32,33,34,35]

The following databases were searched for relevant studies:

- MEDLINE (Ovid): 1946 to November Week 2, 2016
- MEDLINE In-Process Citations (Ovid): to 22 November 2016
- MEDLINE Daily Update (Ovid): to 22 November 2016

- MEDLINE Epub Ahead of Print (Ovid): to 22 November 2016
- EMBASE (Ovid): 1974 to 23 November 2016
- Cochrane Database of Systematic Reviews (CDSR) (Wiley): to Issue 11 of 12, November 2016
- Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): to Issue 10 of 12, October 2016
- Database of Abstracts of Reviews of Effects (DARE) (Wiley): to Issue 2 of 4, April 2015
- Health Technology Assessment Database (HTA) (Wiley): to Issue 4 of 4, October 2016
- International Network of Agencies for Health Technology Assessment (INAHTA) Publications (Internet) http://www.inahta.org/publications/: to 25 November 2016
- NIHR Health Technology Assessment Programme (Internet) http://www.nets.nihr.ac.uk/programmes/hta: to 25 November 2016
- Aggressive Research Intelligence Facility (ARIF) database (Internet) http://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/index.aspx: to 25 November 2016
- PROSPERO (International Prospective Register of Systematic Reviews) (Internet) http://www.crd.york.ac.uk/prospero/: to 25 November 2016

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

- NIH ClinicalTrials.gov (Internet) http://www.clinicaltrials.gov/: to 24 November 2016
- EU Clinical Trials Register (Internet) https://www.clinicaltrialsregister.eu/ctr-search/search: to 25 November 2016
- WHO International Clinical Trials Registry Platform (ICTRP) (Internet) http://www.who.int/ictrp/en/: to 24 November 2016

The following key conference proceedings, were identified in consultation with clinical experts, and were screened for the last three 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.

3.1.2 Inclusion and exclusion criteria

Inclusion criteria for each of the clinical effectiveness questions are summarised in Table 4. Studies which fulfilled these criteria were eligible for inclusion in the review.

Table 4: Inclusion criteria

Question	**What is the accuracy of alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current practice), where the target condition is histologically confirmed ovarian?**	**What are the effects of using alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current practice), on clinical management decisions and clinical outcomes?**
Participants:	People 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[*]	
Interventions (index test):	Alternative methods of risk scoring or RMI used at thresholds other than 250, as described in section 2.3 above[**]	
Comparators:	Risk malignancy Index (RMI)[$]	
Reference standard:	Histological examination of surgically resected tissue sample[$$]	Not applicable
Outcomes:	Diagnostic accuracy (the numbers of true positive, false negative, false positive and true negative test results), where 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[$]:	Diagnostic cohort studies, directly comparing one or more interventions (index tests) and the comparator[~]	Prediction modelling studies, randomised and non-randomised controlled trials

[*] Studies will be included if the setting is unclear, but the population is described as people with suspected ovarian cancer

[**] Any data on the accuracy of risk scores used in combination or in sequence with one or more additional tests (e.g. RMI and HE4, IOTA simple rules and CA125) will also be included

[$] Not applicable for prediction modelling studies

[$$] Studies which use histological examination of a biopsy sample or follow-up (ideally for a minimum of 12 months) of people 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

[~] Studies assessing the accuracy of individual risk scores will also be included

3.1.3 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 two reviewers independently assessed these for inclusion; any disagreements were resolved by consensus. Details of studies excluded at the full paper screening stage are presented in Appendix 5.

Where studies reported insufficient information (e.g. tumour marker assay details not specified, or incomplete accuracy data), 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 process described above.

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 where these were reported), details of the risk score and it's component tests (manufacturer, antibody, detection method [including analyser used], ultrasound method, definition of a positive risk score), details of reference standard (details of methods where these were reported, definition of disease positive, details of the final histopathological diagnoses of study participants where 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 (MW and SL or SD); any disagreements were resolved by consensus.

3.1.4 Quality assessment

The methodological quality of included test accuracy studies was assessed using QUADAS-2[36] and the methodological quality of prediction model studies was assessed using the PROBAST 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 (section 3.2.2) 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.

3.1.5 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 (HSROC) model[38-40] could not be applied because data sets were to small and/or homogeneous. Heterogeneity was assessed visually using summary receiver operating characteristic plots or ROC space plots. Analyses were performed in MetaDisc.[41]

Studies were grouped by risk score, manufacturer of the tumour marker assays (where 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.

3.2 Results of the assessment of clinical effectiveness assessment

The searches of bibliographic databases identified 2,456 records, after deduplication. Following 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, 42-104]

Additionally, one set of slides from a conference presentation was provided, through NICE, by the manufacturer of Overa (MIA2G),[105] and an unpublished interim report of phase 5 of the IOTA study was provided AiC (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017). 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 5 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.

One hundred and sixty-five publications were excluded after full text screening. Six articles could not be obtained,[106-112] and a further three ongoing studies, reported in four references, were identified as potentially relevant to future updates of this assessment.[113-116] Of particular note is ROCkeTS,[114, 115] a large, prospective, phase 3 study, which is funded by NIHR and which is due to report in 2019/2020. The ROCkeTS study is evaluating the clinical utility as well as accuracy of RMI, ROMA score, IOTA simple ultrasound rules, other models and novel models not included in the scope of this assessment and will consider delivery of test in the UK NHS (where an imaging service is predominantly delivered by sonographers rather than expert gynaecologists or radiologists). We identified trial registry entries for two additional diagnostic test accuracy studies: One ongoing study

is comparing the diagnostic performance or IOTA simple ultrasound rules to that of ultrasound pattern recognition in women undergoing surgery for adnexal mass (the reference standard is histopathological diagnosis) and the estimated completion date is September 2017;[113] the second trial registry entry referred to a study assessing the diagnostic performance of a twostep triage process involving RMI 1 (threshold 200) and IOTA simple ultrasound rules, which has been terminated without publication.[116]

We contacted the authors of 11 studies, which were reported as conference abstracts with insufficient detail to determine whether they met our inclusion criteria or where outcomes were unclearly reported in the full paper,[45, 53, 60, 84, 85, 91, 95, 117-120] four of whom provided additional information that allowed the study to be included in this review.[84, 85, 91, 95]

3.2.1 Overview of included studies

Details of the 51 included studies and their associated references are provided in Table 5. The following sections of this report cite studies using the primary publication and, where 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 which reported data on the diagnostic accuracy of one or more ovarian cancer risk scores (ROMA score, IOTA simple ultrasound roles, the ADNEX model, or Overa (MIA2G)), or which provided data on the accuracy of the RMI 1 at different decision thresholds (including 250, as specified in current NICE guidelines[1]). Although ten studies reported an age range which included participants under 18 years,[42, 44, 48, 51, 52, 62, 65, 66, 84, 104] no study reported separate test performance data for this age group or indicated how many participants were under 18 years. Sixteen studies reported data on the accuracy of the ROMA score,[82-84, 87, 90, 91, 95-100, 102-104] five of which reported data to support a direct comparison of the ROMA score to the RMI 1, using a decision threshold of 200.[84, 90, 99, 100, 104] There were no studies which 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 IOTA simple ultrasound rules,[44, 47-53, 55, 58, 60, 62-67] six of which reported data to support a direct comparison of IOTA simple rules to the RMI 1 using a decision threshold of 200.[44, 48, 50, 62, 63, 66] One study compared IOTA simple ultrasound rules to the RMI 1, using a decision threshold of 250, but this study was only reported as a conference abstract and 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 to the RMI 1 using a decision threshold of 200.[44] The unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) provided

data to support a direct comparison between IOTA 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),[69, 71, 105] one of which also provided comparative accuracy data for Overa (MIA2G) versus the ROMA score.[105] There were no studies comparing the accuracy of Overa(MIA2G) to the RMI 1, at any decision threshold. Finally, 10 studies provided data on the accuracy of the RMI 1 at different decision thresholds.[72-81]

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

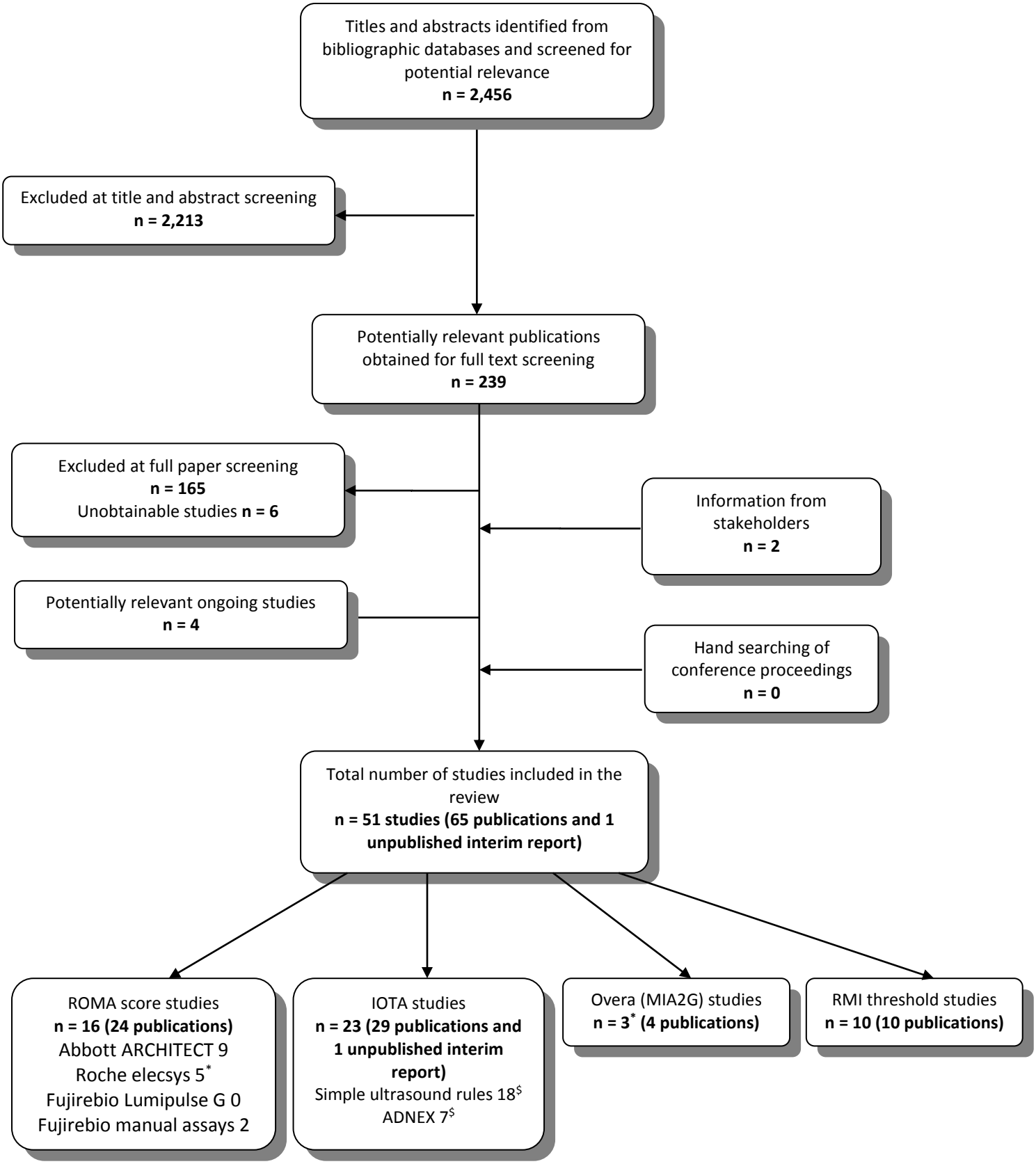
Approximately half of the included published studies (25/49) were conducted in Europe,[17, 42-46, 48-50, 52, 58, 60, 63-65, 67, 79-82, 84, 87, 98, 99, 121] of which, six were conducted solely in the UK[45, 60, 63, 67, 79, 80] and a further two were multi-national studies that included a UK centre.[17, 42] The unpublished interim analysis (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) [redacted]. There were two multi-national, worldwide studies, both of which included UK centre.[62, 66] Four studies were conducted in the USA,[69, 71, 102, 105] 13 in Asia,[47, 51, 73, 74, 76, 78, 83, 90, 91, 96, 97, 100, 103] two in Turkey,[72, 75] one on Oman,[104] and one in Brazil.[55] Two studies, which were published only as conference abstracts, did not report information about geographic location.[53, 95]

Seventeen published studies[17, 44, 46, 47, 50, 51, 62, 63, 66, 79, 82, 87, 96-99, 103] and the unpublished study for which an interim report was provided AiC (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) were publicly funded, and four studies reported receiving some funding from manufacturers (including supply of test kits, reagents and analysers).[71, 83, 84, 102] The remaining 29 included studies did not report any information about funding,[42, 43, 45, 48, 49, 52, 53, 55, 58, 60, 64, 65, 67, 69, 72-78, 80, 81, 90, 91, 95, 100, 105] or stated that they were unfunded.[104]

All studies included women with adnexal/ovarian mass, however, studies frequently reported analyses which excluded some participants based on their final histopathological diagnosis (information which 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' can 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 2 (Tables 34 and 35).

Figure 1: Flow of studies through the review process



*One study reported data for both the ROMA score using Roche elecsys tumour marker assays and Overa (MIA2G); $Two studies reported data for both ADNEX and IOTA simple ultrasound rules

Table 5: Details of included studies

Details	Country	N	Main target condition reported
ROMA score			
Abbott ARCHITECT			
Aarenstrup 2012[84]	Denmark	579	All ovarian malignancies excluding borderline
Al Musalhi 2016[104]	Oman	213	All malignant tumours including borderline
Chan 2013[83]	Multi-national (Asia)	387	All epithelial ovarian malignancies including borderline
Clemente 2015[91]	Philippines	62	Ovarian malignancies (undefined – not clear whether borderline tumours were included)
Li 2016[97]	China	917	Ovarian malignancies (undefined – not clear whether borderline tumours were included)
Moore 2011a[102] Moore 2011b[88] Moore 2012[89]	USA	450	All epithelial ovarian malignancies including borderline
Novotny 2012[87]	Czech Republic	277	All malignant tumours including borderline
Presl 2012[82]	Czech Republic	552	Ovarian malignancies (undefined – not clear whether borderline tumours were included)
Winarto 2014[100]	Indonesia	128	All epithelial ovarian malignancies including borderline
Fujirebio			
Langhe 2013[95]	NR	377	All malignant tumours including borderline
Van Gorp 2012[99] Kaijser 2013a[92] Kaijser 2013b[93] Kaijser 2013c[56] Kaijser 2014[101] Van Gorp 2011[86]	Belgium	374	All malignant tumours including borderline
Roche			
Janas 2015[98]	Poland	259	All malignant tumours including borderline
Shulman 2016[105]	USA	993	All malignant tumours including borderline
Xu 2016[96]	China	521	All epithelial ovarian malignancies excluding borderline
Yanaranop 2016[90]	Thailand	260	All malignant tumours – borderline tumours classified as disease negative

Details	Country	N	Main target condition reported
Zhang 2015b[103] Chen 2015[94]	China	612	All epithelial ovarian malignancies
Simple ultrasound rules (IOTA)			
Abdalla 2013a[48] Abdalla 2013b[57]	Poland	87	All malignant tumours including borderline
Alcázar 2013[52]	Spain	340	All malignant tumours including borderline
Baker 2013[67]	UK	28	All ovarian malignancies
Di Legge 2012a[62] Di Legge 2012b[61]	Multi-national (worldwide)	2445	All malignant tumours including borderline
Fathallah 2011[64]	France	109	All malignant tumours including borderline
IOTA 2017*[redacted]	[redacted]	[redacted]	[redacted]
Knafel 2015a[49] Knafel 2013b[54]	Poland	226	All malignant tumours including borderline
Meys 2016[44]	Netherlands	326	All malignant tumours including borderline
Murala 2014[60]	UK	51	All malignant tumours (undefined – not clear whether borderline tumours were included)
Piovano 2016[58] Piovano 2015[85]	Italy	391	All malignant tumours including borderline
Ruiz de Gauna 2015[65]	Spain	154	All malignant tumours including borderline
Sayasneh 2013b[63]	UK	255	All malignant tumours including borderline
Silvestre 2015[55]	Brazil	75	All malignant tumours including borderline
Tantipalakorn 2014[51]	Thailand	319 (masses)	All malignant tumours including borderline
Testa 2014[50]	Multi-national (Europe)	2403	All malignant tumours including borderline
Timmerman 2010[66] Ameye 2012[68]	Multi-national (worldwide)	1938	All malignant tumours including borderline
Tinnangwattana 2015[47] Tongsong 2016[59]	Thailand	94	All malignant tumours including borderline
Weinberger 2013[53]	NR	347	All ovarian malignancies including borderline
ADNEX model			
IOTA 2017*[redacted]	[redacted]	[redacted]	[redacted]

Details	Country	N	Main target condition reported
Joyeux 2016[43]	France	284	Ovarian malignancies including borderline
Meys 2016[44]	Netherlands	326	All malignant tumours including borderline
Moffatt 2016[45]	UK		Ovarian malignancies (undefined – not clear whether borderline tumours were included)
Sayasneh 2016[46]	UK and Spain	255	All malignant tumours including borderline
Szubert 2016[42]	Poland and Spain	327	All ovarian malignancies including borderline
van Calster 2014[17]	Multi-national (Europe)	2403	All malignant tumours including borderline
Overa (MIA2G)			
Coleman 2016[71] Wolff 2015[70]	USA	493	All malignant tumours including borderline
Shulman 2016[105]	USA	993	All malignant tumours including borderline
Zhang 2015a[69]	USA	305	All malignant tumours including borderline
RMI threshold variation			
Aktürk 2011[72]	Turkey	100	All ovarian malignancies excluding borderline
Asif 2004[78]	Pakistan	100	All malignant tumours (undefined – not clear whether borderline tumours were included)
Davies 1993[80]	UK	124	All malignant tumours including borderline
Jacobs 1990[79]	UK	139	All malignant tumours including borderline
Lou 2010[74]	China	223	All malignant tumours including borderline
Manjunath 2001[76]	India	148	All malignant tumours excluding borderline
Morgante 1999[81]	Italy	124	All malignant tumours including borderline
Tingulstad 1996[77]	Norway	173	All malignant tumours including borderline
Ulusoy 2007[75]	Turkey	296	All malignant tumours including borderline
Yamamoto 2009[73]	Japan	253	All ovarian malignancies including borderline

* Personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017
Note that some studies evaluated multiple risk scores

3.2.2 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 6 and Figure 2). One of these studies[17] reported the development and validation of the ADNEX model, in addition to test accuracy results. This study was assessed using PROBAST, a tool specifically developed to assess the methodological quality of prediction modelling studies, (Table 7) as well as 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.

Eight studies were reported only as conference abstracts or meeting slides, with limited descriptions of methods,[45, 53, 60, 67, 69, 91, 95, 105] and study methods were generally poorly reported. Thirty-seven (73%) studies were rated as 'unclear' risk of bias on at least one QUADAS-2 domain whilst 24 (47%) studies were rated as 'unclear' for applicability on at least one domain.

Two studies[65, 71] were rated as '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 (22%) of studies were rated as having 'low' concerns regarding all applicability domains.[43, 52, 65, 71-73, 76, 78, 83, 84, 90]

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

The main potential sources of bias, across the published included studies, concerned flow and timing. Fifteen (30%) studies were rated as 'high' risk of bias on the flow and timing domain. For most studies, 13/15,[45, 47, 51, 60, 64, 66, 67, 79, 83, 95, 99, 102, 103] this was because not all included patients were included in the analysis. In five studies, the included patients did not all receive the same reference standard.[51, 76, 79, 80, 84]

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 (28%) studies were rated as having 'high' concerns regarding the applicability of the index test; for six studies this was because all or part of the index test was performed before referral,[48, 50, 62, 66, 95, 99] in three studies the index test was applied retrospectively to existing patient data,[45, 53, 67] and in seven studies 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.[42, 44, 49, 50, 53, 66, 75] Eighteen (35%) studies 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 over estimation of test performance. Appendix 2 (Table 36) lists final histological diagnoses (where reported) of study participants. These data illustrate the between study variation in the definitions of disease positive used, which could include borderline, non-ovarian cancers, metastatic cancers, non-ovarian metastatic cancers. To take into account as much of this heterogeneity as possible, we analysed the results according to whether disease positive (target condition) was defined as 'ovarian malignancy' or 'any malignant tumour' and whether or not this definition included borderline tumours.

[redacted]

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

The PROBAST prediction score (Table 7) for Van Calster 2014[17] indicated that there was high risk of bias for the applicability of patient selection. The high risk of bias was due 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 since not all aspects of the model development were clearly described.

Table 6: QUADAS-2 results for accuracy studies of risk scores

Study	RISK OF BIAS				APPLICABILITY		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW & TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Aarenstrup 2012[84]	?	+	?	-	+	+	+
Abdalla 2013[48]	?	+	?	+	-	-	-

Aktürk 2011[72]	?	+	?	?	+	+	+
Al Musalhi 2016[104]	?	+	?	?	+	+	-
Alcazar 2013[52]	?	+	?	+	+	+	+
Asif 2004[78]	+	?	?	?	+	+	+
Baker 2013[67]	-	?	?	-	+	-	?
Chan 2013[83]	+	+	+	-	+	+	+
Clemente 2015[91]	?	?	+	+	?	+	+
Coleman 2016[71]	+	+	+	+	+	+	+
Davies 1993[80]	+	+	-	-	+	+	?
Di Legge 2012[62]	+	+	?	+	-	-	?
Fathallah 2011[64]	+	+	+	-	?	+	+
IOTA5 2017*							
Jacobs 1990[79]	+	+	-	-	+	+	?
Janas 2015[98]	?	+	?	+	?	+	+
Joyeux 2016[43]	?	?	?	+	+	+	+
Knafel 2015[49]	+	+	?	+	?	-	-
Langhe 2013[95]	?	+	?	-	?	-	?
Li 2016[97]	+	+	?	?	?	+	?
Lou 2010[74]	?	?	+	?	+	+	-
Manjunath 2001[76]	+	+	+	-	+	+	+
Meys 2016[44]	+	-	+	+	?	-	+
Moffatt 2016[45]	?	?	+	-	+	-	?
Moore 2011[102]	?	+	?	-	-	+	+
Morgante 1999[81]	+	?	?	?	+	+	?
Murala 2014[60]	+	-	?	-	+	?	?
Novotny 2012[87]	?	+	?	?	+	+	?
Piovano 2016[58]	+	+	+	+	?	+	-
Presl 2012[82]	?	?	?	?	+	?	?
Ruiz de Gauna 2015[65]	+	+	+	+	+	+	+
Sayasneh 2013b[63]	+	+	?	+	-	+	-
Sayasneh 2016[46]	+	?	+	+	+	+	-
Shulman 2016[105]	?	?	?	?	+	+	-
Silvestre 2015[55]	+	+	+	+	+	+	-
Szubert 2016[42]	?	+	?	+	+	-	-
Tantipalakorn 2014[51]	?	+	?	-	+	+	-
Testa 2014[50]	+	+	+	+	-	-	-
Timmerman 2010[66]	+	+	+	-	-	-	-
Tingulstad 1996[77]	-	+	?	?	+	+	-
Tinnangwattana 2015[47]	+	+	+	+	+	+	-
Ulusoy 2007[75]	+	+	?	+	-	-	-
Van Calster 2014[17]	+	+	+	?	+	+	-
Van Gorp 2012[99]	+	+	?	-	?	-	+
Weinberger 2013[53]	?	?	?	?	?	-	?
Winarto 2014[100]	?	+	?	+	?	+	+
Xu 2016[96]	-	+	?	+	?	+	+
Yamamoto 2009[73]	?	+	?	+	+	+	+

Yanaranop 2016[90]	?	+	+	+	+	+	+
Zhang 2015[69]	?	?	?	?	+	+	?
Zhang 2015b[103]	-	+	?	-	?	+	+
+ Low Risk; - High Risk; ? Unclear Risk							

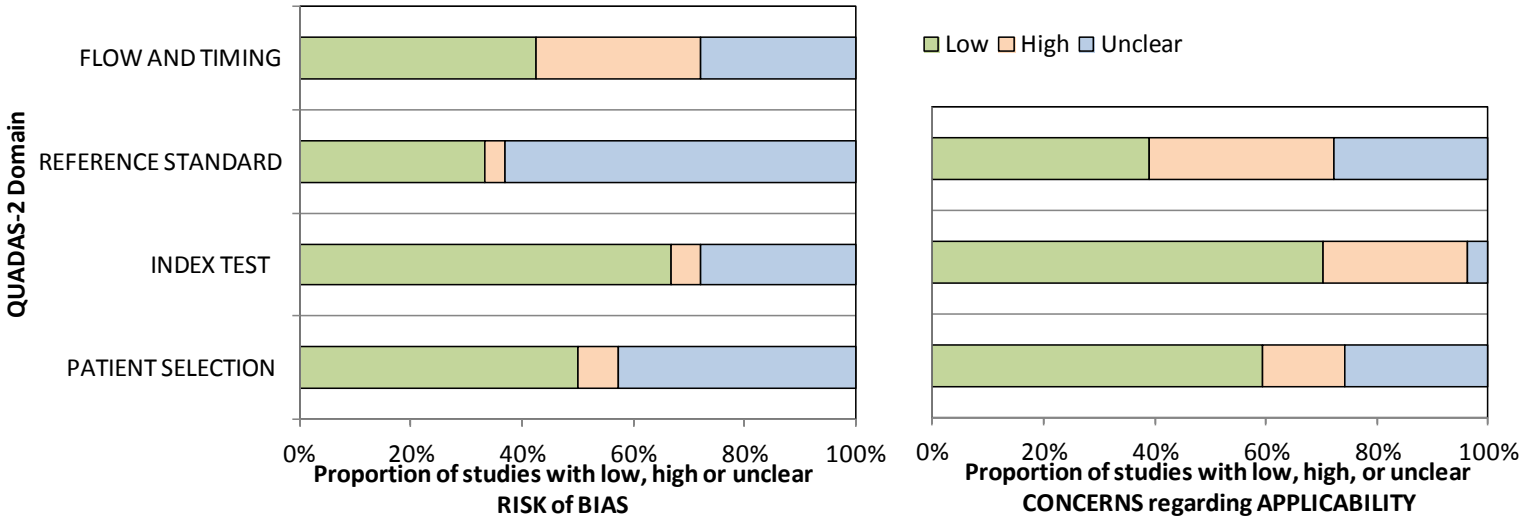
*:personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

Table 7: PROBAST results for studies reporting the development and validation of risk scores

Study	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 2014[17]	+	+	+	+	?	?	?	+	?	-	-	+	+	+	+	+

+ Low Risk; - High Risk; ? Unclear Risk

Figure 2: Summary of QUADAS-2 results for accuracy studies of risk scores



3.2.3 Diagnostic performance of the ROMA score

Details of ROMA studies

Sixteen diagnostic cohort studies,[82-84, 87, 90, 91, 95-100, 102-105] reported in 24 publications[56, 82-84, 86-105] provided data on the diagnostic performance of the ROMA score for identifying people who have an adnexal mass, who are at high risk of ovarian cancer. Nine studies used a ROMA score based on Abbott ARCHITECT tumour marker assays,[82-84, 87, 91, 97, 100, 102, 104] of which, six evaluated a decision threshold for the ROMA score consistent with the manufacturer's recommendations.[82, 83, 87, 97, 100, 104] None of the included studies used the Fujirebio Lumipulse G automated CLEIA system. For information, we have include two studies which used a ROMA score based on manual Fujeribio tumour marker EIA assays (Appendix 4, Tables 41 and 42),[95, 99] 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 used a ROMA score based on Roche Elecsys tumour marker assays,[90, 96, 98, 103, 105] all of which used the manufacturer's recommended decision threshold for the ROMA score.

None of the ROMA score studies, which used Abbott ARCHITECT tumour marker assays, were conducted in the UK; three studies were conducted in European countries,[82, 84, 87] four were conducted in Asia,[83, 91, 97, 100] one was conducted in the USA[102] and one in Oman.[104] None of the ROMA score studies, which used Roche Elecsys tumour marker assays, were conducted in the UK and only one was conducted in a European country.[98] Three of the remaining studies were conducted in Asia[90, 94, 96] and one in the USA.[105]

This assessment is primarily concerned with providing a comparison between the RMI 1,[79] used with a decision threshold of 250 (current standard practice in the UK NHS[1]) and the specified alternative risk scoring methods (see section 2.3). We did not identify any studies which reported a direct comparison (both tests used to assess the same patient cohort) between the ROMA score and RMI 1, used with a decision threshold of 250. Five studies reported direct comparisons between the ROMA score and RMI 1, used with a decision threshold of 200; three used Abbott ARCHITECT tumour marker assays,[84, 100, 104] one used Roche Elecsys assays,[90] and one used Fujirebio manual EIAs.[99] 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 consistent with the manufacturers' recommendations are used. Additional accuracy data, for alternative decision thresholds, are reported in Appendix 4, 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 people 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. Additionally, some studies included patients with non-ovarian primarily cancers/metastases to the ovary,[98, 99, 104] and germ cell tumours.[104] Where the target condition is described as 'all ovarian malignancy', those participants whose post-operative, histological diagnosis identified a non-ovarian primary were excluded from estimates of test performance. Conversely, where the target condition is 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, which 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 included studies). Full details of the final histopathological diagnoses of study participants who had a malignant mass are reported in Appendix 2 (Table 36).

Accuracy of the ROMA score using Abbott ARCHITECT tumour marker assays

Three[84, 100, 104] of the nine[82-84, 87, 91, 97, 100, 102, 104] ROMA score studies, which used Abbott ARCHITECT tumour marker assays, reported a direct comparison of the ROMA score with the RMI 1. Only one study included all participants in 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 manufacture (13.1% in pre-menopausal women and 27.7% in post-menopausal women, as opposed to the manufacturer's recommendation of 7.4% and 25.3%).[104] One study was a retrospective study, which excluded women with histopathological diagnoses other than epithelial ovarian cancer.[100] A second study excluded from the analysis nine (1%) participants with non-epithelial ovarian cancer, 69 (6%) participants with non-ovarian cancers, and 252 (21%) participants with borderline tumours;[84] the distribution of positive and negative ROMA score results, in these patients, was not reported.

The sensitivity estimate for the ROMA score was highest, 96.4% (95% CI: 93.6 to 98.2%), where analyses excluded participants with borderline tumours and those with malignancies other than epithelial ovarian cancer and lowest, 75.0% (95% CI: 60.4 to 86.4%), where all participants were included in the analysis, regardless of their final histopathological

diagnosis (see Table 8). Conversely, the specificity estimate for the ROMA score was highest, 87.9% (95% CI: 81.9 to 92.4%), in the study which included all participants[104] and lowest, 53.3% (95% CI: 50.0 to 56.7%), where analyses excluded participants with borderline tumours and those with malignancies other than epithelial ovarian cancer (see Table 8). Where participants 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), whilst specificity estimates were significantly lower (see Table 8). In contrast, the study which included all participants in the analysis reported similar sensitivity and specificity estimates for the ROMA score and the RMI 1, 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.[104] This study also reported lower sensitivity and higher specificity estimates, for both the ROMA score and the RMI 1, in pre-menopausal women compared to post-menopausal women (see Table 8).

One study reported test performance estimates calculated both with and without the inclusion of participants with borderline tumours.[100] Although the number of participants involved was small, these data indicated that around half of false negative risk scores were accounted for by patients with borderline tumours, 3/6 (50%) using the ROMA score and 7/13 (54%) using the RMI 1 (threshold 200).[100] Approximately 13% (17/128) of the participants in this study had borderline tumours, whilst 39% (50/128) had malignant tumours; i.e. a higher proportion of patients with borderline tumours had a negative ROMA score, 17.6% (3/17), than was the case for patients with malignant tumours, 3/50 (6%).[100] A similar pattern was observed for RMI1; the proportion of patients with borderline tumours who had a negative RMI 1 was approximately 41% (7/17), compared to 12% (6/50) for those with malignant ovarian tumours.[100]

One additional study reported performance estimates for the ROMA score, excluding patients with borderline tumours and those with none ovarian malignancies, without a comparison to the RMI1.[83] When data from this study were combined with the ROMA data from the two similar comparative accuracy studies,[84, 100] the summary estimates of sensitivity did not change significantly, 95.1% (95% CI: 92.4 to 97.1%) based on three studies (see Table 9), versus 96.4% (95% CI: 93.6 to 98.2%) based on two studies (see Table 8). The summary estimate of specificity, based on all three studies 62.5% (95% CI: 59.7 to 65.3%) (see Table 9), was higher than that derived from the two comparative accuracy studies alone

53.3% (95% CI: 50 to 56.7%) (see Table 8). There were no additional studies which evaluated the performance of the ROMA score alone and included all participants in the analysis (target condition: all malignant tumours including borderline).

One study assessed the variation in the performance of the ROMA score with different stages of epithelial ovarian cancer (see Table 9).[83] The sensitivity estimate was highest, 92.1% (95% CI: 78.6 to 98.3%), where the target condition was stage III/IV epithelial ovarian cancer and patients with stage I/II and borderline disease were excluded from the analysis.[83] There was small, but non-significant, fall in sensitivity, 82.6% (95% CI: 61.2 to 95.0%), where the target condition was stage I/II epithelial ovarian cancer and patients with borderline and higher stage disease were excluded from the analysis.[83] When the target condition was borderline epithelial tumours and all patients with higher stage disease were excluded from the analysis, the sensitivity estimate was significantly lower, 56.3% (95% CI: 29.9 to 80.2%).[83] These data are consistent with the observation that the proportion of patients with a negative ROMA score is higher amongst those with borderline disease than amongst those with ovarian malignancies and may also be higher amonst those with lower stage epithelial ovarian cancer than those with higher stage.

Two studies reported accuracy data for ovarian malignancy but without clarifying whether or not the definition of malignancy included borderline tumours (Appendix 4, Table 43).[82, 97] Accuracy data for thresholds other than those recommended by the manufacturer (7.4% in pre-menopausal women and 25.3% in post-menopausal women) are reported in Appendix 4, 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.

Table 8: Comparative accuracy of the ROMA score using Abbott ARCHITECT tumour marker assays versus the RMI

study ID	Subgroup	ROMA Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	RMI	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline																	
Al Musalhi 2016[104]	All	13.1%/27.7%	36	12	20	145	213	75.0 (60.4, 86.4)	87.9 (81.9, 92.4)	200	37	11	30	135	213	77.1 (62.7, 88.0)	81.8 (75.1, 87.4)
	Pre-menopausal	13.1%	11	10	14	127	162	52.4 (29.8, 74.3)	90.1 (83.9, 94.5)	200	12	9	21	120	162	57.1 (34.0, 78.2)	85.1 (78.1, 90.5)
	Post-menopausal	27.7%	25	2	5	19	51	92.6 (75.7, 99.1)	79.2 (57.8, 92.9)	200	22	2	9	18	51	91.7 (73.0, 99.0)	66.7 (46.0, 83.5)
Target condition: Epithelial ovarian malignancies including borderline																	
Winarto 2014[100]	All	7.4%/25.3%	61	6	35	26	128	91.0 (81.5, 96.6)	42.6 (30.0, 55.9)	200	54	13	21	40	128	80.6 (69.1, 89.2)	65.6 (52.3, 77.3)
Target condition: Epithelial ovarian malignancies excluding borderline																	
Aarenstrup 2012[84]	All	7.4%/25.3%	244	8	371	438	1061	96.8 (93.8, 98.6)	54.1 (50.6, 57.6)	200	238	14	150	659	1061	94.4 (90.9, 96.9)	81.5 (78.6, 84.1)
Winarto 2014[100]	All	7.4%/25.3%	47	3	35	26	111	94.0 (83.5, 98.7)	42.6 (30.0, 55.9)	200	44	6	21	40	111	88.0 (75.7, 95.5)	65.6 (52.3, 77.3)
Summary estimates								**96.4 (93.6, 98.2)**	**53.3 (50.0, 56.7)**	**Summary estimates**						**93.4 (90.0, 95.9)**	**80.3 (77.5, 82.9)**
Aarenstrup 2012[84]	Pre-menopausal	7%	46	3	251	279	579	93.9 (83.1, 98.7)	52.6 (48.3, 57.0)	200	41	8	42	488	579	83.7 (70.3, 92.7)	92.1 (89.4, 94.2)
	Post-menopausal	25.3%	198	5	120	159	482	97.5 (94.3, 99.2)	57.0 (51.0, 62.9)	200	196	7	108	171	482	96.6 (93.0, 98.6)	61.3 (55.3, 67.0)

Table 9: Accuracy of the ROMA score using Abbott ARCHITECT tumour marker assays at the manufacturer's recommended thresholds

study ID	Subgroup	Threshold	TP	FN	FP	TN	Total N	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: Epithelial ovarian malignancies including borderline									
Winarto 2014[100]	All	7.4%/25.3%	61	6	35	26	128	91.0 (81.5, 96.6)	42.6 (30.0, 55.9)
Target condition: Epithelial ovarian malignancies excluding borderline									
Aarenstrup 2012[84]	All	7.4%/25.3%	244	8	371	438	1061	96.8 (93.8, 98.6)	54.1 (50.6, 57.6)
Chan 2013[83]	All	7.4%/25.3%	58	7	41	281	387	89.2 (79.1, 95.6)	87.3 (83.1, 90.7)
Winarto 2014[100]	All	7.4%/25.3%	47	3	35	26	111	94.0 (83.5, 98.7)	42.6 (30.0, 55.9)
Summary estimates								**95.1 (92.4, 97.1)**	**62.5 (59.7, 65.3)**
Aarenstrup 2012[84]	Pre-menopausal	7%	46	3	251	279	579	93.9 (83.1, 98.7)	52.6 (48.3, 57.0)
Chan 2013[83]	Pre-menopausal	7%	18	4	34	235	291	81.8 (59.7, 95.9)	87.4 (82.8, 91.1)
Summary estimates								**90.1 (80.7, 95.9)**	**64.3 (60.9, 67.7)**
Aarenstrup 2012[84]	Post-menopausal	25.3%	198	5	120	159	482	97.5 (94.3, 99.2)	57.0 (51.0, 62.9)
Chan 2013[83]	Post-menopausal	25.3%	40	3	7	46	96	93.0 (80.9, 98.5)	86.8 (56.3, 67)
Summary estimates								**96.7 (93.7, 98.6)**	**61.7 (56.3, 67.0)**
Target condition: Epithelial ovarian malignancies (stage III/IV) – borderline and stage I/II tumours excluded									
Chan 2013[83]	All	7.4%/25.3%	35	3	41	281	360	92.1 (78.6, 98.3)	87.3 (83.1, 90.7)
	Pre-menopausal	7%	10	2	34	235	281	83.3 (51.6, 97.9)	87.4 (82.8, 91.1)
	Post-menopausal	25.3%	24	1	7	46	78	96.0 (79.6, 99.9)	86.8 (74.7, 94.5)

study ID	Subgroup	Threshold	TP	FN	FP	TN	Total N	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: Epithelial ovarian malignancies (stage I/II) – borderline and stage III/IV tumours excluded									
Chan 2013[83]	All	7.4%/25.3%	19	4	41	281	345	82.6 (61.2, 95.0)	87.3 (83.1, 90.7)
	Pre-menopausal	7%	6	2	34	235	277	75.0 (34.9, 96.8)	87.4 (82.8, 91.1)
	Post-menopausal	25.3%	12	2	7	46	67	85.7 (57.2, 98.2)	86.8 (74.7, 94.5)
Target condition: Ovarian borderline tumours – higher stage tumours excluded									
Chan 2013[83]	All	7.4%/25.3%	9	7	41	281	338	56.3 (29.9, 80.2)	87.3 (83.1, 90.7)
	Pre-menopausal	7%	6	2	34	235	277	75.0 (34.9, 96.8)	87.4 (82.8, 91.1)
	Post-menopausal	25.3%	12	2	7	46	67	85.7 (57.2, 98.2)	86.8 (74.7, 94.5)

[c]: 2x2 data were calculated (other studies reported 2x2 data); *:calculated confidence intervals

Accuracy of the ROMA score using Fujirebio tumour marker assays

None of the included studies used the Fujirebio Lumipulse G automated CEIA system and hence there were no studies of the ROMA score, using Fujirebio assays, which met the inclusion criteria for this assessment. We have included two studies that evaluated a ROMA score based on manual Fujeribio tumour marker EIA assays in this report.[95, 99] These studies are included for information only. Both of these studies included all participant in the analysis, regardless of their final histopathological diagnosis (target condition: all malignant tumours including borderline). One study reported a direct comparison of the ROMA score with the RMI 1 (threshold 200).[99] The results of these studies are provided in Appendix 4, Tables 41 and 42.

Accuracy of the ROMA score using Roche tumour marker assays

Only one[90] of the five[90, 96, 98, 103, 105] ROMA score studies, which used Roche Elecsys tumour marker assays, reported a direct comparison of the ROMA score with the RMI 1 (threshold 200). This study classified patients found to have borderline ovarian tumours as disease negative and included patients whose final histopathological diagnoses were epithelial ovarian cancer, non-epithelial ovarian cancer and metastases form non-ovarian primaries (target condition all malignant tumours).[90] This study may be considered more applicable to clinical practice, if it is considered preferable to manage patients with borderline tumours in non-specialist settings. In these patients, the sensitivity estimate for the ROMA score appeared slightly higher than that for the RMI 1, 83.8% (95% CI: 73.4 to 91.3%) versus 78.4 (95% CI: 67.3 to 87.1%) and the specificity estimate for the ROMA score appeared slightly lower than that for the RMI 1, 68.8% (95% CI: 61.6 to 75.4%) versus 79.6% (95% CI: 73.1 to 85.1%), but neither difference was statistically significant.[90] A similar pattern was observed when data were stratified by menopausal status (see Table 10). The same study also reported test performance data, where eight (3%) patients 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 RMI1 (see Table 10). Although the number involved was small, it should be noted that patients with malignancies other than epithelial ovarian cancer accounted for four (50%) of the false negative results using the ROMA score and three (37.5%) using the RMI 1.[90]

The comparative accuracy study described above also assessed the variation in the performance of the ROMA score with different stages of epithelial ovarian cancer (see Table 10).[90] The sensitivity estimate was highest, for both the ROMA score 97.2% (95% CI: 95.5 to 99.9%) and the RMI 1 88.9% (95% CI: 73.9 to 96.9%), where the target condition was stage II to IV epithelial ovarian cancer and patients with stage I disease were excluded from the analysis.[90] As with the Abbott ARCHITECT

ROMA score, sensitivity estimates were lower, for both the ROMA score 76.7% (95% CI: 57.7 to 90.1%) and the RMI 1 70.0% (95% CI: 50.6 to 85.3%), where the target condition was stage I epithelial ovarian cancer and patients with higher stage disease were excluded from the analysis.[90] This indicates that the proportion of patients with a negative ROMA score may be higher amonst those with lower stage epithelial ovarian cancer than those with higher stage.

Two[98, 105] of the four[96, 98, 103, 105] additional studies, which evaluated the performance of the ROMA score but did not provide a comparison with the RMI 1, included all study participants in the analysis regardless of their final histopathological diagnoses (target condition all malignant tumours including borderline), (see Table 11). 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 described above, where participants with borderline tumours were classified as disease negative,[90] and the summary specificity estimate 79.1% (95% CI: 76.3 to 81.6%), but these differences were not statistically significant. Two studies reported test performance data for the ROMA score, where patients found to have borderline tumours and those with non-ovarian primaries were excluded from the analyses.[96, 98] The sensitivity estimates derived from these two studies were very different (see Table 11) and hence no summary estimates were calculated. One of these studies reported test performance estimates calculated both with and without the inclusion of participants with borderline tumours and those with non-ovarian primaries.[98] As with the ROMA score using Abbott ARCHITECT tumour marker assays, these data indicated that patients with borderline tumours and those with non-ovarian primaries accounted for a high proportion, 12/14 (86%), of the false negative risk scores observed.[98]

One study[103] provided performance estimates for the ROMA score, using Roche Elecsys tumour marker assays, for the target condition 'ovarian malignancy', where it was not clear whether or not the definition of malignancy included of borderline tumours. (Appendix 4, Table 44). Accuracy data for thresholds other than those recommended by the manufacturer (11.4% in pre-menopausal women and 29.9% in post-menopausal women) are reported in Appendix 4, 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 manufacturer's tumour marker assays. However, between study comparisons indicate that, where all study participants were included in the analyses regardless of final histopathological diagnosis (target condition all malignant tumours including borderline), estimates of sensitivity did not differ

significantly between the two manufacturers' assays for which data were available (see Figure 3). The sensitivity estimate for the ROMA score, using Abbott ARCHIECT was 75.0% (95% CI: 60.4 to 86.4%), derived from one study,[104] compared to 79.1% (95% CI: 74.2 to 83.5%) Roche Elecsys, derived from two studies.[98, 105] However, the specificity estimate for Abbott ARCHIECT, 87.9% (95% CI: 81.9 to 92.4%), was higher than that for Roche Elecsys, 79.1% (95% CI: 76.3 to 81.6%). There were no studies of the ROMA score using Fujirebio 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 manufacturer's tumour marker assays for detecting different stages of disease.

Table 10: Comparative accuracy of the ROMA score using Roche tumour marker assays versus the RMI

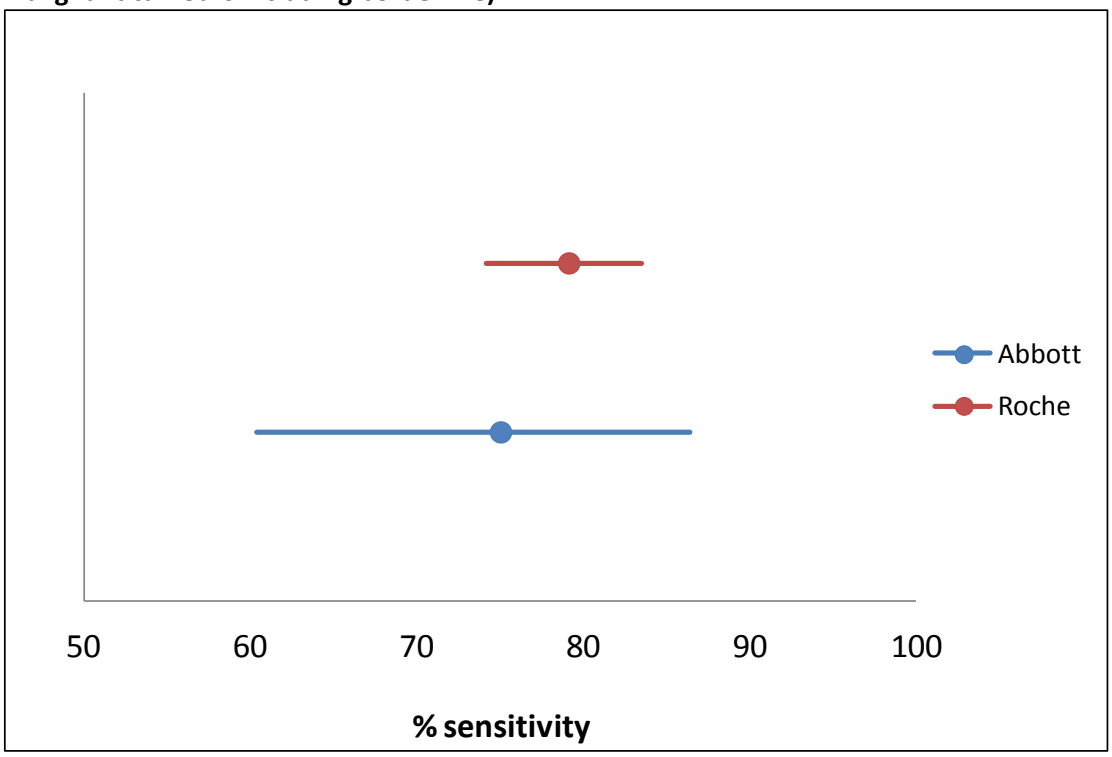
study ID	Sub-group	ROMA Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	RMI	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours – borderline tumours classified as disease negative																	
Yanaranop 2016[90]	All	11.4%/29.9%	62	12	58	128	260	83.8 (73.4, 91.3)	68.8 (61.6, 75.4)	200	58	16	38	148	260	78.4 (67.3, 87.1)	79.6 (73.1, 85.1)
	Pre-menopausal	11.4%	24	4	35	85	148	85.7 (67.3, 96.0)	70.8 (61.8, 78.8)	200	21	7	23	97	148	75.0 (55.1, 89.3)	80.8 (72.6, 87.4)
	Post-menopausal	29.9%	38	8	23	43	112	82.6 (68.6, 92.2)	65.2 (52.4, 76.5)	200	37	9	15	51	112	80.4 (66.1, 90.6)	77.3 (65.3, 86.7)
Target condition: Epithelial ovarian malignancies – borderline tumours classified as disease negative																	
Yanaranop 2016[90]	All	11.4%/29.9%	58	8	58	128	252	87.9 (77.5, 94.6)	68.8 (61.6, 75.4)	200	53	13	38	148	252	80.3 (68.7, 89.1)	79.6 (73.1, 85.1)
Target condition: Epithelial ovarian malignancies (stage I) – borderline tumours classified as disease negative and higher stage tumours excluded																	
Yanaranop 2016[90]	All	11.4%/29.9%	23	7	58	128	216	76.7 (57.7, 90.1)	68.8 (61.6, 75.4)	200	21	9	38	148	216	70.0 (50.6, 85.3)	79.6 (73.1, 85.1)
Target condition: Epithelial ovarian malignancies (II-IV) – borderline tumours classified as disease negative and stage I tumours excluded																	
Yanaranop 2016[90]	All	11.4%/29.9%	35	1	58	128	222	97.2 (85.5, 99.9)	68.8 (61.6, 75.4)	200	32	4	38	148	222	88.9 (73.9, 96.9)	79.6 (73.1, 85.1)

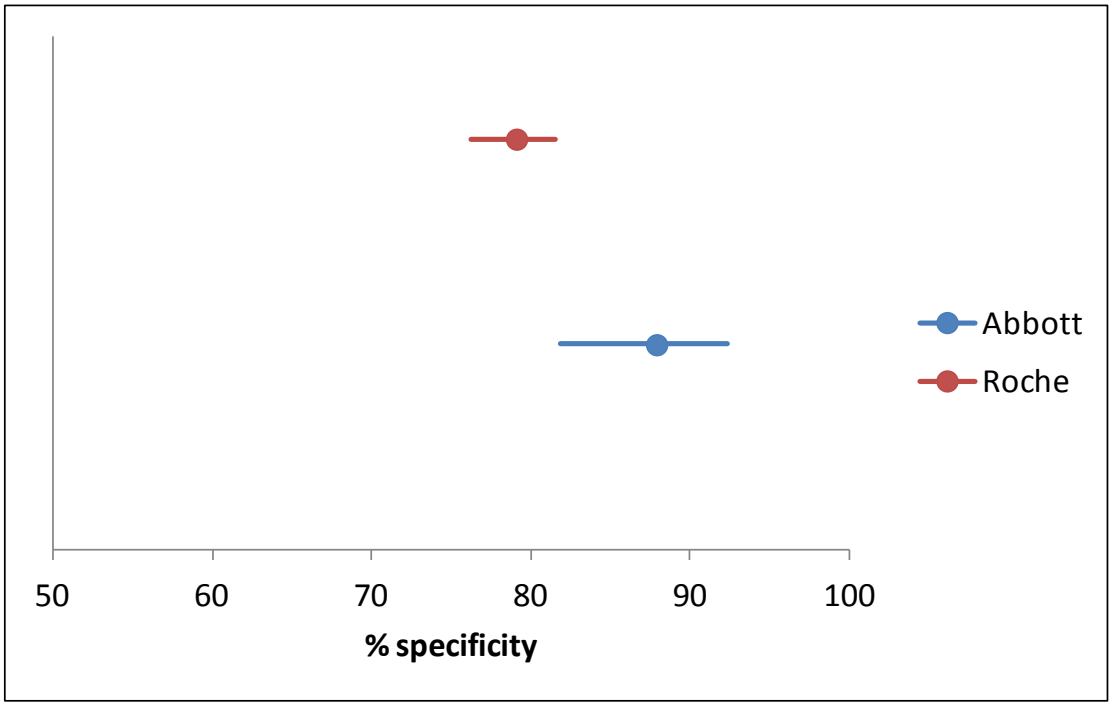
Table 11: Accuracy of the ROMA score using Roche tumour marker assays at the manufacturer's recommended thresholds

study ID	Subgroup	Thresh-old	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline									
[C]Janas 2015[98]	All	11.4%/29.9%	52	14	39	154	259	78.8 (67.0, 87.9)	79.8 (73.4, 85.2)
Shulman 2016[105]	All	11.4%/29.9%	194	51	158	590	993	79.2 (73.7, 83.8)	78.9 (75.8, 81.7)
Summary estimates								**79.1 (74.2, 83.5)**	**79.1 (76.3, 81.6)**
[C]Janas 2015[98]	Pre-menopausal	11.4%	9	1	22	100	132	90.0 (55.5, 99.7)	82.0 (74.0, 88.3)
[C]Janas 2015[98]	Post-menopausal	29.9%	44	12	17	54	127	78.6 (65.6, 88.4)	76.1 (64.5, 88.4)
Target condition: Ovarian malignancies excluding borderline									
[C]Janas 2015[98]	All	11.4%/29.9%	42	2	39	154	237	95.5 (84.5, 99.4)	79.8 (73.4, 85.2)
Xu 2016[96]	All	11.4%/29.9%	113	97	39	272	521	53.8 (46.8, 60.7	87.5 (83.3, 90.9)
[C]Janas 2015[98]	Pre-menopausal	11.4%	6	0	22	100	128	100 (54.1, 100)	82.0 (74.0, 88.3)
Xu 2016[96]	Pre-menopausal	11.4%	56	51	38	226	371	54.9 (42.5, 62.1)	85.6 (80.8, 89.6)
[C]Janas 2015[98]	Post-menopausal	29.9%	36	2	17	54	109	94.7 (82.3, 99.4)	76.1 (64.5, 85.4)
Xu 2016[96]	Post-menopausal	29.9%	57	46	1	46	150	53.3 (45.2, 65.1)	97.9 (88.7, 99.9)

[C]: 2x2 data were calculated (other studies reported 2x2 data); [*]:calculated confidence intervals

Figure 3: Comparison of the accuracy of ROMA Abbott versus ROMA Roche (Target condition all malignant tumours including borderline)





3.2.4 Diagnostic performance of IOTA simple ultrasound rules and the ADNEX model

Details of ADNEX studies

Six published studies,[17, 42-46] and one unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) 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 are at high risk of ovarian cancer and all used a version of ADNEX which included CA125 level. Four of the six published studies did not report any details about the experience of those performing the ultrasound examinations.[42-45] One study reported that ultrasound examinations were performed by ultrasound examinations performed by EFSUMB level 2 ultrasound examiners (nonconsultant gynaecology specialist, gynaecology trainees doctors and gynaecology sonographers),[46] and the remaining study used EFSUMB level 2 or 3 practitioners with 8 to 20 years experience in gynaecological sonography.[17] [redacted]

This section reports only accuracy data for the 10% threshold. Three studies provided accuracy data for additional thresholds and these are reported in Appendix 4, Table 38.[17, 43, 46] All studies in this section were conducted in Europe; one was conducted solely in the UK,[45] and two were multi-centre studies which included UK participants.[17, 46]

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 people with one or more adnexal mass and all but one[45] included borderline tumours in their definition of malignancy; this study was only reported as a conference abstract and it was not clear whether any borderline tumours were included (Appendix 4, Table 45). Three published studies[17, 44, 46] and the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) included participants with 'other malignancies', metastases from non-ovarian sites and 'non-ovarian cancers'. Where the target condition is described as 'all ovarian malignancy', those participants whose post-operative, histological diagnosis identified a non-ovarian primary were excluded from estimates of test performance. Conversely, where the target condition is 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, which 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 included studies). Full details of the final histopathological diagnoses of study participants who had a malignant mass are reported in Appendix 2 (Table 36).

Accuracy of the ADNEX model for determining high risk of ovarian cancer

Three published studies[17, 44, 46] and the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) 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%) (see Table 12). These estimates did not differ significantly from those calculated from only those studies of the ADNEX model that reported a direct comparison with RMI 1 (threshold 200 or 250) (see Table 15). Two further studies, reporting three data sets, excluded women with histopathological diagnoses other than primary ovarian cancer.[42, 43] The summary estimate of sensitivity, 94% (95% CI: 88.6, 97.4), derived from these studies, did not differ significantly from that derived from the studies which included all participants in their analyses. However, the summary estimate of specificity, 77.6% (95% CI: 73.6, 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 pre-menopausal women than in post-menopausal women. (see Table 12).

Accuracy data for thresholds other than the 10% validated threshold (1%, 3%, 5%, 15%, 20% and 30%) are reported in Appendix 4, Table 38. As might be expected, sensitivity estimates increase and specificity estimates decrease with decreasing threshold.

Details of IOTA simple ultrasound rules studies

Seventeen published studies, [44, 47-52, 55, 58-60, 62-66, 68] and the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) provided data on the diagnostic performance of the IOTA simple ultrasound rules, for the identification of women with an adnexal mass, who are at high risk of ovarian cancer. The majority (11/17) of the published studies,[44, 48-50, 52, 58, 60, 63-65, 67] were conducted in Europe; three of these were conducted in the UK.[60, 63, 67] Two further, worldwide, multi-national studies included UK participants.[62, 66] Two of the remaining studies were conducted in Thailand[47, 51] and one in Brazil.[55] One study did not report sufficient detail to determine geographic location.[53] [redacted]

[redacted]

Three published studies were clearly conducted by the IOTA study core group,[50, 62, 66] 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 AiC (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017).

Ten published studies,[44, 48-50, 52, 55, 58, 62, 63, 66] as well as the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017), included all participants in the analysis; participants with inconclusive IOTA 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 where all participants were included in the analysis. Six further studies excluded participants with inconclusive IOTA simple ultrasound rules assessments from their analysis.[47, 51, 53, 64, 65, 67] The results of these studies are provided in Appendix 4, Table 39. One study did not report sufficient information to determine how participants with inconclusive IOTA simple ultrasound rules assessments were handled.[60]

Accuracy of IOTA simple ultrasound rules for determining high risk of 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, 63, 66] and the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) provided accuracy data for the IOTA simple ultrasound rules, where participants 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 simple ultrasound rules, where participants with inconclusive assessments were assumed to have malignant tumours, which reported a direct comparison with RMI 1 (threshold 200 or 250) (see Tables 15 and 16). Four studies of these studies reported accuracy data stratified by menopausal status.[44, 49, 50, 63] Menopausal status did not significantly affect sensitivity, however, the specificity estimate was significantly higher in pre-menopausal women than in post-menopausal women. (see Table 13).

Seven studies [44, 49, 50, 52, 58, 63, 66] provided accuracy data for the IOTA simple ultrasound rules, where participants with inconclusive assessments were classified by an expert subjective assessment. In this analysis, we have only included studies where the subjective assessment was done by experts or by level 2/3 examiners as per the European Federation Of Societies For Ultrasound In Medicine And Biology (EFSUMB) classification system. 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 13). One of these studies[49] also assessed the effect of the training level of examiners on the diagnostic performance of the IOTA simple ultrasound rules and found no significant differences in test performance between EFSUMB level 2 examiners (see Table 13) and EFSUMB level 1 examiners (see Table 14). However, it should be noted that all examiners received a half day of practical training in IOTA simple rules before the study.

Five of these studies reported accuracy data stratified by menopausal status.[44, 49, 50, 58, 63] Menopausal status did not significantly affect sensitivity, however, the specificity estimate was significantly higher in pre-menopausal women than in post-menopausal women (see Table 13). One study[58] (Appendix 4, Table 39) also assessed whether the addition of biomarkers to the IOTA simple ultrasound rules could improve the diagnostic performance. Where a positive index test was defined as a malignant classification by IOTA simple ultrasound rules (with subjective assessment of inconclusives) and a ROMA score >11.4%/29.9%, the sensitivity and specificity estimates were 90.5% (95% CI: 82.1, 95.8) and 80.1% (95% CI: 75.2, 84.4), respectively. Where a positive index test was defined as a malignant classification by IOTA simple ultrasound rules (with subjective assessment of inconclusives) and a HE4 level ≥70/140, the sensitivity and specificity estimates were 86.9% (95% CI: 77.8, 93.3) and 86.3% (95% CI: 82, 90), respectively. Neither the addition of the ROMA score or the addition of HE4 alone significantly affected estimates of test performance. Finally, where a positive index test was defined as a malignant classification by IOTA simple ultrasound rules (with subjective assessment of inconclusives) and a CA125 level ≥35, the sensitivity estimate was similar to that for IOTA 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, 73.3).

Comparison of IOTA simple ultrasound rules where 'inconclusive treated as malignant' to those where 'inconclusive were classified by an expert' indicates that sensitivity estimates were significantly higher for 'inconclusive treated as malignant', whist specificity was significantly higher for 'inconclusive were classified by an expert'. Therefore 'inconclusive treated as malignant' is better if we do not want to miss people with ovarian cancer whilst 'inconclusive treated as malignant' is better if we do not want to miss-classify people with ovarian cancer.

Comparisons between IOTA simple rules, ADNEX and the RMI 1

This assessment is primarily concerned with providing a comparison between the RMI 1,[79] used with a decision threshold of 250 (current standard practice in the UK NHS[1]) and the specified alternative risk scoring methods (see section 2.3). Our searches did not identify any studies which reported a direct comparison (both tests used to assess the same patient cohort) between the ADNEX model or IOTA simple rules and RMI 1, used with a decision threshold of 250. One published study[44] and the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) reported direct comparisons between the ADNEX model at the 10% threshold, IOTA simple ultrasound rules (where patients with an inconclusive assessment were assumed to have malignant tumours) and RMI 1, used with a decision threshold of 200 (see Table 15). 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 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 RMI 1 at a decision threshold of 200, 66% (95% CI: 62.9 to 69%), was significantly lower than both the ADNEX model and ultrasound simple rules estimates. Conversely, the specificity estimate for RMI 1 at a decision threshold of 200 was significantly higher, 89% (95%CI: 87 to 90.7%) than both the ADNEX model and ultrasound simple rules estimates (see Figure 4). The unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) also reported direct comparisons between the ADNEX model at the 10% threshold, IOTA simple ultrasound rules (where patients with an inconclusive assessment were assumed to have malignant tumours) and 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 15).

Only the published study[44] reported accuracy data stratified by menopausal status. In pre-menopausal women, the ADNEX model at the 10% threshold and the IOTA simple ultrasound rules had similar sensitivities of 100% (95% CI: 86 to 100%) and 94% (95% CI: 77 to 99%) respectively, in comparison to the overall population. The specificities were significantly lower at 71% (95% CI: 61 to 80%) and 76% (95% CI: 66 to 84%), respectively. In post-menopausal women, the ADNEX model at the 10% threshold and the IOTA simple ultrasound rules had similar sensitivities of 98% (95% CI: 91 to 100%) and 93% (95% CI: 85 to 97%) respectively, in comparison to the overall population. The

specificities of 54% (95% CI: 44 to 63%) and 61% (95% CI: 52 to 70%), respectively, were significantly lower than the overall population. RMI 1, using a decision threshold of 200, had a significantly lower sensitivity of 42% (95% CI: 25 to 61%) and significantly higher specificity of 94% (95% CI: 86 to 97%) in pre-menopausal women. Conversely, the sensitivity estimate was higher, 82% (95% CI: 72 to 89%), and the specificity estimate lower, 66% (95% CI: 56 to 74%), in post-menopausal women (but this was not significant).

Four published studies[44, 48, 50, 63] and the unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) reported direct comparisons IOTA simple ultrasound rules (where patients with an inconclusive assessment were assumed to have malignant tumours) and RMI 1, used with a decision threshold of 200 (see Table 16). 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 IOTA 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 IOTA 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%).

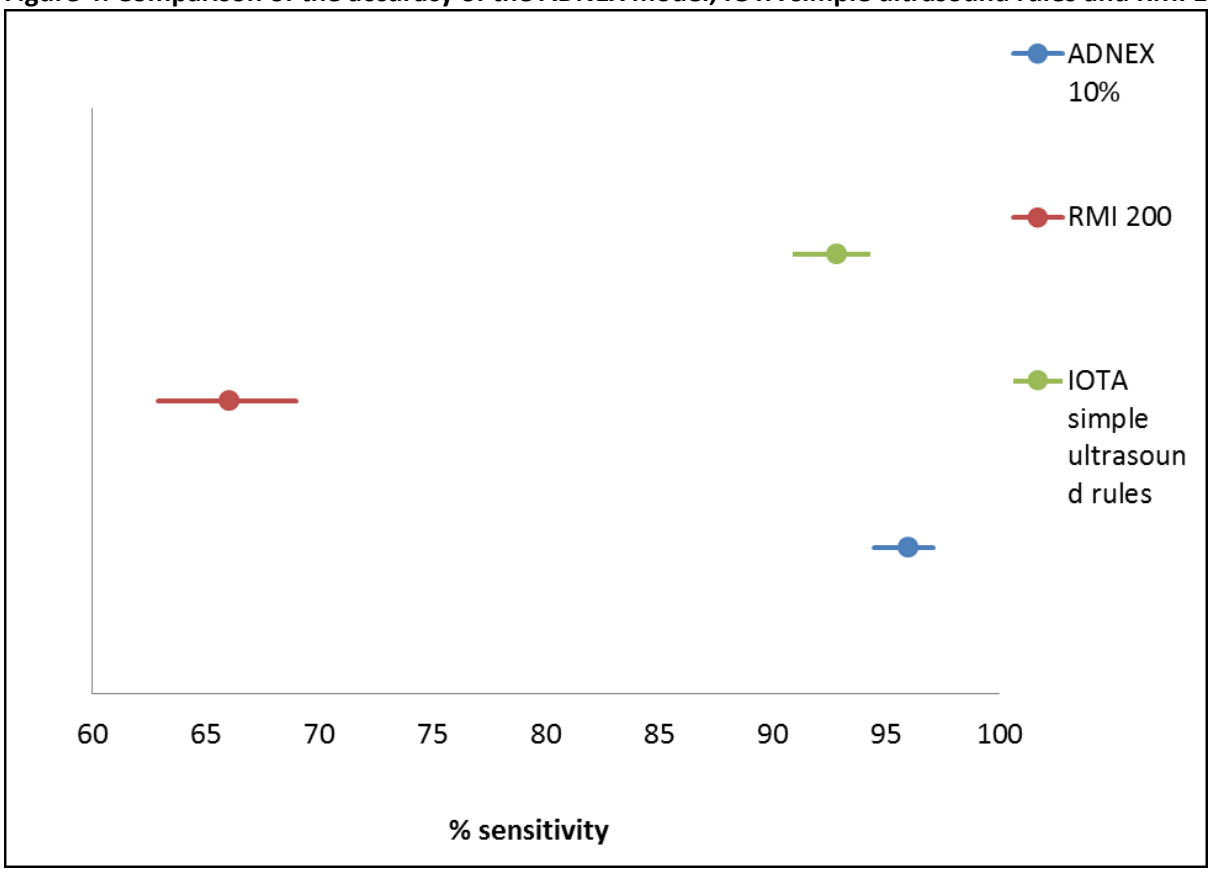
Three of the above studies[44, 50, 63] also reported comparative accuracy data for IOTA simple ultrasound rules versus the RMI 1 (threshold 200), where participants with inconclusive IOTA assessments were classified by an expert subjective assessment of the ultrasound images (see Table 16). The summary estimate of sensitivity for IOTA 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 IOTA 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 16).

In pre-menopausal women using IOTA simple ultrasound rules, where participants with an inconclusive assessment were assumed to have a malignant tumour, 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 post-menopausal women, 95.5% (95% CI: 93.7 to 96.9%) and 72.3% (95% CI: 68.9 to 75.5%), respectively. Where participants with inconclusive IOTA simple rules assessments were classified by an expert subjective assessment, the summary estimates of sensitivity were similar (for both pre- and post-menopausal

women) to those obtained when inconclusive assessments were assumed to be malignant (see Table 16). However, in both pre- and post-menopausal women, the use of subjective assessment significantly increased the summary estimate of specificity to 92% (95% CI: 90.3 to 93.5%) and 80.3% (95% CI: 76.9 to 83.4%), respectively. In pre-menopausal women, the summary sensitivity estimate for RMI 1, 52.2% (95% CI: 47.4 to 56.9%), was significantly lower than those for IOTA simple ultrasound rules and the summary specificity estimates, 94.2% (95% CI: 92.7 to 95.5%), was significantly higher. In post-menopausal women, the summary sensitivity estimate for RMI 1, 78.8% (95% CI: 75.7 to 81.7%), was also significantly lower than those for IOTA simple ultrasound, however, there were no significant differences between the specificity estimates (see Table 16).

One further study, which was only reported as a conference abstract,[60] did not report sufficient information to determine how participants with inconclusive IOTA 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 IOTA 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.

Figure 4: Comparison of the accuracy of the ADNEX model, IOTA simple ultrasound rules and RMI 1



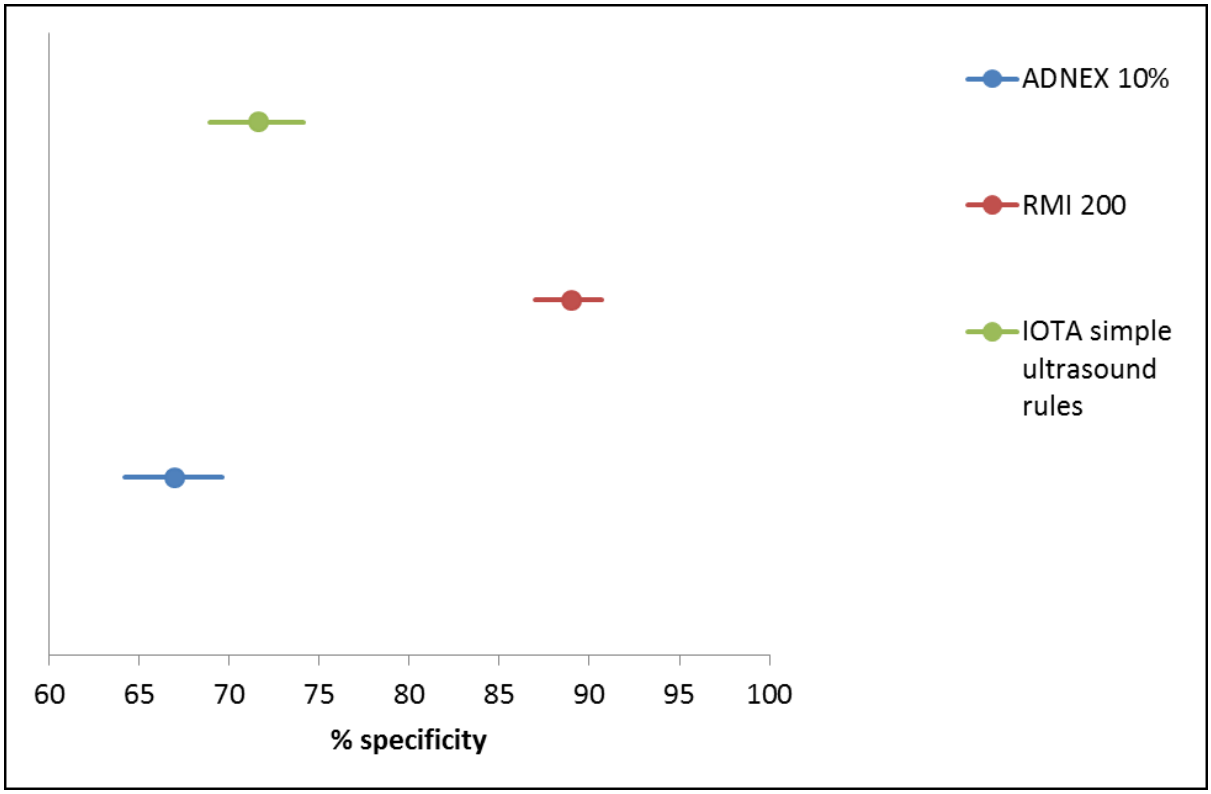

ADNEX 10%
RMI 200
IOTA simple ultrasound rules
60
65
70
75
80
85
90
95
100
% specificity

Table 12: Accuracy of the ADNEX model at a threshold of 10%

Study ID	Subgroup	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline									
IOTA 2017[*,$]	All	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Meys 2016[44]		113	2	80	131	326	calculated	98.0 (93.0, 100)	62.0 (55.0, 68.0)
Sayasneh 2016[46]		177	5	138	290	610	calculated	97.3 (93.5, 98.9)	67.7 (63.0, 72.0)
Van Calster[$] 2014[17]		946	34	408	1015	2403	calculated	96.5 (95.2, 97.6)	71.3 (68.9, 73.7)
Summary estimates								**96.3 (95.3, 97.1)**	**69.1 (67.4, 70.8)**
Meys 2016[44]	pre-menopausal	31	0	28	69	128	calculated	100 (86.0, 100)	71.0 (61.0, 80.0)
Meys 2016[44]	post-menopausal	82	2	52	62	198	calculated	98.0 (91.0, 100)	54.0 (44.0, 63.0)
Target condition: Ovarian malignancies including borderline									
Joyeux 2016[43]	All	27	3	48	206	284	calculated	90 (73.5, 97.9)	81.1 (75.7, 85.7)
Szubert 2016[42]	All- Poland	66	4	37	97	204	reported	94.3 (88.5, 98.7)	72.4 (65.1, 79.7)
	All- Spain	33	1	22	67	123	reported	97.1 (89.7, 100)	75.3 (65.2, 84.7)
Summary estimates								**94 (88.6, 97.4)**	**77.6 (73.6, 81.2)**
Szubert 2016[42]	Poland, pre-menopausal	29	3	23	83	138	calculated	90.6 (77.0, 100)	78.3 (70.7, 85.9)
	Spain pre-menopausal	15	0	11	51	66	calculated	100 (78.2, 100)	82.3 (71.6, 91.1)
Summary estimates								**93.6 (82.5, 98.7)**	**79.8 (72.9, 85.6)**
Szubert 2016[42]	Poland post-menopausal	37	1	14	14	77	calculated	97.4 (91.7, 100)	50.0 (32.1, 69.8)
	Spain Post-menopausal	18	1	11	16	46	calculated	95.8 (85.7, 100)	59.3 (41.5, 77.6)
Summary estimates								**96.5 (87.9, 99.6)**	**54.5 (40.6, 68)**

[*]personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017
[$]data from the IOTA cohort

Table 13: Accuracy of IOTA simple ultrasound rules, where inconclusive results were assumed to be malignant or classified by subjective assessment

Threshold	Study ID	Subgroup	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline										
Malignant (inconclusive were treated as malignant)	Adballa 2013[48]	All	16	1	7	63	87	reported	94.1 (71.3, 99.9)	90.0 (80.5, 95.9)
	Alcazar 2013[52]		51	4	54	231	340	reported	92.7 (82.4, 98.0)	81.1 (76.0, 85.4)
	IOTA 2017[* $]		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Knafel 2015[49]		78	4	15	129	226	reported	95.1 (88.0, 98.7)	89.6 (83.4, 94.1)
	Meys 2016[44]		107	8	67	144	326	calculated	93.0 (86.0, 97.0)	68.0 (61.0, 70.0)
	Sayasneh 2013b[63]		67	7	24	157	255	calculated	91.0 (82.0, 95.0)	87.0 (82.0, 91.0)
	Silvestre 2015[55]		32	0	17	26	75	reported	100 (89.1, 100)	60.5 (44.4, 75.0)
	Testa 2014[$50]		934	46	369	1054	2403	calculated	95.3 (93.1, 96.19)	74.1 (67.7, 79.7)
	Timmerman 2010[$66]		515	27	307	1089	1938	calculated	95.0 (92.0, 96.0)	78.0 (75.0, 80.0)
	Summary estimates								**94.2 (93.3, 95.1)**	**76.1 (74.9, 77.3)**
	Knafel 2015[49]	Pre-menopausal	32	1	9	101	143	calculated	96.9 (84.2, 99.9)	91.9 (85.0, 96.2)
	Meys 2016[44]		29	2	23	74	128	calculated	94.0 (77.0, 99.0)	76.0 (66.0, 84.0)
	Sayasneh 2013b[63]		24	4	16	121	165	calculated	86.0 (69.0, 94.0)	88.0 (83.0, 93.0)
	Testa 2014[$50]		359	19	225	751	1354	calculated	95.0 (91.0, 97.0)	77.0 (70.0, 83.0)
	Summary estimates								**94.5 (92.0, 96.4)**	**79.3 (77.0, 81.5)**
	Knafel 2015[49]	Post-menopausal	46	3	6	28	83	calculated	94 (83.1, 98.7)	81.8 (65.5, 93.2)
	Meys 2016[44]		78	6	44	70	198	calculated	93.0 (85.0, 97.0)	61.0 (52.0, 70.0)
	Sayasneh 2013b[63]		43	3	7	37	90	calculated	93.0 (82.0, 98.0)	84.0 (71.0, 92.0)
	Testa 2014[$50]		578	24	152	295	1049	calculated	96.0 (93.0, 97.0)	66.0 (59.0, 73.0)
	Summary estimates								**95.4 (93.7, 96.8)**	**67.3 (63.5, 70.9)**
Malignant (inconclusive were classified by expert SA)	Alcazar 2013[52]	All	49	6	11	274	340	reported	89.1 (77.8, 95.9)	96.1 (93.2, 98.1)
Malignant (inconclusive	Knafel 2015[49]		78	4	9	135	226	calculated	95.1 (88.0, 98.7)	93.8 (88.5, 97.1)
	Meys 2016[44]		102	13	21	190	326	calculated	89.0 (81.0, 94.0)	90.0 (85.0, 94.0)

Threshold	Study ID	Subgroup	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
were classified by level 2 or level 3 SA)	Piovanono 2016[58]		69	15	23	284	391	calculated	82.1 (72.3, 89.6)	92.5 (89, 95.2)
	Sayasneh 2013b[63]		64	10	11	170	255	calculated	86.0 (77.0, 92.0)	94.0 (90.0, 97.0)
	Testa 2014[$50]		900	80	157	1266	2403	calculated	91.8 (89.1, 93.9)	89.0 (85.2, 92.0)
	Timmerman 2010[$66]		494	102	48	1294	1938	calculated	91.0 (88.0, 93.0)	93.0 (91.0 94.0)
	Summary estimates								**88.4 (86.9, 89.8)**	**92.5 (91.6, 93.4)**
	Knafel 2015[49]	Pre-menopausal	32	1	5	105	143	calculated	96.9 (84.2, 99.9)	95.5 (89.7, 98.5)
	Meys 2016[44]		27	4	4	93	128	calculated	87.0 (69.0, 96.0)	96.0 (89.0, 99.0)
	Piovanono 2016[58]		18	3	6	194	221	calculated	86.0 (71.0, 100)	97.0 (94.0, 99.0)
	Sayasneh 2013b[63]		23	5	5	132	165	calculated	82.0 (64.0, 92.0)	96.0 (91.0, 98.0)
	Testa 2014[$50]		348	24	88	888	1354	calculated	92.0 (86.0, 95.0)	91.0 (87.0, 94.0)
	Summary estimates								**92.4 (89.6, 94.6)**	**92.9 (91.5, 94.1)**
	Knafel 2015[49]	Post-menopausal	46	3	1	33	83	calculated	94.0 (83.1, 98.7)	97.9 (84.7, 99.9)
	Meys 2016[44]		75	9	39	75	198	calculated	89.0 (80.0, 95.0)	85.0 (77.0, 91.0)
	Piovanono 2016[58]		51	12	17	90	170	calculated	81.0 (71.0, 91.0)	84.0 (77.0, 91.0)
	Sayasneh 2013b[63]		41	5	4	40	90	calculated	89.0 (77.0, 95.0)	91.0 (79.0, 96.0)
	Testa 2014[$50]		560	42	76	371	1049	calculated	93.0 (90.0, 95.0)	83.0 (78.0, 87.0)
	Summary estimates								**91.6 (89.5, 93.4)**	**81.6 (78.7, 84.4)**

SA: subjective assessment

[*]personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

[$]data from the IOTA cohort

Table 14: Accuracy of IOTA simple ultrasound rules using and EFSUMB Level 1 examiner

Threshold	Study ID	Subgroup	Index Test variations	TP	FN	FP	TN	Total N	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline											
Malignant (inconclusive were treated as malignant)	Knafel 2015[49]	All	level 1 examiner	79	3	26	118	226	reported	96.3 (89.7, 99.2)	81.9 (74.7, 87.9)
	Knafel 2015[49]	Post-menopausal	level 1 examiner	46	3	12	22	83	calculated	94 (83.1, 98.7)	63.6 (46.5, 80.3)
		Pre-menopausal	level 1 examiner	33	0	14	96	143	calculated	100 (89.4, 100)	87.4 (79.6, 92.9)
Malignant (inconclusives were classified by SA)	Knafel 2015[49]	All	level 1 examiner	79	3	7	137	226	calculated	96.3 (89.7, 99.2)	95.1 (90.2, 98.0)
		Post-menopausal	level 1 examiner	46	3	3	31	83	calculated	93.9 (83.1, 98.7)	90 (76.3, 98.1)
		Pre-menopausal	level 1 examiner	33	0	4	106	143	calculated	100 (89.4, 100)	96.4 (91.0, 99.0)

Table 15: Comparative accuracy of ADNEX, IOTA simple ultrasound rules and RMI 1

Study ID	Subgroup	Index Test	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	RMI threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline																		
Meys 2016[44]	All	ADNEX	≥10%	113	2	80	131	326	98.0 (93.0, 100)	62.0 (55.0, 68.0)	200	82	33	44	167	326	71.0 (62.0, 79.0)	79.0 (72.0, 84.0)
IOTA 2017*,$				[redacted]					[redacted]	[redacted]	200	[redacted]					[redacted]	[redacted]
Summary estimates									**96.0 (94.5, 97.1)**	**67.0 (64.2, 69.6)**	**Summary estimates**						**66.0 (62.9, 69.0)**	**89.0 (87.0, 90.7)**
IOTA 2017*	All	ADNEX	≥10%	[redacted]					[redacted]	[redacted]	250	[redacted]					[redacted]	[redacted]
Meys 2016[44]	All	IOTA simple rules	inconclusive = malignant	107	8	67	144	326	93.0 (86.0, 97.0)	68.0 (61.0, 70.0)	200	82	33	44	167	326	71.0 (62.0, 79.0)	79.0 (72.0, 84.0)
IOTA 2017*,$				[redacted]					[redacted]	[redacted]	200	[redacted]					[redacted]	[redacted]
Summary estimates									**92.8 (90.9, 94.3)**	**71.6 (68.9, 74.1)**	**Summary estimates**						**66.0 (62.9, 69.0)**	**89.0 (87.0, 90.7)**
IOTA 2017*	All	IOTA simple rules	inconclusive = malignant	[redacted]					[redacted]	[redacted]	250	[redacted]					[redacted]	[redacted]
Meys 2016[44]	All	IOTA simple rules	inconclusive = SA	102	13	21	190	326	89.0 (81.0, 94.0)	90.0 (85.0, 94.0)	200	82	33	44	167	326	71.0 (62.0, 79.0)	79.0 (72.0, 84.0)
	Pre-menopausal	ADNEX	≥10%	31	0	28	69	128	100 (86.0, 100)	71.0 (61.0, 80.0)	200	13	18	6	91	128	42.0 (25.0, 61.0)	94.0 (86.0, 97.0)
		IOTA simple rules	inconclusive by SA	27	4	4	93	128	87.0 (69.0, 96.0)	96.0 (89.0, 99.0)								
			inconclusive = malignant	29	2	23	74	128	94.0 (77.0, 99.0)	76.0 (66.0, 84.0)								
	Post-menopausal	ADNEX	≥10%	82	2	52	62	198	98.0 (91.0, 100)	54.0 (44.0, 63.0)	200	69	15	39	75	198	82.0 (72.0, 89.0)	66.0 (56.0, 74.0)
		IOTA simple rules	inconclusive = SA	75	9	39	75	198	89.0 (80.0, 95.0)	85.0 (77.0, 91.0)								
			inconclusive = malignant	78	6	44	70	198	93.0 (85.0, 97.0)	61.0 (52.0, 70.0)								

*personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017; $data from the IOTA cohort

Table 16: Comparative accuracy of IOTA simple ultrasound rules and the RMI 1

Study ID	Subgroup	IOTA simple rules threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	RMI threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline																	
Adballa 2013[48]	All	inconclusive = malignant	16	1	7	63	87	94.1 (71.3, 99.9)	90.0 (80.5, 95.9)	200	15	2	8	62	87	88.2 (63.6, 98.5)	88.6 (78.7, 94.9)
IOTA 2017[*$]			[redacted]					[redacted]	[redacted]	200	[redacted]					[redacted]	[redacted]
Meys 2016[44]			107	8	67	144	326	93.0 (86.0, 97.0)	68.0 (61.0, 70.0)	200	82	33	44	167	326	71.0 (62.0, 79.0)	79.0 (72.0, 84.0)
Sayasneh 2013b[63 C]			67	7	24	157	255	91.0 (82.0, 95.0)	87.0 (82.0, 91.0)	200	53	21	11	170	255	72.0 (60.0, 81.0)	94.0 (90.0, 97.0)
Testa 2014[$50 C]			934	46	369	1054	2403	95.3 (93.1, 96.19)	74.1 (67.7, 79.7)	200	657	323	134	1289	2403	67.1 (61.4, 72.4)	90.6 (87.3, 93.1)
Summary estimates								**93.9 (92.8, 94.9)**	**74.2 (72.6, 75.8)**	**Summary estimates**						**66.9 (64.8, 68.9)**	**90.1 (88.9, 91.2)**
Meys 2016[44]	All	inconclusive = SA	102	13	21	190	326	89.0 (81.0, 94.0)	90.0 (85.0, 94.0)	200	82	33	44	167	326	71.0 (62.0, 79.0)	79.0 (72.0, 84.0)
Sayasneh 2013b[63 C]			64	10	11	170	255	86.0 (77.0, 92.0)	94.0 (90.0, 97.0)	200	53	21	11	170	255	72.0 (60.0, 81.0)	94.0 (90.0, 97.0)
Testa 2014[$50 C]			900	80	157	1266	2403	91.8 (89.1, 93.9)	89.0 (85.2, 92)	200	657	323	134	1289	2403	67.1 (61.4, 72.4)	90.6 (87.3, 93.1)
Summary estimates								**91.2 (89.4, 92.8)**	**89.6 (88.1, 91)**	**Summary estimates**						**67.8 (65.0, 70.4)**	**98.5 (98.3, 98.7)**
Meys 2016[44]	Pre-menopausal	inconclusive = malignant	29	2	23	74	128	94.0 (77.0, 99.0)	76.0 (66.0, 84.0)	200	13	18	6	91	128	42.0 (25.0, 61.0)	94.0 (86.0, 97.0)
Sayasneh 2013b[63 C]			24	4	16	121	165	86.0 (69.0, 94.0)	88.0 (83.0, 93.0)	200	15	13	5	132	165	54.0 (36.0, 70.0)	96.0 (92.0, 98.0)
Testa 2014[$50 C]			359	19	225	751	1354	95.0 (91.0, 97.0)	77.0 (70.0, 83.0)	200	200	178	59	917	1354	53.0 (45.0, 61.0)	94.0 (92.0, 96.0)
		Summary estimates						**94.3 (91.7, 96.3)**	**78.2 (75.7, 80.5)**	**Summary estimates**						**52.2 (47.4, 56.9)**	**94.2 (92.7, 95.5)**
Meys 2016[44]		inconclusive = SA	27	4	4	93	128	87.0 (69.0, 96.0)	96.0 (89.0, 99.0)	200	13	18	6	91	128	42.0 (25.0, 61.0)	94.0 (86.0, 97.0)
Sayasneh 2013b[63 C]			23	5	5	132	165	82.0 (64.0, 92.0)	96.0 (91.0, 98.0)	200	15	13	5	132	165	54.0 (36.0, 70.0)	96.0 (92.0, 98.0)
Testa 2014[$50 C]			348	24	88	888	1354	92.0 (86.0, 95.0)	91.0 (87.0, 94.0)	200	200	178	59	917	1354	53.0 (45.0, 61.0)	94.0 (92.0, 96.0)
		Summary estimates						**92.3 (89.4, 94.7)**	**92 (90.3, 93.5)**	**Summary estimates**						**52.2 (47.4, 56.9)**	**94.2 (92.7, 95.5)**
Meys 2016[44]	Post-menopausal	inconclusive = malignant	78	6	44	70	198	93.0 (85.0, 97.0)	61.0 (52.0, 70.0)	200	69	15	39	75	198	82.0 (72.0, 89.0)	66.0 (56.0, 74.0)
Sayasneh 2013b[63 C]			43	3	7	37	90	93.0 (82.0, 98.0)	84.0 (71.0, 92.0)	200	38	8	5	39	90	83.0 (69.0, 91.0)	89.0 (76.0, 95.0)

Study ID	Subgroup	IOTA simple rules threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	RMI threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Testa 2014[$50 C]			578	24	152	295	1049	96.0 (93.0, 97.0)	66.0 (59.0, 73.0)	200	470	132	85	362	1049	78.0 (72.0, 83.0)	81.0 (76.0, 85.0)
		Summary estimates						**95.5 (93.7, 96.9)**	**72.3 (68.9, 75.5)**	**Summary estimates**						**78.8 (75.7, 81.7)**	**78.7 (75.2, 81.9)**
Meys 2016[44]		inconclusive = SA	75	9	39	75	198	89.0 (80.0, 95.0)	85.0 (77.0, 91.0)	200	69	15	39	75	198	82.0 (72.0, 89.0)	66.0 (56.0, 74.0)
Sayasneh 2013b[63 C]			41	5	4	40	90	89.0 (77.0, 95.0)	91.0 (79.0, 96.0)	200	38	8	5	39	90	83.0 (69.0, 91.0)	89.0 (76.0, 95.0)
Testa 2014[$50 C]			560	42	76	371	1049	93.0 (90.0, 95.0)	83.0 (78.0, 87.0)	200	470	132	85	362	1049	78.0 (72.0, 83.0)	81.0 (76.0, 85.0)
		Summary estimates						**92.3 (90.2, 94.2)**	**80.3 (76.9, 83.4)**	**Summary estimates**						**78.8 (75.7, 81.7)**	**78.7 (75.2, 81.9)**
Di Legge 2012[$62]	Tumour size <4 cm	inconclusive = malignant	42	9	13	332	396	82.0 (69.0, 92.0)	96.0 (94.0, 98.0)	200	29	22	16	329	396	56.0 (43.0, 70.0)	95.0 (93.0, 98.0)
	Tumour size ≥10 cm		281	23	66	222	592	92.0 (89.0, 95.0)	77.0 (72.0, 82.0)	200	224	80	38	250	592	74.0 (69.0, 79.0)	87.0 (83.0, 91.0)
	Tumour size 4 to 9.9 cm		303	27	60	1067	1457	92.0 (88.0, 95.0)	95.0 (93.0, 96.0)	200	220	110	68	1059	1457	67.0 (62.0, 72.0)	94.0 (92.0, 95.0)
Target condition: Ovarian borderline tumours – higher stage malignancies excluded																	
Testa 2014[$50 C]	All	inconclusive = malignant	133	20	367	1056	1576	87.5 (79.3, 92.8)	74.2 (66.5, 80.7)	200	45	108	134	1289	1576	29.6 (21.2, 39.7)	90.6 (87.1, 93.2)
		inconclusive = SA	121	32	152	1271	1576	79.5 (70.8, 86.1)	89.3 (84.7, 92.7)	200	45	108	134	1289	1576	29.6 (21.2, 39.7)	90.6 (87.1, 93.2)

[*]personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

[$]data from the IOTA cohort

[C] = 2x2 data were calculated, studies not denoted had reported 2x2 data

MP = menopausal; SA = subjective assessment

3.2.5 Diagnostic performance of Overa (MIA2G)

Details of Overa (MIA2G) studies

Three diagnostic cohort studies reported in four publications[69-71, 105] provided data on the diagnostic performance of the Overa (MIA2G) score, for the identification of women with an adnexal mass, who are at high risk of ovarian cancer. All were conducted in the USA. Only one study[71] was reported as a full paper; the remaining two studies were reported in the form of meeting slides[105] and a conference abstract,[69] respectively.

One study used an Overa (MIA2G) score based on Roche assays and a Roche analyser[71], the other two studies did not report assay details.[69, 105]

The target condition for this assessment is ovarian cancer (epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma, and borderline ovarian cancer). All studies in this section included people with one or more adnexal mass, and used a definition of malignancy which included borderline cancers. Histopathology indicated that all of the studies also included some patients 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 2 (Table 36).

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

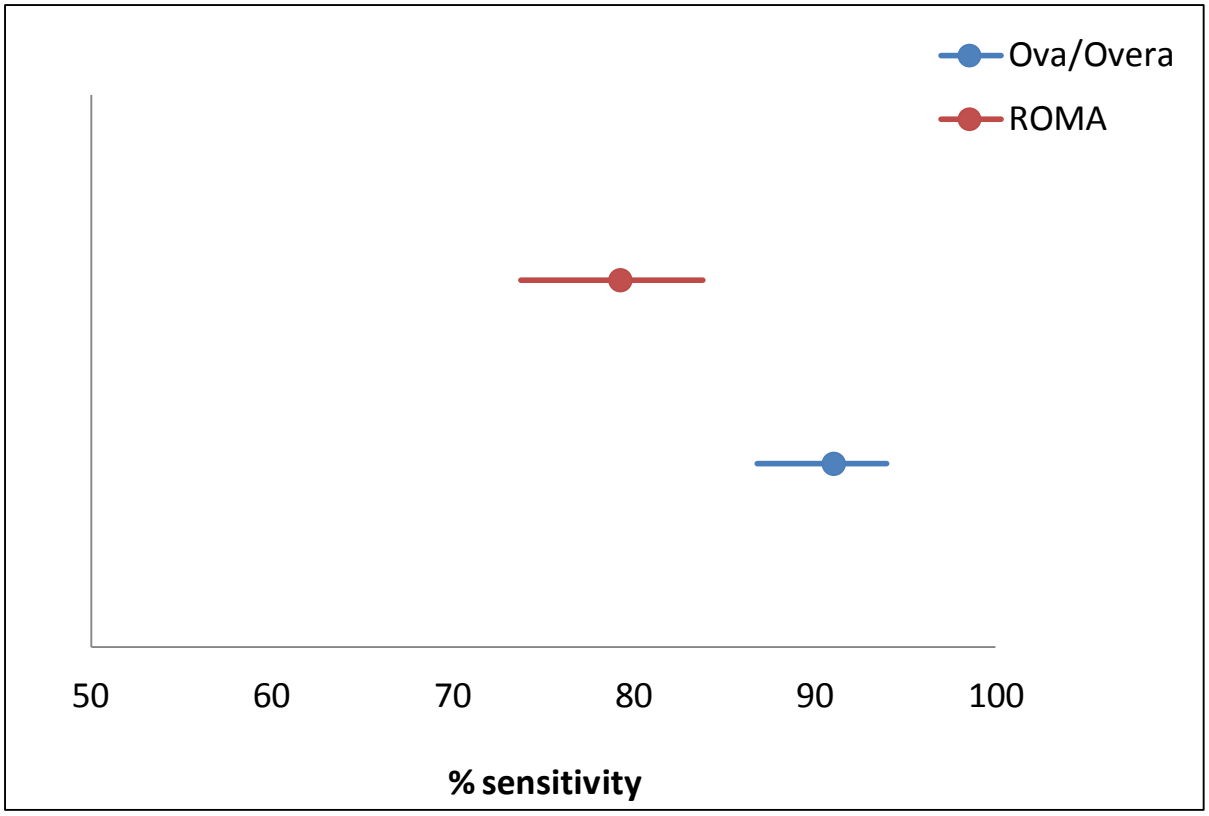
No studies were identified that directly compared Overa (MIA2G) to RMI 1 at either decision threshold (200 or 250).

One study[105] reported comparative accuracy data for Overa (MIA2G) versus the ROMA score, using Roche Elecsys tumour marker assays (see Table 17). This study included all participants in the analysis, regardless of their final histopathological diagnosis (i.e. target condition all malignancies including borderline). At a threshold of five units, the sensitivity estimate for Overa (MIA2G) was, 91% (95% CI: 86.8, 94) and the specificity estimate was, 65.5% (95% CI: 62.0 to 68.8%). In comparison the sensitivity estimate for ROMA (threshold 11.4%/29.9%) was, 79.2% (95% CI: 73.7 to 83.8%) and the specificity estimate was, 78.9% (95% CI: 75.8, 81.7). These data indicate that the sensitivity of the Overa (MIA2G) score was significantly higher than the ROMA score, whilst the specificity of the Overa (MIA2G) score was significantly lower than the ROMA score (see Figure 5).

The two remaining studies reported data on the accuracy of Overa (MIA2G) without comparison to RMI 1 or any other risk score (see Table 18).[69, 71] analysed data for any malignant tumour plus borderline (see Table 18). At a threshold of five 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 18).[71]

Figure 5: Comparison of the summary estimates for Overa (MIA2G) and ROMA Roche (all malignant tumours plus borderline)



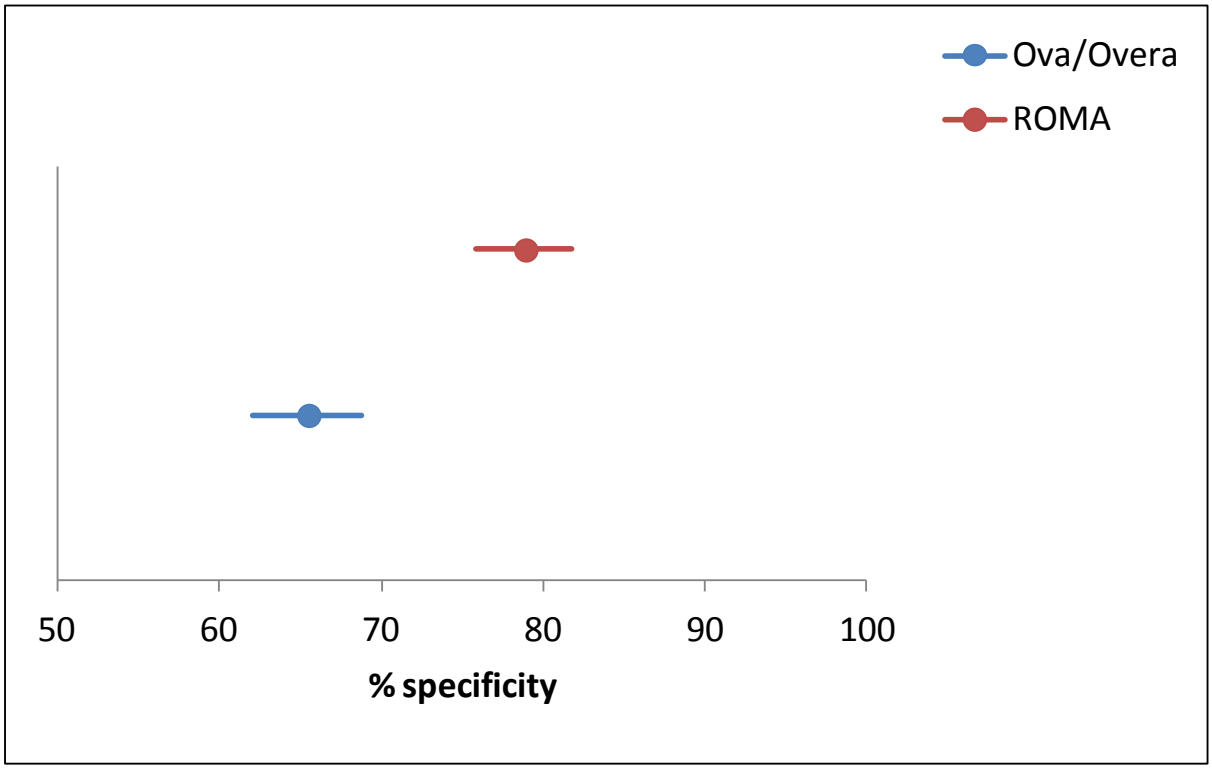


Table 17: Comparative accuracy of Overa (MIA2G) versus ROMA score

study ID	Index Test	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline									
Shulman 2016[105]	Overa (MIA2G)	5 Units	223	22	258	490	993	91.0 (86.8, 94.0)	65.5 (62.0, 68.8)
	ROMA Roche	11.4%/29.9%	194	51	158	590	993	79.2 (73.7, 83.8)	78.9 (75.8, 81.7)

Table 18: Accuracy of the Overa (MIA2G) score at a threshold of five units

study ID	Subgroup	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline									
Coleman 2016[71]	All	84	8	124	277	493	reported	91.3 (83.8, 95.5)	69.1 (64.4, 73.4)
Zhang 2015[69]	All	64	8	93	140	305	reported	88.9 (79.3, 95.1)	60.1 (53.5, 66.4)
Summary estimates								**90.2 (84.6, 94.3)**	**65.8 (61.9, 69.5)**
Coleman 2016[71]	Pre-menopausal	28	3	70	175	276	reported	90.3 (75.1, 96.7)	71.4 (65.5, 76.7)
	Post-menopausal	56	5	54	102	217	reported	91.8 (82.2, 96.4)	65.4 (57.6, 72.4)

3.2.6 Diagnostic performance of the RMI using decision thresholds other than 250

Details of RMI studies

Ten diagnostic cohort studies,[72-81] 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 are at high risk of ovarian cancer.

Two studies specifically included patients from the UK[79, 80], two were European (Italy, Norway)[77, 81] and six studies were from five non-European countries (Turkey, Pakistan, China, India, Japan).[72-76, 78]

Three studies used a RMI score based on an Abbott CA125 assay,[76, 77, 79] three used a Roche [72, 73, 75], one used an Immulite assay[78], one used CIS bioindustries[80], one used Centocor[81] and one did not report the CA125 assay used.[74]

This assessment is primarily concerned with providing a comparison between the RMI 1,[79] used with a decision threshold of 250 (current standard practice in the UK NHS[1]) and the specified alternative risk scoring methods (see section 2.3). The identified studies for the RMI 1 reported test performance data for multiple thresholds and full data are reported in Appendix 4, Table 40. All of the identified studies which provide comparative accuracy data for alternative risk scoring methods versus 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 RMI 1, using decision thresholds of 200 and 250.

The target condition for this assessment is ovarian cancer (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 people with one or more adnexal mass. Seven studies used a definition of malignancy which included borderline tumours,[73-75, 77, 79-81] two studies excluded patients found to have borderline tumours from the analyses,[72, 76] and, in the remaining study,[78] it was unclear whether patients with borderline tumours were included in the analysis (no histopathology was reported with which to confirm the tumour type). Six studies included all study participants in the analyses, and included some patients with 'other malignancies', metastases from non-ovarian sites, 'non-ovarian cancers.'[74, 75, 77, 79-81] Full details of the final histopathological diagnoses of study participants who had a malignant mass are reported in Appendix 2 (Table 26).

Accuracy of RMI for determining high risk of ovarian cancer using different decision thresholds

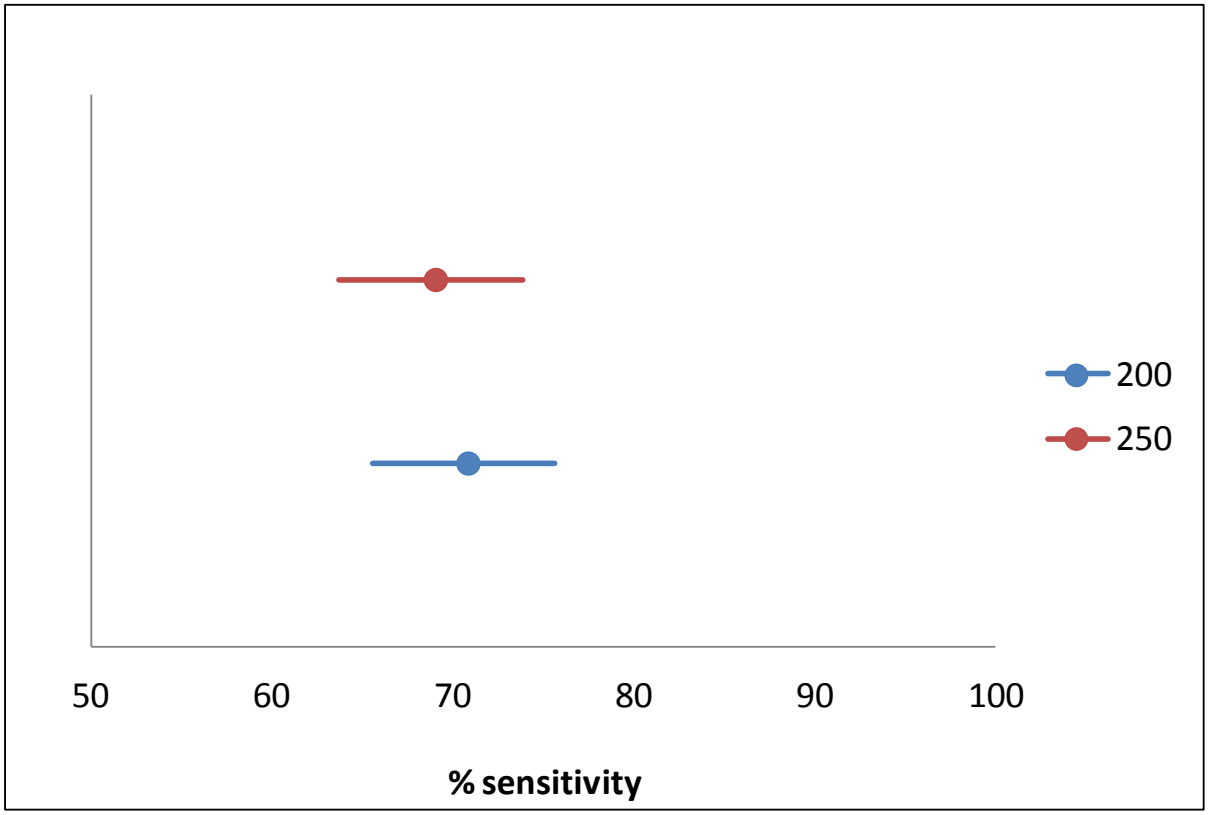
Six studies included all study participants in the analyses, regardless of final histopathological diagnosis (target condition all malignant tumours including borderline), (see Table 19).[74, 75, 77, 79-81] At the decision threshold of 200, the summary estimate of sensitivity estimate derived from these studies was 70.8% (95% CI: 65.6 to 75.6%) and the summary estimate of specificity 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) (see Figure 6). Studies compared multiple thresholds (between 25 and 500); as would be expected, the sensitivity estimate for the RMI 1 increased and specificity decreased with decreasing threshold (see Appendix 4, Table 40).

One study[73] reported a direct comparison of the RMI 1 at decision thresholds of 250 and 200 and excluded patients with a final histopathological diagnosis other than primary ovarian cancer from the analysis (i.e. target condition ovarian malignancies including borderline) (see Table 19). 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 above.

Two further studies[72, 76] excluded participants found to have borderline tumours from the analysis (i.e. 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 did not differ significantly from those derived from the six studies which included all participants in their analyses.

One study[78] included participants with malignant tumours but it was unclear whether borderline tumours were included or not (Appendix 4, Table 43) .

Figure 6: Comparison of the summary estimates for RMI at 200 and 250 thresholds (all malignant tumours plus borderline)



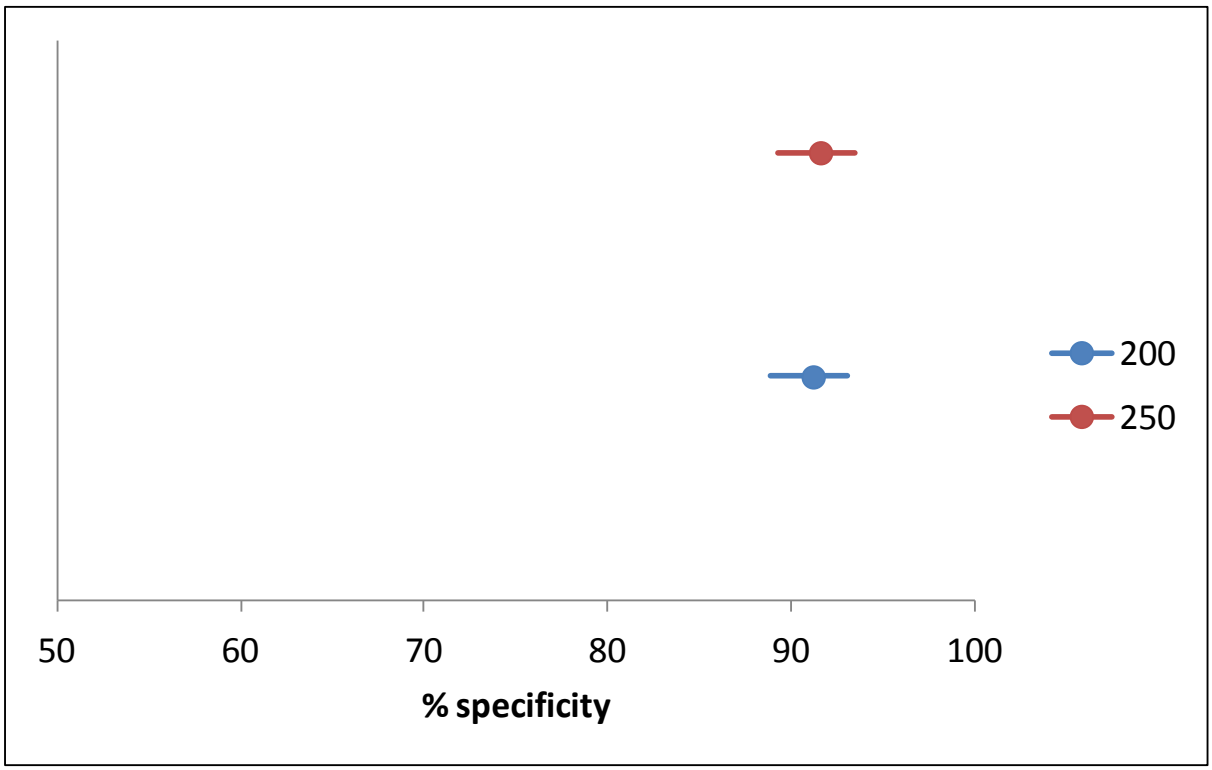


Table 19: Comparative accuracy of the RMI 1 at decision thresholds of ≥200 and ≥250

Threshold		**200**						**250**					
study ID	**Total**	**TP**	**FN**	**FP**	**TN**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**TP**	**FN**	**FP**	**TN**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
Target Condition: All malignant tumours including borderline													
Davies 1993[80]	124	33	4	11	76	89.2 (74.6, 97.0)	87.4 (78.5, 93.5)	34	3	21	66	91.9 (78.1, 98.3)	75.9 (65.5, 84.4)
Jacobs 1990[79]	139	35	6	3	95	85.4 (70.8, 94.4)	96.9 (91.3, 99.4)	32	9	1	97	78.0 (62.4, 89.4)	99.0 (94.5, 100)
Lou 2010[74]	223	34	27	5	157	55.7 (42.4, 68.5)	96.9 (92.9, 99.0)	35	26	3	159	57.4 (44.1, 70.0)	98.1 (94.7. 99.6)
Morgante 1999[81]	124	18	13	5	88	58.1 (39.1, 75.5)	94.6 (87.9, 98.2)	17	14	4	89	54.8 (36.0, 72.7)	95.7 (89.4, 98.8)
Tingulstad 1996[77]	173	40	16	5	112	71.4 (57.8, 82.7)	95.7 (90.3, 98.6)	38	18	5	112	67.9 (54.0, 79.7)	95.7 (90.3, 98.6)
Ulusoy 2007[75]	296	75	31	37	153	71.1 (62.1, 80)	80.5 (74.2, 85.9)	73	33	29	161	68.9 (59.1, 77.5)	84.7 (78.8, 89.5)
Summary estimates						**70.8 (65.6, 75.6)**	**91.2 (88.9, 93.1)**					**69.0 (63.7, 73.9)**	**91.6 (89.3, 93.5)**
Target Condition: Ovarian malignancies including borderline													
Yamamoto 2009[73]	253	32	8	29	184	80.0 (65.2, 89.5)*	86.4 (81.8, 89.9)*	29	11	24	189	72.5 (57.2, 83.9)*	88.7 (84.4, 92)*
All malignant tumours excluding borderline													
Aktürk 2011[72]	100	15	5	9	71	75.0 (50.9, 91.3)	88.8 (79.7, 94.7)	13	7	4	76	65.0 (40.8, 84.6)	95.0 (87.7, 98.6)
Manjunath 2001[76]	148	68	25	5	50	73.1 (62.9, 81.8)	90.9 (80.0, 97.0)	62	31	5	50	66.7 (56.1, 76.1)	90.9 (80.0, 97.0)
Summary estimates						**73.5 (64.3, 81.3)**	**89.6 (83.2, 94.2)**					**66.4 (56.9, 75.0)**	**93.3 (87.7, 96.9)**

*Calculated values

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

We prioritised data for 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, 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 which fall outside the scope of conditions covered in NICE CG122[1] (e.g. germ cell tumours and sex cord stromal tumours of the ovary); we therefore consider that studies which include all participants in their analysis, irrespective of final histological diagnosis, are more likely to produce estimates of risk score performance which are representative of what might be expected in clinical practice.

Comparative accuracy data were available for risk scores ROMA, IOTA simple ultrasound 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. We did not identify any studies that provided a direct comparison of Overa (MIA2G) versus 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 which reported a direct comparison with RMI 1. Cost effectiveness modelling therefore used the summary estimates of diagnostic performance 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 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, pre-menopausal women or post-menopausal women). Where no data were available for the RMI 1 with a decision threshold of 250, we used data for a decision threshold of 200; the analysis reported in section 3.2.6 indicated no significant difference in the performance of RMI 1 at these two thresholds.

4. ASSESSMENT OF COST EFFECTIVENESS

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

What is the cost effectiveness of alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI 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?

4.1 Review of economic analyses of ovarian cancer risk scores

4.1.1 Search strategy

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

Methodological study design filters were included in the search strategy where 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 CADTH 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 (Ovid): 1946 to November Week 2, 2016
- MEDLINE In-Process Citations (Ovid): to 22 November 2016
- MEDLINE Daily Update (Ovid): to 22 November 2016
- MEDLINE Epub Ahead of Print (Ovid): to 23 November 2016
- EMBASE (Ovid): 1974 to 22 November 2016
- NHS Economic Evaluation Database (NHS EED) (Wiley): to Issue 2 of 4, April 2015
- EconLit (EBSCO): 1966 to 25 November 2016
- Cost-Effectiveness Analysis Registry (Internet) http://www.cearegistry.org: to 25 November 2016
- Research Papers in Economics (RePEc) (Internet) http://repec.org/: to 25 November 2016

Full search strategies are presented in Appendix 1.

4.1.2 Inclusion criteria

Studies reporting outcomes of a full cost effectiveness analysis, examining (quality-adjusted) life-years, 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.

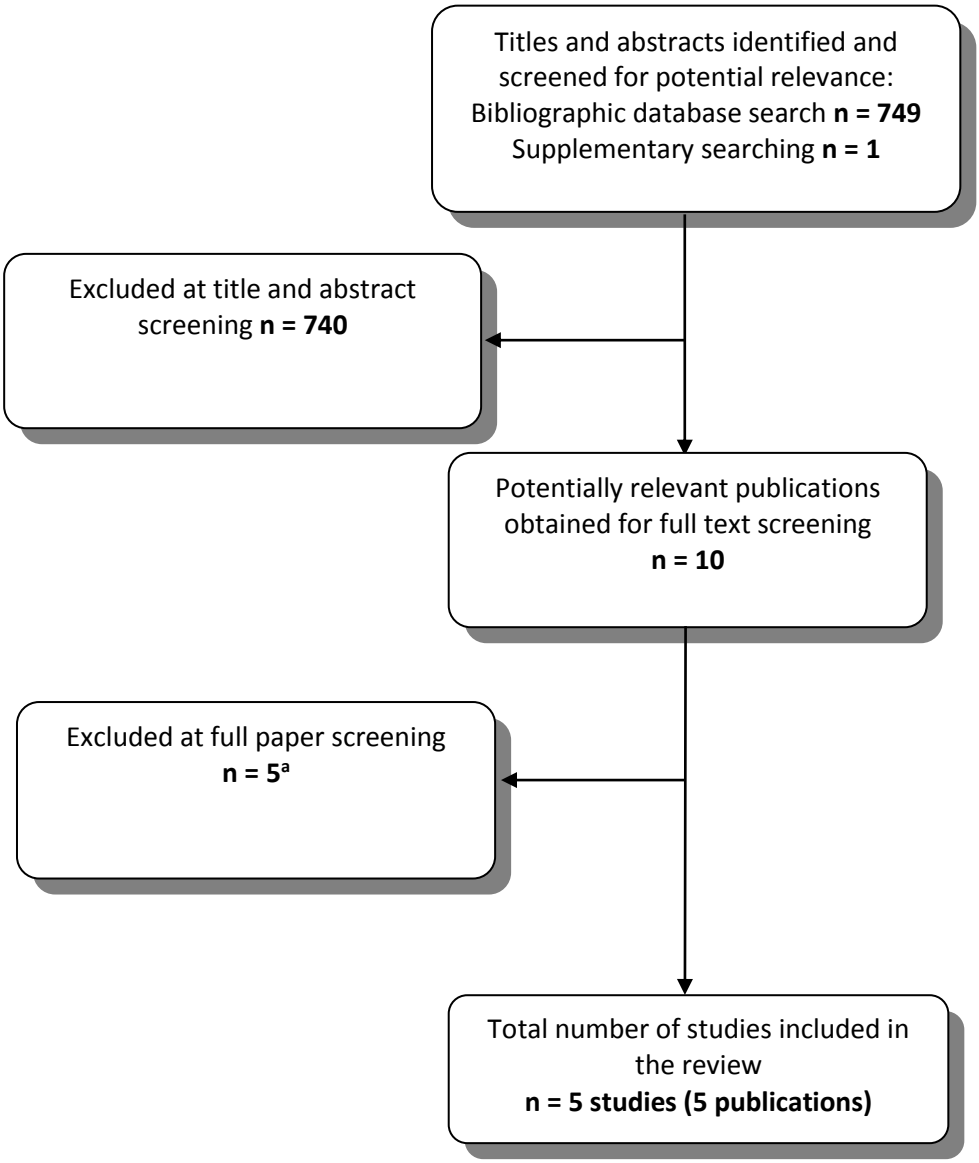
4.1.3 Quality assessment

Included studies are appraised using a quality checklist based on Drummond et al. [122]

4.1.4 Results

The literature search identified 749 records from bibliographic database searches and supplementary searching (e.g. reference/citation checking, additional database searches including the database search for the assessment of clinical effectiveness). After title and abstract screening, 10 records are considered to be potentially relevant; after full text screening five studies (five publications, including one abstract) were considered eligible for inclusion (see Figure 7).[123-127] These studies are described in more detail below and summarised in Table 20. The results of the quality assessment are shown in Table 21.

Figure 7: Flowchart (review of economic analyses)



[a] Reasons for exclusion: did not report outcomes of a full cost effectiveness analysis (n=2) and did not report QALY or LYs as outcome (n=3).

Havrilesky 2015

Havrilesky and colleagues[125] constructed a Markov model (in TreeAge Pro 2013) using alive and death health states. From a societal US perspective, the authors 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:

- American Congress of Obstetricians and Gynaecologists guidelines (ACOG),
- Multivariate Index Assay algorithm (MIA; Vermillion, Austin, TX),
- Risk of Malignancy Algorithm (ROMA),
- CA125 alone with lowered cut-off values to prioritize test sensitivity over specificity (15 U/mL for postmenopausal and 22 U/mL for premenopausal patients)
- referral of all women

The analyses indicated that CA125 is cost effective for willingness to pay thresholds below $9,423 and $10,644 per life year (LY) gained for post- 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. due to capacity constraints), CA125 with lowered cut-off values should be considered.

Drescher 2012

Drescher and colleagues[126] used an unspecified model type and structure to estimate the cost effectiveness of first-line testing of women with an adnexal mass, transvaginal ultrasound (TVS) and CA125, in women aged between 45 and 85 years (US setting, perspective not stated). The following multimodal testing strategies were considered:

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

The analysis indicated that CA125 and TVS led to 1.68 more LYs than no primary care testing, resulting in an ICER of $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 and has similar costs is required for primary care testing to reduce ovarian mortality by at least 25% and be reasonably cost-effective.

Kearns 2016

Using the NHS Personal Social Services perspective, Kearns and colleagues[127] developed a Markov model to estimate the cost-effectiveness of different screening strategies in post-menopausal women and to estimate the value of further research. The following screening strategies were considered:

- multimodal screening (MMS): first line screening with CA125 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 £8,864 per QALY gained, but alternative mortality extrapolation methods increased the ICER. The conclusion was that multimodal screening 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 2016

From the perspective of the public payer, Forde and colleagues[124] 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 multivariate index assay (MIA; Ova1, Vermillion, Austin, TX, USA) based on five biomarkers including CA125
- the modified American College for Obstetricians and Gynaecologists (mACOG) referral guidelines
- CA125 testing alone

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

Ding 2010

Ding and colleagues[123] assess the cost effectiveness of annual multimodal screening (with CA125 marker, followed by transvaginal ultrasound for those at increased risk according to CA125 level) versus no screening for postmenopausal females age 65 to 69 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, multimodal screening was both more costly (incremental costs of $820) and more effective (incremental QALY of 0.004) resulting in an ICER of $221,622 per QALY gained compared to no screening.

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

In total, three out of the five included studies reported QALYs as outcome. Of these studies[123, 124, 127] two considered population screening[123, 127] while the remaining study considered the assessment of women referred to secondary care[124] from the US perspective. The latter study was of reasonable quality (see Table 21). The UK screening study[127] indicated that multimodal triage consisting of CA125 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 indicating that MIA might be cost effective,[124] whereas the other indicated that it was dominated by other strategies (when considering LYs).[125] This latter study was the only one to consider the ROMA score and also indicated that this strategy would be dominated by other strategies (when considering LYs).[125] Moreover, this study indicated that a "refer all" strategy is cost effective for thresholds above $10,644 per LY gained.[125] In conclusion, there is limited and conflicting evidence regarding the cost effectiveness of alternative risk scores, which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥250 (current UK practice) for people 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.

Table 20: Summary of included economic evaluations (all abstracts)

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
Population	Women with an adnexal mass	Women aged between 45 and 85 years	Postmenopausal women aged 50–74 years in the UK	Women with adnexal masses	Postmenopausal females age 65 to 69
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 between 45 and 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 multivariate index assay for use in triaging women with an adnexal mass.	To assess cost effectiveness of annual multimodal screening versus no screening for postmenopausal females age 65–69
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) American Congress of Obstetricians and Gynaecologists guidelines (ACOG), 2) Multivariate Index Assay algorithm (MIA), 3) Risk of Malignancy Algorithm (ROMA), 4) CA125 alone with	1) No screening 2) CA125 and TVS 3) CA125 and hypothetical imaging 4) Hypothetical biomarker and TVS 5) Hypothetical biomarker and hypothetical imaging	1) Ultrasound screening (transvaginal ultrasound by sonographer in first line and by more experienced member of staff in second line). 2) multimodal screening (CA125 interpreted using ROCA in first line and	1) Multivariate index assay (MIA): based on 5 biomarkers, transthyretin, apolipoprotein, A-1, 2-microglobulin, transferrin & CA125 2) Modified American College of Obstetricians	1) Annual multimodal screening (with CA125 marker, followed by transvaginal ultrasound for those at increased risk according to CA125 level) 2) No screening

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
	lowered cut-off values to prioritize test sensitivity over specificity, 5) referral of all women		ultrasound by more experienced member of staff in second line)	& Gynaecologists (mACOG) referral guidelines (including patient's history, physical, pelvic ultrasound and CA125) 3) CA125 alone	
Costs items	-Test costs -Surgery costs -Subspecialist costs -End-of-life costs	-Test costs -Surgery costs -Treatment costs -End-of-life costs	Multimodal and ultrasound screening drop-outs and complete screening, screening invitation, diagnosis and treatment of borderline or stages 1-4 ovarian cancer, end of life cost	Chemotherapy with different cycle lengths and for colorectal cancer; DRG costs and professional fees for surgery for malignancy, non-malignancy, staging surgery; CT scan; CA125; mACOG; MIA	Not stated
Main measure of benefit	LY	LY	QALY	QALY	QALYs
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 specialty of the original surgeon). -The "average" postsurgical treatment, including chemotherapy, is similar for women with	-Women alive 15 years after 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	Lognormal for modelling survival in screening arms and Weibull in no screening arm; disutility associated with diagnosis relates to treatment and lasts only for one year; no disutility associated with screen, use of ROCA does not increase costs.	Major treatment related costs occur during first year of treatment; QoL utility weights change with disease progression and differ by stage and type of cancer	Not stated

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
	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 false-negative 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 cases. - 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	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) is detected in early stage (I or II) by screening.			

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
	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)				
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 and PSA	Yes, one-way	Yes
Monetary outcomes	2013 US$	2010 US$	£	2014 US$	2009 US$
Outcomes per comparator	**Postmenopausal** (costs, LYs) CA125: $17,428; 16.93 ACOG: $17,469; 16.92 ROMA: $17,485; 16.91 Refer all: $17,510; 16.94 MIA: $18,004; 16.92 **Premenopausal**	No screening: $865 CA125 and TVS: $1,741 Absolute LYs are not provided (see results section of the paper for the results of hypothetical	MMS vs USS vs no screening QALYs: 14.357 vs 14.297 vs 14.29 Costs: 598 vs 824 vs 179	MIA vs mACOG (only direct costs; direct and indirect costs): 35,094; dominating MIA vs CA125 (only direct costs; direct and indirect costs): 12,189; dominating	Not reported

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
	(costs, LYs) CA125: $9,876; 28.58 ACOG: $9,892; 28.55 ROMA: $9,897; 28.57 Refer all: $9,999; 28.60 MIA: $10,354; 28.58	strategies)			
Summary of incremental analysis	CA125 is cost effective for willingness to pay thresholds below $9,423 and $10,644 per LY gained for post- and premenopausal women respectively. Refer all is cost effective above these thresholds. The other strategies are dominated.	CA125 and TVS led to 1.68 more LY than no screening, resulting in an ICER of $88,993 per LY gained. (see results section of the paper for the results of hypothetical strategies)	MMS and USS likely to be associated with benefits for patients but also with additional costs. The ICER of MMS vs no screening was 8,864 and USS was dominated my MMS	MIA and referral to gynaecologic oncologist (instead of surgery by gynaecologist) for all patients are the most cost effective triage strategies for women with adnexal masses.	Multimodal screening resulted in additional costs and QALYs of $820 and 0.0037 respectively versus no screening. This resulted in an ICER of $226,622 per QALY gained.

Abbreviations: NR, not reported; NA, not applicable; QALY, quality adjusted life-years; LY, Life years; vs, versus; PSA, probabilistic sensitivity analysis; ACOG, American Congress of Obstetricians and Gynaecologists guidelines; MIA, Multivariate Index Assay algorithm; ROMA, Risk of Malignancy Algorithm; TVS, transvaginal sonography; ICER, incremental cost effectiveness ratio;

Table 21: Study quality checklist for included studies

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
Study design					
The research question is stated	√	√	√	√	√
The economic importance of the research question is stated	√	√	√	√	X
The viewpoint(s) of the analysis are clearly stated and justified	√	X	√	√	√
The rationale for choosing alternative programmes or interventions compared is stated	√	√	√	√	X
The alternatives being compared are clearly described	√	√	√	√	X
The form of economic evaluation used is stated	√	√	√	√	√
The choice of form of economic evaluation is justified in relation to the questions addressed	√	X	√	√	√
Data collection					
The source(s) of effectiveness estimates used are stated	√	√	√	√	
Details of the design and results of effectiveness study are given (if based on a single study)	X	X	√	X	X
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	X
The primary outcome measure(s) for the economic evaluation are clearly stated	√	√	√	√	√
Methods to value benefits are stated	NA	NA	√	X	X
Details of the subjects from whom valuations were obtained were given	NA	NA	X	X	X
Productivity changes (if included) are reported separately	NA	NA	NA	√	X
The relevance of productivity changes to the study question is discussed	NA	NA	NA	√	X
Quantities of resource use are reported separately from their unit costs	X	X	√	X	X
Methods for the estimation of quantities and unit costs are described	X	X	√	X	X

	Havrilesky 2015[125]	Drescher 2012[126]	Kearns 2016[127]	Forde 2016[124]	Ding 2010[123]
Currency and price data are recorded	√	√	√	√	√
Details of currency of price adjustments for inflation or currency conversion are given	√	√	√	√	X
Details of any model used are given	√	X	√	√	X
The choice of model used and the key parameters on which it is based are justified	X	X	√	√	X
Analysis and interpretation of results					
Time horizon of costs and benefits is stated	X	X	√	√	√
The discount rate(s) is stated	√	√	√	√	√
The choice of discount rate(s) is justified	X	X	√	√	X
An explanation is given if costs and benefits are not discounted	NA	NA	NA	NA	NA
Details of statistical tests and confidence intervals are given for stochastic data	√	X	√	X	X
The approach to sensitivity analysis is given	√	√	√	√	√
The choice of variables for sensitivity analysis is justified	X	√	√	√	X
The ranges over which the variables are varied are justified	X	√	√	√	X
Relevant alternatives are compared	√	√	√	√	√
Incremental analysis is reported	√	√	√	√	√
Major outcomes are presented in a disaggregated as well as aggregated form	NA	NA	√	√	X
The answer to the study question is given	√	√	√	√	√
Conclusions follow from the data reported	√	√	√	√	√
Conclusions are accompanied by the appropriate caveats	√	X	√	√	√

Abbreviations: NA, not applicable

Symbols: X, No; √, yes

4.2 Model structure and methodology

4.2.1 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:

- RMI 1 (threshold 250)
- ROMA Abbott ARCHITECT
- ROMA Roche Elecsys
- Overa (MIA2G) Vermillion (threshold five units)
- IOTA Simple Rules (inconclusive assumed to be malignant)
- IOTA ADNEX model (threshold most commonly used in studies, 10%)
- RMI 1 (threshold 200)

An optimised risk assessment that reduces the number of people with ovarian cancer who are not referred for further specialist care (i.e. those with a 'false-negative' 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 patients 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 at a decision thresholds of ≥250) has been reported as having poor sensitivity (69%) for the prediction of malignancy (see Table 19). If referral decisions are based on the RMI 1 at this threshold, there remains the potential for significant numbers of people with ovarian cancer to remain un-referred and to 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) 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, the Overa (MIA2G) (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 section 3.2.3). Alternative threshold values for ADNEX were not considered since the 10% threshold is the most commonly studied and has been used in model validation studies.[42, 46]

For the IOTA simple ultrasound rules risk score, it is 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 specialist ultrasonographer needed). Concerning the alternative decision thresholds for RMI, a threshold of ≥200 (used in the original publication[79]) was used in the base-case analysis and other RMI thresholds were considered in scenario analyses.

4.2.2 Model structure

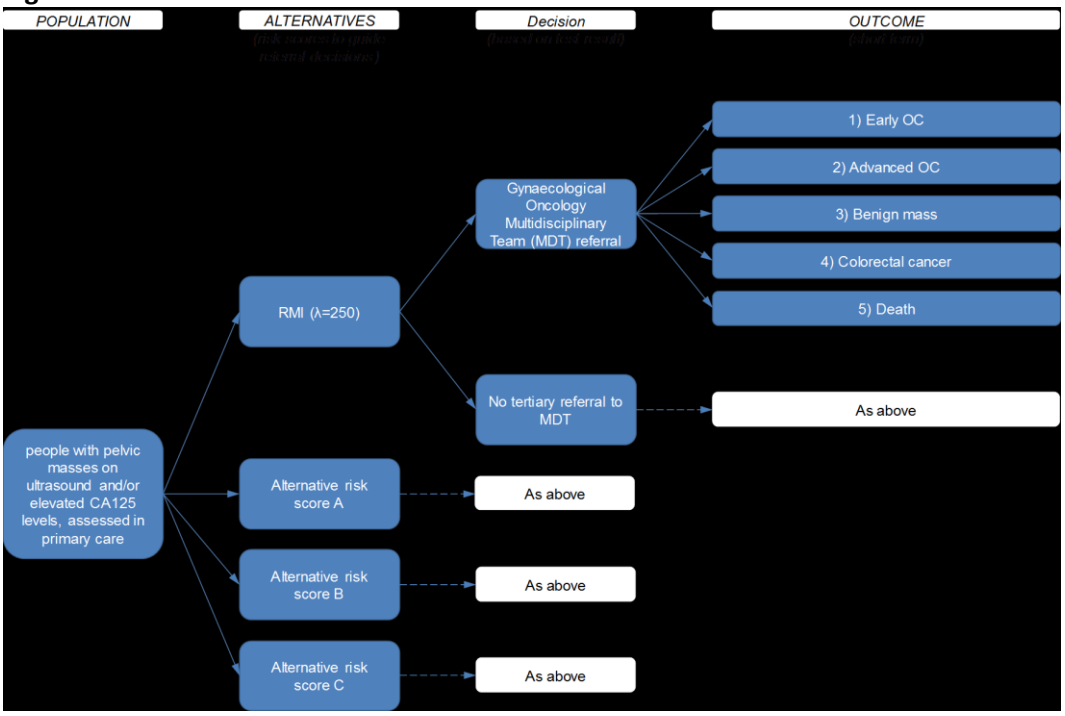
This assessment uses the economic model from CG122 as a starting point. CG122 reviewed clinical and economic questions that involve the detection in primary care, diagnosis in secondary care and initial management of early and advanced stage ovarian cancer.[1] The CG122 model consisted of a decision tree outlining the assessment strategies and a Markov process to model the progression and survival of patients 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 model was constructed using TreeAge Pro (2009) software. The assessment group used the description of this model as a starting point to develop a de novo model (in Microsoft Excel) adapted to better fit the scope of the current assessment. Consistent with the CG122 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) and postmenopausal (mean age of 68) patient groups. The mean age for both groups was estimated based on information on the distribution of ovarian cancer patients pre- and post-age 50 from Cancer Research UK, assuming that menopause occurs approximately at the age of 50.[128]

In the de novo health economic model the mean expected costs and quality adjusted life years (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 an SMDT, or no tertiary care referral. It was also taken into account that a small proportion of the patients with pelvic masses and having tested positive based on the risk score are ultimately diagnosed with colorectal cancer (consistent with CG 122). These patients were therefore included in the model and prognosis for colorectal cancer patients was included in the Markov model.

We developed a decision tree and a Markov model. The decision tree was used to model the short-term (up to 30 days after surgery) outcomes. It was assumed that patients who receive a high-risk test result (either true or false) are referred to SMDT, and patients who receive a low risk test result (either true or false) are not referred to SMDT. After the risk assessment and referral decision, patients in the decision tree are allocated to ‘early OC’, ‘advanced OC’, ‘benign mass’, ‘colorectal

cancer' and 'death'. Death was included as an outcome in the decision tree to account for 30 days' post-surgery mortality. Patients referred to SMDT receive surgery by gynaecological oncology specialists that has shown to achieve better patient outcomes compared with patients not referred to SMDT and, for a proportion of patients, this surgery is extensive. Patients not referred to SMDT receive surgery by secondary care gynaecologists. In the case of a false negative diagnosis, i.e. where a patient has a malignancy and is incorrectly classified as having a low risk score, there is an increased risk of progressing to advanced ovarian cancer (AOC) and / or death. This increased risk is likely due to a combination of factors, such as a delay in appropriate treatment because the patient would be operated on and then referred to SMDT for another surgery (based on clinical experts' feedback). Patients with a benign mass incorrectly classified as at high risk and referred to SMDT receive surgery and have their benign mass removed. This incorrect referral has only cost implications, as patients are identified as having a benign mass at surgery. Alternatives in the patients' care pathway are explored in scenario analyses. The decision tree is shown in Figure 8.

Figure 8: Decision tree structure



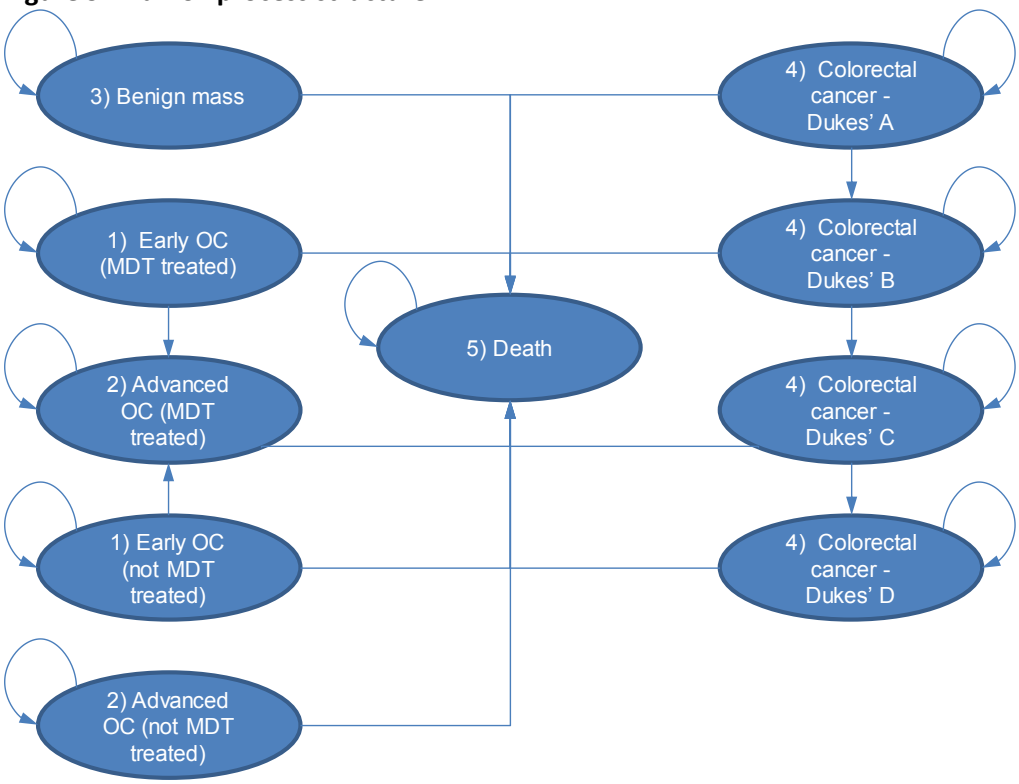
The long-term consequences in terms of costs and QALYs were estimated using a Markov cohort model (see Figure 9) with a lifetime time horizon. The cycle time was one year. The following health states were included:

- Benign mass
- Early ovarian cancer, not referred to tertiary care SMDT
- Early ovarian cancer, referred to tertiary care SMDT

- Advanced ovarian cancer, not referred to tertiary care SMDT
- Advanced ovarian cancer, referred to tertiary care SMDT
- Colorectal cancer Dukes' A
- Colorectal cancer Dukes' B
- Colorectal cancer Dukes' C
- Colorectal cancer Dukes' D
- Death

A distinction between SMDT referral or not was only made for 'early OC' and 'advanced OC'. This was done as it was assumed that a referral to the SMDT would only have an impact on the long-term outcomes in terms of life years and QALYs for patients with ovarian cancer.

Figure 9: Markov process structure



4.2.3 Model parameters

Estimates for the model input parameters were retrieved from the literature and by consulting experts for unpublished data. For consistency and where the same parameters were required, the same sources as used in CG122 were used.[1] Accuracy estimates were derived from the systematic review component of this assessment (see section 3.2). 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 22. 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 proportion of patients diagnosed with other malignancies (assumed to be colorectal cancer) 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 early vs. advanced stage ovarian cancer
- 30 day post-surgery ovarian cancer mortality
- 10 year OS and PFS of ovarian cancer (using updates of the same trials: the ICON1 study was used to model PFS and OS for early ovarian cancer.[129] The ICON3 study was used to model these outcomes for advanced ovarian cancer.[130]

The following parameters were estimated as in the most recent diagnostic appraisal review (DAR) in CRC:[131]

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

The effect of the SMDT (i.e. with gynaecologic oncologists on site) versus patients treated in secondary care was estimated from Woo et al.,[132] who report a HR (0.90; 95%CI: 0.820-0.990) for OS of patients with ovarian cancer treated in teaching versus general hospitals. This HR was also assumed for PFS since the analyses by Woo et al.,[132] indicated that the HR for overall and progression-free survival for teaching versus general hospitals are very similar. This study was obtained from a focused literature search, which was pragmatic in design. For this, the following resources were searched:

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

- Database of Abstracts of Reviews of Effects (DARE) (Wiley): to Issue 2 of 4, April 2015
- Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): to Issue 11 of 12, November 2016
- NHS Economic Evaluation Database (NHS EED) (Wiley): 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 patients with a benign mass, after the 30 day post-surgery period. All input parameters for the Markov model are reported in Table 22.

Table 22: Ovarian cancer and CRC probabilities

	Estimate	Se	Distribution	Source
Decision tree (short-term)				
Prevalence (all malignancies)	21.3%	1.0%	Beta	Akturk 2011,[72] Colombo 2009,[133] Coleman 2016,[71] Di Legge 2012,[62] Jacobs 1990,[79] Janas 2015,[98] Knafel 2016,[49] Lou 2010,[74] Meys 2016,[44] Moore 2011,[102] Morgante 1999,[81] Sayasneh 2016,[46] Shulman 2016,[105] Testa 2014,[50] Timmerman 2010,[66] van Gorp 2012,[99] Xu 2016,[96] Yanaranop 2016,[90]
Prevalence (all malignancies) - pre menopausal	16.2%	2.0%	Beta	Al Musahli 2016,[104] Coleman 2016,[71] Janas 2015,[98] Knafel 2016,[49] Meys 2016,[44] Piovano 2016,[58] Sayasneh 2013,[63] Testa 2014,[50] van Gorp 2012,[99] Yanaranop 2016,[90]
Prevalence (all malignancies) - post menopausal	45.9%	3.3%	Beta	Al Musahli 2016,[104] Coleman 2016,[71] Janas 2015,[98] Knafel 2016,[49] Meys 2016,[44] Piovano 2016,[58] Sayasneh 2013,[63] Testa 2014,[50] van Gorp 2012,[99] Yanaranop 2016,[90]
Prevalence non-ovarian malignancies (colorectal) within malignancies	2.9%	0.3%	Beta	Akturk 2011,[72] Colombo 2009,[133] Coleman 2016,[71] Di Legge 2012,[62] Jacobs 1990,[79] Janas 2015,[98] Knafel 2016,[49] Lou 2010,[74] Meys 2016,[44] Moore 2011,[102] Morgante 1999,[81] Sayasneh 2016,[46] Shulman 2016,[105] Testa 2014,[50] van Gorp 2012,[99] Xu 2016,[96] Yanaranop 2016,[90]
Advanced stage if ovarian malignancy*	75%		Fixed	Bell 1998[134]
IF CRC proportion of Dukes A	13.2%	0.1%	Dirichlet	National cancer Registration and Analysis Service 2010[135]
IF CRC proportion of Dukes B	36.9%	0.1%	Dirichlet	
IF CRC proportion of Dukes C	35.9%	0.1%	Dirichlet	
IF CRC proportion of Dukes D	14.0%		Dirichlet	
If FN proportion of ovarian cancer	100.0%		Fixed	assumption

	Estimate	Se	Distribution	Source
If FP proportion having benign mass	100.0%		Fixed	assumption
If TN proportion having benign mass	100.0%		Fixed	assumption
30 day post-surgery mortality early ovarian cancer	1.1%	0.5%		National Collaborating Centre for Cancer 2011,[1] Venesmaa 1992[136]
30 day post-surgery mortality advanced ovarian cancer	2.9%	0.3%		National Collaborating Centre for Cancer 2011,[1] Gerestein 2009[137]
30 day post-surgery mortality related to benign surgery	0.2%	0.0%		National Collaborating Centre for Cancer 2011,[1] Loft 1991[138]
Markov model (long-term)				
10 year progression free survival for early ovarian cancer	70.0%	4.7%	Beta	ICON 1, Collinson 2014[129]
10 year overall survival for early ovarian cancer	73.0%	4.0%	Beta	ICON 1, Collinson 2014[129]
2 year overall survival for advanced ovarian cancer	62.6%	1.8%	Beta	ICON 3 2002[130]
HR overall survival SMDT vs no SMDT treatment	0.900	0.048 = se ln(HR)	LogNormal	Woo 2012[132]
HR progression free survival SMDT vs no SMDT treatment	Assumed equal to HR for overall survival given that OS and PFS HRs for teaching vs general hospitals are very similar[132]			
Annual progression for Dukes A to B	58.3%	0.5%	Beta	Westwood 2016,[131] Tappenden 2007[139]
Annual progression for Dukes B to C	65.6%	0.8%	Beta	
Annual progression for Dukes C to D	86.7%	0.8%	Beta	
Mortality CRC	Time dependent, see Appendix 7 in[131] for more details			

* Only for true positives (see text)
Se: standard error

Risk score accuracy parameters

The proportions of patients testing positive (and thus referred to SMDT) or negative were based on the estimated accuracy of the risk scores considered (see section 3.2.7 and Table 23) and the estimated prevalence of all malignancies detected in this population (21.3% with standard error 1.0%). The proportion of true positives (TP), false positive (FP), false negative (FN) and true negatives (TN) were calculated as follows:

- TP = prevalence × sensitivity
- FP = (1 –prevalence) × (1 – specificity)
- FN = prevalence × (1 – sensitivity)
- TN = (1 –prevalence) × specificity

Subsequently, the proportions of patients who are referred to SMDT (TP + FP), and who are not referred to SMDT (TN + FN) were calculated. The results are listed in Table 24.

Table 23: Test accuracy

Risk score	Sensitivity	Se	Specificity	Se	Source (systematic review chapter 3)
RMI 1 (threshold 250)	64.4%	(1.4%)	91.8%	(0.7%)	Summary estimate derived from all studies, 6 published studies[74, 75, 77, 79-81] and 1 un-published study that reported data for RMI 1 (threshold 250) and the target condition 'all malignant tumours'
ROMA Abbott ARCHITECT	75.0%	(6.6%)	87.9%	(2.7%)	Al Musalhi 2016[104] (see Table 8)
ROMA Roche Elecsys	79.1%	(2.4%)	79.1%	(1.4%)	Summary estimate derived from 2 studies[98, 105] (see Table 11)
Overa (MIA2G) Vermillion (threshold 5 units)	90.2%	(2.5%)	65.8%	(1.9%)	Summary estimate derived from 2 studies[69, 71] (see Table 18)
IOTA Simple Rules (inconclusive assumed to be malignant)	94.2%	(0.5%)	76.1%	(0.6%)	Summary estimate derived from 8 published studies,[44, 48-50, 52, 55, 63, 66] and 1 un-published study (see Table 13)
IOTA ADNEX model (threshold 10%)	96.3%	(0.5%)	69.1%	(0.9%)	Summary estimate derived from 3 published studies,[17, 44, 46] and 1 un-published study (see Table 12)
RMI 1 (threshold 200)	68.1%	(0.9%)	90.1%	(0.5%)	Summary estimate derived from all studies, 12 published studies[44, 48, 50, 63, 74, 75, 77, 79-81, 99, 104] and 1 un-published study that reported data for RMI 1 (threshold 200) and the target condition 'all malignant tumours'

Table 24: Test outcomes

Risk score	TP	FP	FN	TN	PPV	NPV	LR+	LR-
RMI 1 (threshold 250)	13.7%	6.5%	7.6%	72.2%	0.68	0.90	7.85	0.39
ROMA Abbott ARCHITECT	16.0%	9.5%	5.3%	69.2%	0.63	0.93	6.20	0.28
ROMA Roche Elecsys	16.9%	16.4%	4.5%	62.2%	0.51	0.93	3.78	0.26
Overa (MIA2G) Vermillion (threshold 5 units)	19.2%	26.9%	2.1%	51.8%	0.42	0.96	2.64	0.15
IOTA 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 ADNEX model (threshold 10%)	20.5%	24.3%	0.8%	54.4%	0.46	0.99	3.12	0.05
RMI 1 (threshold 200)	14.5%	7.8%	6.8%	70.9%	0.65	0.91	6.88	0.35

After the risk assessment and referral decision, patients in the decision tree are allocated to 'early OC', 'advanced OC', 'benign mass', 'colorectal cancer' and 'Death'. One of the main assumptions in the decision tree was that the patients categorised as false negative all had early stage disease based on expert opinion i.e. the value of 25% for early stage only refers to the true positives.

Health state utilities

For patients with a benign mass, age-dependent general population utility estimates were used.[140] The utilities for early and advanced ovarian cancer were taken from Havrilesky et al. [141] and Grann et al.,[142] 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.[131, 143] was used to inform utilities for the four stages of colorectal cancer.

Table 25: Utility scores

Population	Estimate	Se	Distribution	Source
Benign mass (assumed equal to general population)	Age dependent			Ara 2010[140]
Early ovarian cancer SMDT treated	0.830	0.063	Beta	Havrilesky 2009[141]
Early ovarian cancer not SMDT treated	Assumed equal to SMDT treated			
Advanced ovarian cancer SMDT treated	0.630	0.247	Beta	Grann 1998[142]
Advanced ovarian cancer not SMDT treated	Assumed equal to SMDT treated			
Colorectal cancer Dukes' A	0.740	0.023	Beta	Ness 1999[143]
Colorectal cancer Dukes' B	0.670	0.026	Beta	
Colorectal cancer Dukes' C	0.500	0.031	Beta	
Colorectal cancer Dukes' D	0.250	0.028	Beta	

Resource use and costs related to the risk scores

Risk score costs are listed in Table 26.

Table 26: Risk score costs

Risk score	Estimate	Se / Range	Distribution	Source
RMI 1	£102	£20	Gamma	More detail on the calculation of these costs is provided below and in Appendix 6
ROMA Abbott ARCHITECT	£130	£26	Gamma	
ROMA Roche Elecsys	£126	£25	Gamma	
Overa (MIA2G) Vermillion	£176	£35	Gamma	
IOTA simple ultrasound rules	£77	£15	Gamma	
IOTA ADNEX model	£102	£20	Gamma	

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

<u>Ultrasound costs</u>

All risk scores entail, or are intended to be derived partly from, transvaginal ultrasound scans. Costs for these were informed by costs for transvaginal ultrasound scans used in CG122[1], and inflated to 2015/16 PSSRU values, resulting in a cost of £77.

CA125 test costs

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

RMI 1 costs

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

IOTA simple ultrasound rules

IOTA simple ultrasound rules only entails the costs of ultrasound scans (£77). The IOTA group stated that using the simple rules algorithm would not add to the time needed to conduct the scan or interpret results (personal communication: e-mail via. Thomas Walker, Technical Analyst, NICE Diagnostic Assessment Programme to Bram Ramaekers, Health Economics lead, Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University, 20/12/2016).

IOTA ADNEX model

This consists of the costs of ultrasound and CA125 and was therefore estimated to be £102.

Vermillion Overa (MIA2G)

Vermillion provided a cost estimate of £99 chargeable for the provision of its test. It was unclear whether this cost included all materials. The cost of ultrasound was added to it, resulting in £176. This cost was added because the company highlighted in their submission that Vermillion Overa (MIA2G) "is to be used in conjunction with a physician's independent clinical and radiological (imaging) evaluation. The test is not intended as a screening or stand-alone diagnostic assay."

ROMA (Abbott ARCHITECT and Roche 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 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. Below is a list of these assumptions:

Cost per HE4 test kit:

- The Abbott ARCHITECT HE4 test kit cost provided by the manufacturer was £21.33 per single test (Personal communication: e-mail from Abbott diagnostics to Thomas Walker, Technical Analyst, Diagnostics Assessment Programme, National Institute for Health and Care Excellence (NICE), 27th January 2017)
- The Roche Elecsys HE4 test kit cost provided by the manufacturer was £1,594.65, 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 Lumipulse G HE4, Abbott ARCHITECT HE4 and Roche Elecsys HE4 and they were based on the average analyser equipment cost of Fujirebio 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 seven work hours per day was assumed, and that two tests on average would be run per hour. This resulted in 3,542 HE4 tests run on one analyser per year. However, many tests are run at the same time and some labs would run these tests only weekly whilst some other 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 over-estimate. It, however, only amounted to £1.92 per test and therefore did not significantly affect model outcomes.

Maintenance costs:

- Only Fujirebio provided cost estimates for maintenance. The average of maintenance costs for Lumipulse G1200 and G600 II were assumed to be representative of maintenance costs for all analysers of different manufacturers. The two options of maintenance cover (fully comprehensive and preventative) were assumed to be adopted in equal proportions.

Quality control:

- Abbott and Roche provided cost estimates for their quality control. These were applied for each.

Calibration:

- Calibration costs were provided by Abbott and Roche, but it was not clear how often the calibration costs provided by Abbott were going to be incurred. The Roche calibration costs were therefore applied to both tests.

Shipping:

- Only Fujirebio provided costs for shipping each month. These were assumed to apply to Roche. Abbott stated that one shipment per month was free of charge and that further shipments were unlikely, so no further shipment costs were added to Abbott 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 over-estimate. The personnel cost was assumed to be that of a healthcare scientist derived from CURTIS (Unit Costs of Health and Social Care 2014) at £2.76 per test. Given that these are still relatively small, the potential over-estimation of personnel costs was not likely to affect model outcomes. These costs were therefore added to both Roche and Abbott tests.

Roche Elecsys HE4 test costs amounted to £23.75 and Abbott ARCHITECT HE4 test costs to £27.97. The main difference in costs between Roche Elecsys HE4 and Abbott 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 patients 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 an SMDT meeting, assuming no additional cost of the surgery or other investigations. The cost of this meeting (£116) is estimated to be that of SMDT Meetings, from the latest NHS reference costs.[144]

Treatment of ovarian cancer may consist of surgery and/or chemotherapy or supportive care, depending on disease stage. As assumed in CG122, chemotherapy consists of six cycles of carboplatin for patients with early ovarian cancer, and six cycles of carboplatin/paclitaxel for patients with advanced stage disease.[1] Surgery costs were also based on CG122[1], and calculated as a weighted average of relevant NHS reference costs, taking into account the probability of complications and the underlying disease (early or advanced ovarian cancer or benign mass).[144] The proportions of patients receiving each type of care were based on CG122[1], and unit costs of care from the latest Personal Social Services Research Unit (PSSRU) publication.[145] Frequency of follow-up costs for patients with ovarian cancer were based on CG122[1], and unit costs on the PSSRU publication.[145] Annual costs for

the CRC states were estimated from lifetime costs of CRC and mean survival as was done in the most recent DAR in CRC[131, 139] (see Table 27). Patients with a high risk score, but with a benign mass, are identified at the SMDT meeting to be false positive cases, and would undergo SMDT surgery without any further costs incurred in the tertiary care setting.

Table 27: Health state costs, event costs and unit prices

	Estimate (£)	Se	Distribution	Source
Health state costs colorectal cancer				
Colorectal cancer Dukes' A (lifetime costs)	£10,683	£3,959	Gamma	Westwood 2016,[131] Tappenden 2007[139]
Colorectal cancer Dukes' B (lifetime costs)	£18,015	£6,677	Gamma	
Colorectal cancer Dukes' C (lifetime costs)	£29,141	£10,800	Gamma	
Colorectal cancer Dukes' D (lifetime costs)	£19,392	£7,187	Gamma	
Colorectal cancer Dukes' A (annual costs)	£264			Calculated (using mean survival as in previous DAR), Westwood 2016[131]
Colorectal cancer Dukes' B (annual costs)	£609			
Colorectal cancer Dukes' C (annual costs)	£2,039			
Colorectal cancer Dukes' D (annual costs)	£8,391			
SMDT visit				
SMDT visit (necessary or unnecessary for benign)	£116		Fixed	calculated
Treatment costs ovarian cancer / benign mass				
Chemotherapy early ovarian cancer (6 cycles of carboplatin)	£1,898	£380	Gamma	calculated
Unit costs 1 cycle carboplatin	£316.29		Fixed	National Collaborating Centre for Cancer 2011,[1] British National Formulary 2016[146]
Chemotherapy administration early ovarian cancer	£1,417	£283		National Collaborating Centre for Cancer 2011[1]
Simple parenteral chemotherapy administration (per cycle)	£236		Fixed	NHS reference cost SB12Z[144]
Chemotherapy advanced ovarian cancer (6 cycles carboplatin + paclitaxel)	£5,905	£1.181	Gamma	National Collaborating Centre for Cancer 2011[1]
Unit costs paclitaxel	£667.88			National Collaborating Centre for Cancer 2011,[1] British National Formulary 2016[146]
Chemotherapy administration advanced ovarian cancer	£1,918	£384		calculated
More complex parenteral chemotherapy administration (per cycle)	£320		Fixed	NHS reference cost SB13Z[144]
Laparotomy malignancy without complication	£3,615	£723	Gamma	NHS reference cost M06C[144]
Laparotomy malignancy with complication	£4,566	£913	Gamma	NHS reference cost MA06A & MA06B[144]
Laparotomy benign mass without complication	£3.301	£660	Gamma	NHS reference cost MA07G & MA08B[144]
Laparotomy benign mass with complication	£4,112	£822	Gamma	NHS reference cost MA07E, MA07F & MA08A[144]
Proportion complication laparotomy benign mass	5.0%	1.0%		National Collaborating Centre for Cancer

	Estimate (£)	Se	Distribution	Source
Proportion complication laparotomy early ovarian cancer	5.0%	1.0%		2011[1]
Proportion complication laparotomy advanced ovarian cancer	12.5%	2.5%		
Number of hospital specialist care support	14.0	2.8	Gamma	
Unit costs hospital specialist care support	£100		Fixed	NHS reference cost SD03A[144]
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 cost SD01A[144]
Number of GP visits	1.0	0.2	Gamma	National Collaborating Centre for Cancer 2011[1]
Unit costs GP visit	£76		Fixed	PSSRU[145]
Number of district nurse visits	4.0	0.8	Gamma	National Collaborating Centre for Cancer 2011[1]
Unit costs district nurse visit	£42		Fixed	PSSRU[145]
Number of nurse specialist visits	2.0	0.4	Gamma	National Collaborating Centre for Cancer 2011[1]
Unit costs nurse specialist visit	£50		Fixed	PSSRU[145]
Total costs supportive care	£7,290			Calculated
Proportion chemotherapy early ovarian cancer	50%	10%	Beta	National Collaborating Centre for Cancer 2011[1]
Proportion chemotherapy advanced ovarian cancer	95%	(100% minus proportion supportive care)		
Proportion surgery early ovarian cancer	100%		fixed	
Proportion surgery advanced ovarian cancer	85%	17%	Beta	
Proportion supportive care early ovarian cancer	0%		Fixed	
Proportion supportive care advanced ovarian cancer	5%	1%	Beta	
Total treatment costs benign mass	£3,342			calculated
Total treatment costs early ovarian cancer	£5,320			
Total treatment costs advanced ovarian cancer	£10,606			
Follow up costs				
Annual number follow up visits (year 1-3)	4	0.8	Gamma	National Collaborating Centre for Cancer 2011[1]
Annual number follow up visits (> year 3)	1	0.2	Gamma	
Unit costs follow up visit	£92		Fixed	PSSRU[145]
Total annual follow up costs (year 1-3)	£398			Calculated
Total annual follow up costs (> year 3)	£92			

4.2.4 Overview of main model assumptions

The main assumptions in the health economic analyses were:

- All non-ovarian malignancies are CRC malignancies.

- FN are more likely to be early stage than advanced stage ovarian cancer .
- For IOTA Simple ultrasound Rules, inconclusive assessments would be assumed to be malignant.
- Carboplatin costs (6 cycles) for EOC and carboplatin + paclitaxel costs (6 cycles) for AOC (without avastin).
- List prices are used for carboplatin and paclitaxel.
- The HR of 0.900 retrieved from the Cochrane review by Woo et al [132], focussing on the comparison of institutions with gynaecologic oncologists on site versus community or general hospitals is representative for the relative PFS and OS for SMDT versus no SMDT.
- All FP and FN patients will be operated on for a benign mass.
- No disutility is incorporated for FP patients (i.e. people who are incorrectly told they have ovarian cancer).

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

4.3 Model analyses

Expected costs, life years (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 cost 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 (15,000 simulations) were performed and cost effectiveness acceptability curves (CEACs) were constructed.

4.3.1 Sensitivity analyses

Deterministic one-way sensitivity analyses were performed, using all input parameters incorporated stochastic in probabilistic sensitivity analyses (PSA) 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.

4.3.2 Scenario analyses

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

- Assuming a prevalence of 20% for all malignancies
- Assuming a prevalence of 30% for all malignancies
- Assuming 0% prevalence of non-ovarian (CRC) malignancies

- Assuming an equal proportion of early versus advanced stage ovarian cancer in the FN and TP groups (in the base case it was assumed that FN would all be early stage)
- Assuming, for IOTA Simple ultrasound Rules, that subjective assessment would be used for inconclusive assessments (instead of assumed to be malignant). 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 no ultrasound is performed in conjunction with ROMA and Overa (MIA2G) risk scores, thus reducing the costs of these risk scores
- Assuming additional costs for FP (surgery costs with malignancy instead of without) and additional costs for FN (additional costs of benign surgery)
- Assuming additional costs for FP (surgery costs with malignancy instead of without) and additional costs for FN (additional costs of benign surgery and SMDT costs)
- Assuming a discount of 92% for carboplatin (CG122: discount England 91.8%, discount Wales 92.1%)
- Assuming a discount of 95% for paclitaxel (CG122: discount England 91.0%, discount Wales 95.4%)
- Assuming an alternative HR for progression-free and overall survival for SMDT versus no SMDT (of 0.808)[147] This study was selected as it was not included in the Cochrane review, used in the base-case, and provided an alternative HR (N=275, N=238 for this comparison).
- Assuming an alternative HR for progression-free and overall survival for SMDT versus no SMDT (of 0.990; the upper bound of the confidence interval used in the base-case)
- Assuming that the proportion of patients receiving supportive care (for advanced stage) is 10% (instead of 5%)
- Assuming alternative transvaginal ultrasound cost of £142.46 (MA36Z) (instead of £76.75 based on CG122)
- Assuming alternative transvaginal ultrasound cost of £142.46 (MA36Z) (instead of £76.75 based on CG122) and increasing the transvaginal ultrasound cost for the IOTA risk scores by 20% (to reflect potential training costs)
- Assuming additional SMDT costs of £2,500 to reflect higher surgery costs given that according to expert opinion 1 in 3 to 4 patients referred to SMDT may receive extensive surgery for ovarian cancer (IPG 470) for which the price is unknown
- Assuming 90% of the non-malignancy surgery and complications costs for TN reflecting a scenario wherein 90% of the TN are operated (instead of all)

- Assuming avastin for advanced stage. Assuming additional treatment costs of £17,760 per treated patient [148] and assuming a median survival of 39.7 months (95%CI 36.0–44.2), derived from clinically predefined high-risk subgroup of the ICON7 trial. This subgroup was used given that 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 [149]
- Assuming a disutility for FP during the first year in state transition model of 0.100
- Assuming a disutility for FP during the first year in state transition model of 0.010
- Using the optimal RMI threshold (i.e. RMI threshold cost effective at £20,000 and/or £30,000 per QALY gained in former scenario), based on a comparison of only different RMI thresholds (see Appendix 7 for accuracy estimates)

4.3.3 Subgroup analyses

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

- Pre-menopausal women (mean age of 38, subgroup-specific accuracy data)
- Post-menopausal women (mean age of 68, 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

4.4 Results of 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.

4.4.1 Base-case analysis

The base case analysis includes seven risk scores. Tables 28 and 29, as well as Figure 10, show the probabilistic results of this analysis. The RMI 1, with a threshold of 250, was least effective (16.926 life years, 13.820 QALYs) and second cheapest (£5,669). The IOTA simple ultrasound rules (inconclusive assumed to be malignant), was cheapest (£5,667) and second most effective (16.954 life years, 13.841 QALYs) and thereby dominating the RMI 1 (at both the 200 and 250 thresholds). The IOTA ADNEX model (threshold 10%), costing £5,699, was most effective (16.957 life years, 13.843 QALYs) and compared with the IOTA simple ultrasound rules resulted in an ICER of £15,304 per QALY gained. The remaining risk scores (ROMA Abbott ARCHITECT (threshold 7.4%/25.3%), ROMA Roche Elecsys (threshold 11.4%/29.9%) and Overa (MIA2G) Vermillion (threshold five units)) were dominated. As a result, incremental analysis indicated that up to thresholds of £15,304 per

QALY gained the IOTA Simple Rules is cost effective whereas the IOTA ADNEX model (threshold 10%) is cost effective for higher thresholds.

Table 28: Probabilistic results for base case analysis

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

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 simple ultrasound rules and IOTA ADNEX model (threshold 10%) this was 39% and 60% respectively (£20,000 threshold) and 23% and 75% respectively (£30,000 threshold). The probabilities for the other risk scores were <1% for these thresholds (Figure 11).

Figure 10: Probabilistic results

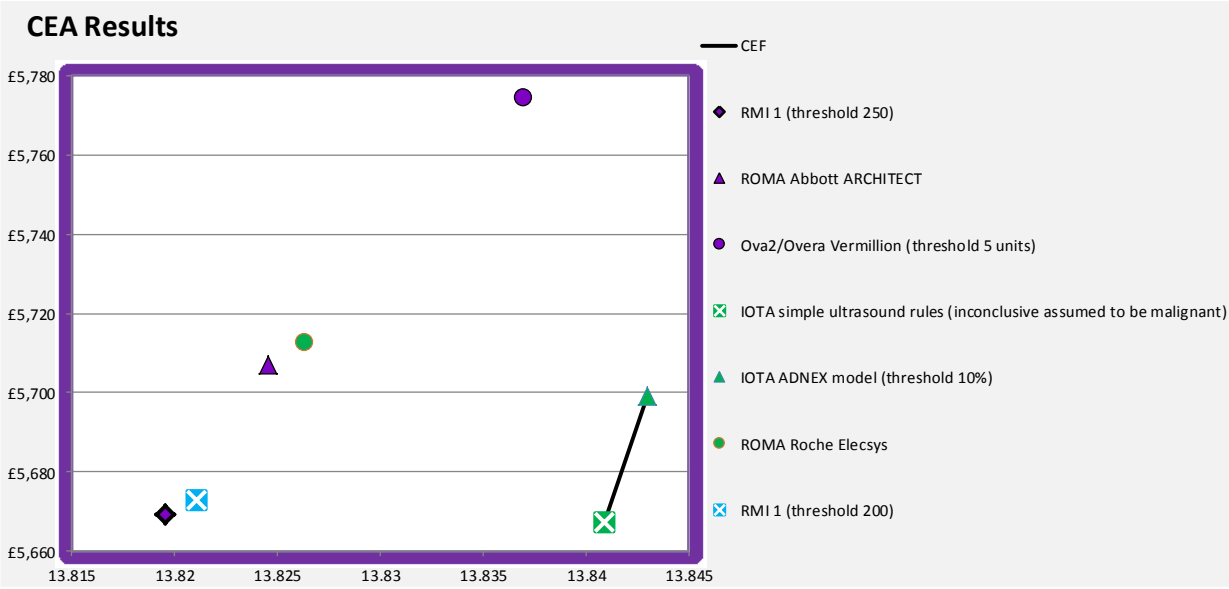
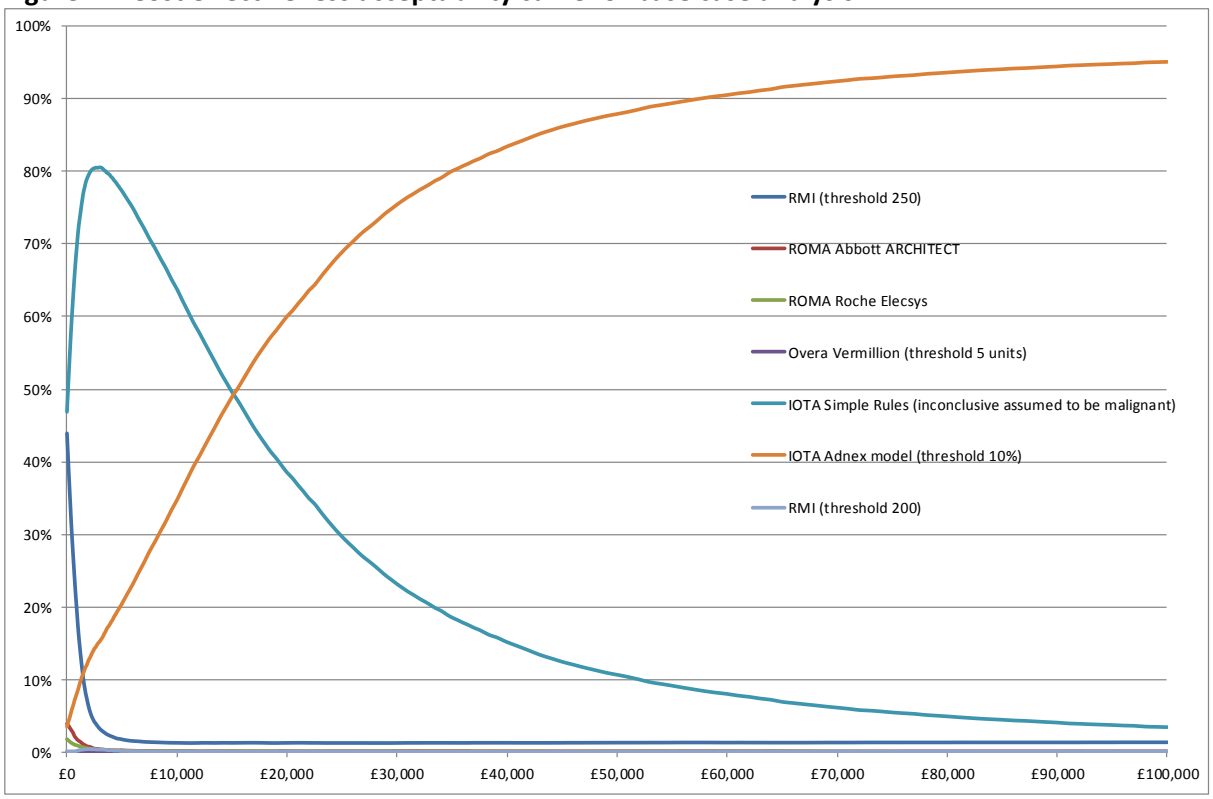


Table 29: Probabilistic results for base case analysis: costs, QALYs and incremental analysis

Risk score	Costs (95% CI)	QALYs (95% CI)	Compared to RMI 250			Full incremental
			ΔCosts	ΔQALYs	ΔCosts / ΔQALYs	ΔCosts / ΔQALYs
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,667 (95% CI: £4,551 to £6,941)	13.841 (95% CI: 13.477 to 14.154)	-£2	0.021	dominant	cheapest
RMI (threshold 250)	£5,669 (95% CI: £4,553 to £6,934)	13.82 0 (95% CI: 13.456 to 14.134)				dominated
RMI (threshold 200)	£5,673 (95% CI: £4,557 to £6,939)	13.821 (95% CI: 13.456 to 14.135)	£4	0.002	£2,483	dominated
IOTA ADNEX model (threshold 10%)	£5,699 (95% CI: £4,585 to £6,973)	13.843 (95% CI: 13.480 to 14.155)	£30	0.023	£1,274	£15,304
ROMA Abbott ARCHITECT	£5,707 (95% CI: £4,593 to £6,976)	13.825 (95% CI: 13.458 to 14.138)	£38	0.005	£7,506	dominated
ROMA Roche Elecsys	£5,713 (95% CI: £4,597 to £6,985)	13.826 (95% CI: 13.461 to 14.141)	£44	0.007	£6,409	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,775 (95% CI: £4,655 to £7,049)	13.837 (95% CI: 13.472 to 14.151)	£105	0.017	£6,038	dominated

Figure 11: Cost effectiveness acceptability curve for base case analysis



4.4.2 Sensitivity analyses

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

1. Progression-free and overall survival hazard ratios for SMDT versus no SMDT referral
2. Test costs
3. Utility of advanced ovarian cancer
4. SMDT costs
5. Test sensitivity
6. Discount rate
7. Test specificity

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

- Upper bound (0.990) for the OS HR for SMDT versus no SMDT
- Upper bound (95.1%) for sensitivity for IOTA Simple ultrasound Rules
- Lower bound (95.3%) for sensitivity for the IOTA ADNEX model

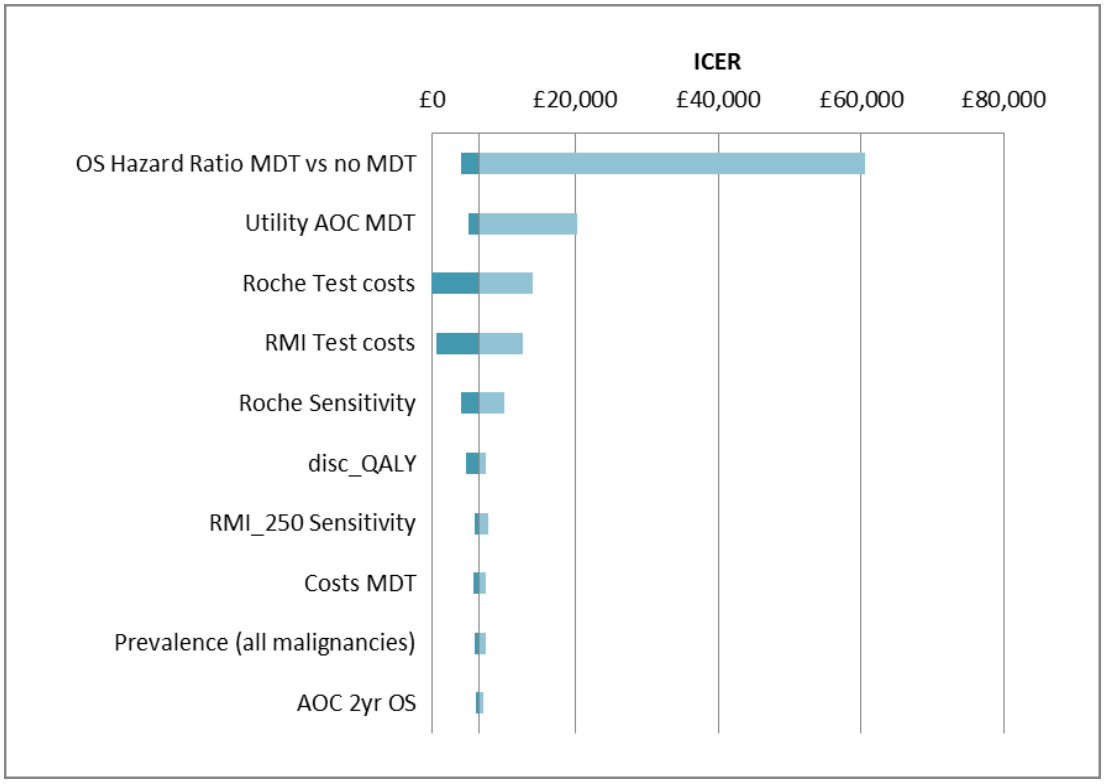
- Lower bound (£47) for IOTA Simple ultrasound Rules costs
- Upper bound (£142) for IOTA ADNEX model costs

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

- Upper bound (0.990) for the OS HR for SMDT versus no SMDT
- Upper bound (£142) for IOTA ADNEX model costs

These results are shown in the tornado diagrams in Appendix 8 and below in Figure for the comparison between RMI 1 (threshold 250) and ROMA Roche Elecsys.

Top 10 influential parameters in the comparison of the RMI 1 (threshold 250) versus ROMA Roche Elecsys



4.4.3 Scenario analyses

The scenario analyses include 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 ADNEX model (threshold 10%) remained the most cost effective strategy. This excludes the scenarios

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

In these scenario analyses, IOTA simple ultrasound rules (inconclusive assumed to be malignant) was cost effective at a threshold of £20,000 per QALY gained, while again the IOTA ADNEX model (threshold 10%) was cost effective at a threshold of £30,000 per QALY gained. Moreover, in the following scenario analyses IOTA 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 versus no SMDT (of 0.990; the upper bound of the confidence interval used in the base-case)
- assuming a disutility for FP during the first year in state transition model of 0.100

Finally, in the scenario with increased SMDT surgery costs, given that according to expert opinion 1 in 3 to 4 patients referred to SMDT may receive extensive surgery for ovarian cancer (IPG 470) for which the price is unknown, RMI (threshold 250) was cost effective at a threshold of £20,000 per QALY gained while IOTA simple ultrasound rules was cost effective at a threshold of £30,000 per QALY gained.

When comparing only different RMI 1 thresholds, it was found that RMI 1 with a threshold of 25 would be cost effective at all thresholds of £2,890 per QALY gained or higher. However, when including this RMI 1threshold , optimal sensitivity, (instead of RMI 1 with a threshold of 200) in the base-case analysis, the RMI 1 was still dominated.

4.4.4 Subgroup analysis

Pre-menopausal

The pre-menopausal subgroup analysis used a different mean starting age (38 years) and drew on subgroup specific accuracy data (see Appendix 7). Tables 30 and 31, as well as Figure 12, show the probabilistic results of the subgroup analysis in pre-menopausal women. ROMA Abbott ARCHITECT was the least effective (18.108 life years, 14.927 QALYs) and the RMI 1 with a threshold of 200 was the cheapest (£5,188), followed by IOTA simple ultrasound rules (inconclusive assumed to be malignant) at £5,189. The most effective options were the IOTA ADNEX model (18.137 life years,

14.948 QALYs) followed by IOTA simple ultrasound rules (18.135 life years, 14.946 QALYs). Consequently, incremental analysis indicated that between thresholds of £15 and £18,304 per QALY gained the IOTA Simple ultrasound Rules is cost effective whereas the IOTA ADNEX model (threshold 10%) is cost effective for higher thresholds.

Table 30: Probabilistic results for pre-menopausal subgroup analysis

Risk score	Life years	Compared to RMI 250
RMI 1 (threshold 200)	18.108 (95% CI: 17.435 - 18.720)	-0.006
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	18.135 (95% CI: 17.470 - 18.740)	0.021
RMI 1 (threshold 250)		
ROMA Abbott ARCHITECT	18.108 (95% CI: 17.434 - 18.720)	-0.006
IOTA ADNEX model (threshold 10%)	18.137 (95% CI: 17.472 - 18.741)	0.024
ROMA Roche Elecsys	18.132 (95% CI: 17.461 - 18.737)	0.018
Overa (MIA2G) Vermillion (threshold 5 units)	18.131 (95% CI: 17.464 - 18.736)	0.018

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

Figure 12: Probabilistic results premenopausal subgroup

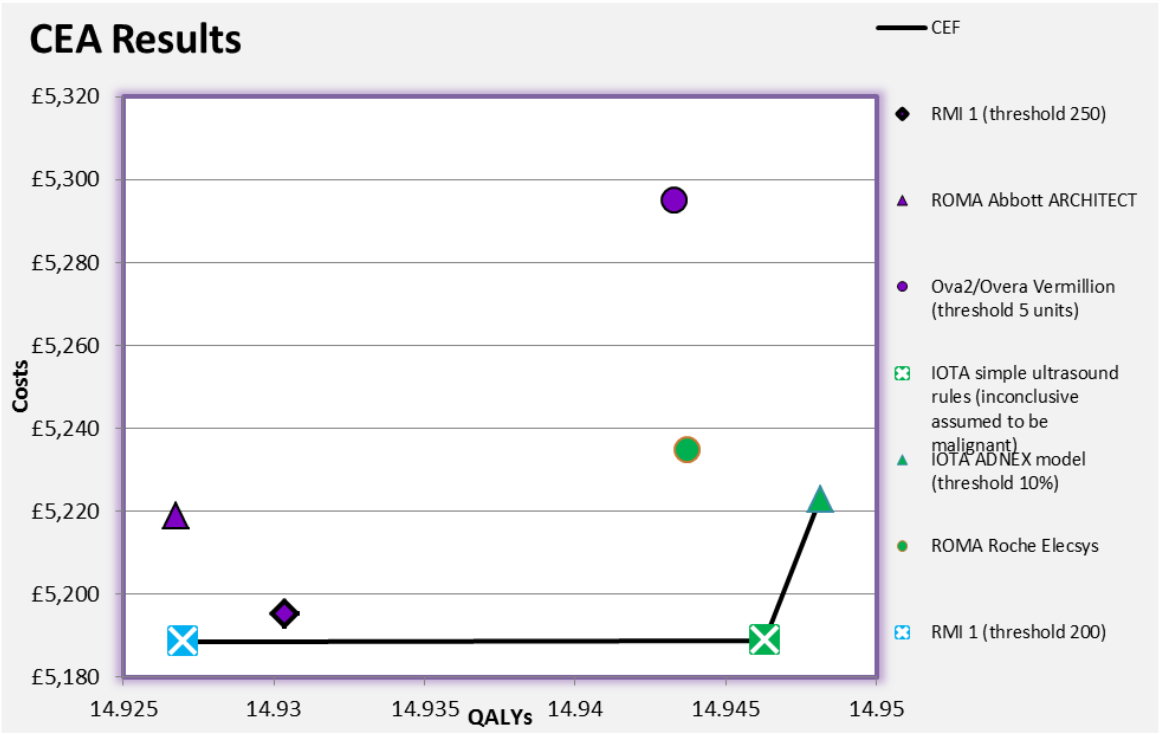
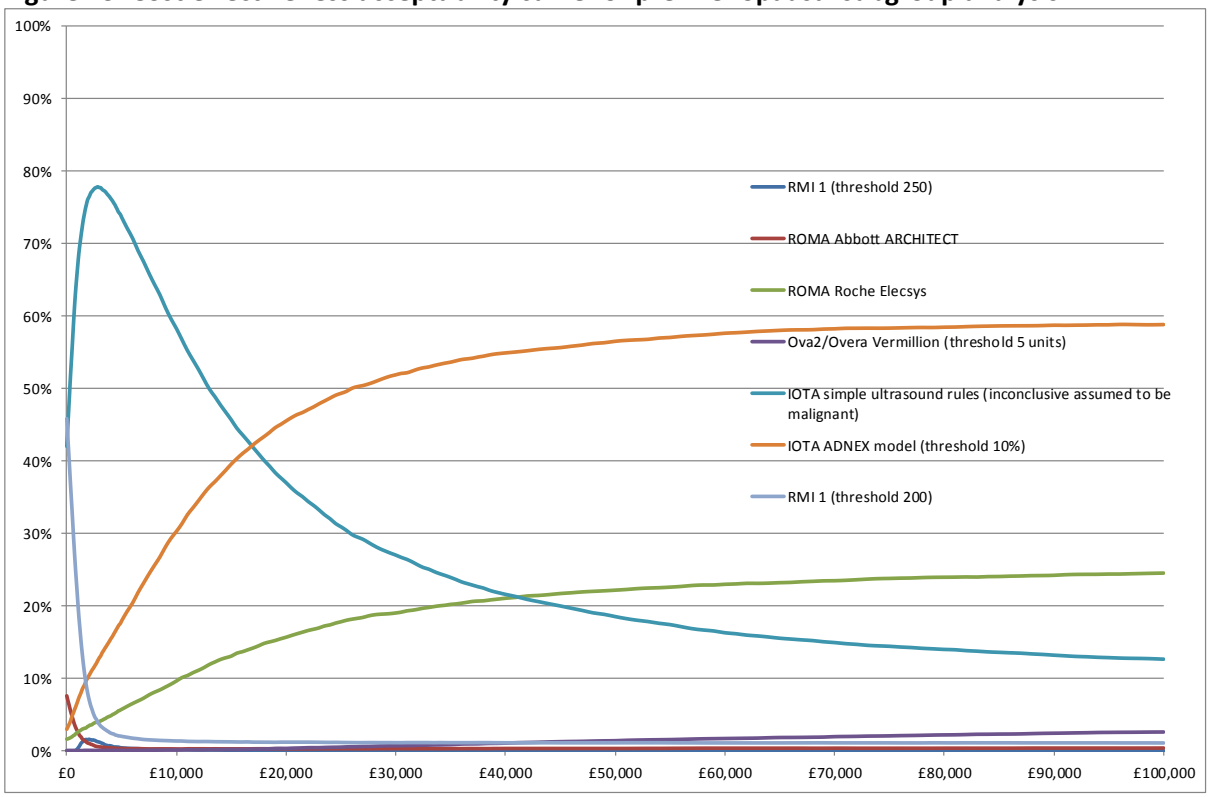


Table 31: Probabilistic results for pre-menopausal subgroup analysis: costs, QALYs and incremental analysis

Risk score	Costs (95% CI)	QALYs (95% CI)	Compared to RMI 250			Full incremental
			ΔCosts	ΔQALYs	ΔCosts / ΔQALYs	ΔCosts / ΔQALYs
RMI 1 (threshold 200)	£5188 (95% CI: £4045 - £6510)	14.927 (95% CI: 14.309 - 15.471)	-£7	-0.003	£1,954	cheapest
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5189 (95% CI: £4046 - £6515)	14.946 (95% CI: 14.331 - 15.486)	-£6	0.016	dominant	£15
RMI 1 (threshold 250)	£5195 (95% CI: £4051 - £6516)	14.93 (95% CI: 14.311 - 15.473)	£0	0.000		dominated
ROMA Abbott ARCHITECT	£5219 (95% CI: £4076 - £6542)	14.927 (95% CI: 14.308 - 15.471)	£24	-0.004	dominated	dominated
IOTA ADNEX model (threshold 10%)	£5223 (95% CI: £4081 - £6549)	14.948 (95% CI: 14.335 - 15.487)	£28	0.018	£1,564	£18,466
ROMA Roche Elecsys	£5235 (95% CI: £4090 - £6571)	14.944 (95% CI: 14.329 - 15.484)	£40	0.013	£2,993	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5295 (95% CI: £4150 - £6617)	14.943 (95% CI: 14.327 - 15.484)	£100	0.013	£7,748	dominated

Figure 13: Cost effectiveness acceptability curve for pre-menopausal subgroup analysis



Post-menopausal

The post-menopausal subgroup analysis used a different mean starting age (68 years) and drew on subgroup specific accuracy data (see Appendix 7). Tables 32 and 33, as well as Figure 14, show the probabilistic results of the subgroup analysis in post-menopausal women, mean age 68. The RMI 1 with a threshold of 250 was the least effective (8.031 life years, 5.690 QALYs). The cheapest risk score was IOTA simple ultrasound rules (£7,742) which was £1 cheaper than RMI 1 with a threshold of 250. The most effective option was the IOTA ADNEX model (8.076 life years, 5.721 QALYs). IOTA simple ultrasound rules were the second most effective (8.072 life years, 5.718 QALYs) and cost effective up to a threshold of £12,876 per QALY gained; thereafter, the IOTA ADNEX model was cost effective. The other risk scores (the RMI 1 at a threshold of 200, ROMA Abbott ARCHITECT, ROMA Roche Elecsys and Overa (MIA2G) Vermillion) were dominated.

Table 32: Probabilistic results for postmenopausal subgroup analysis, mean age 68: life years

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

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 <2%. For the IOTA simple ultrasound rules and IOTA ADNEX model (threshold 10%) this was 40% and 59% respectively (£20,000 threshold) and 24% and 74% respectively (£30,000 threshold). The probabilities for the other risk scores were <1% for these thresholds (Figure 15).

Figure 14: Probabilistic results post-menopausal subgroup, mean age 68

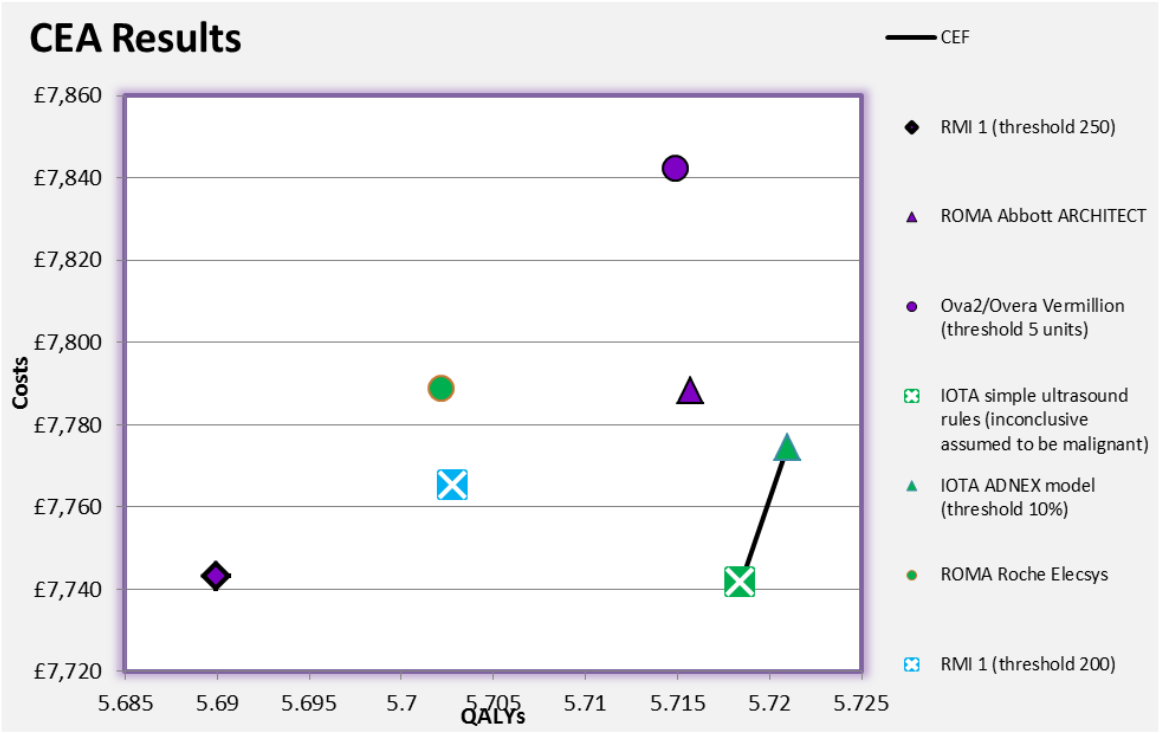
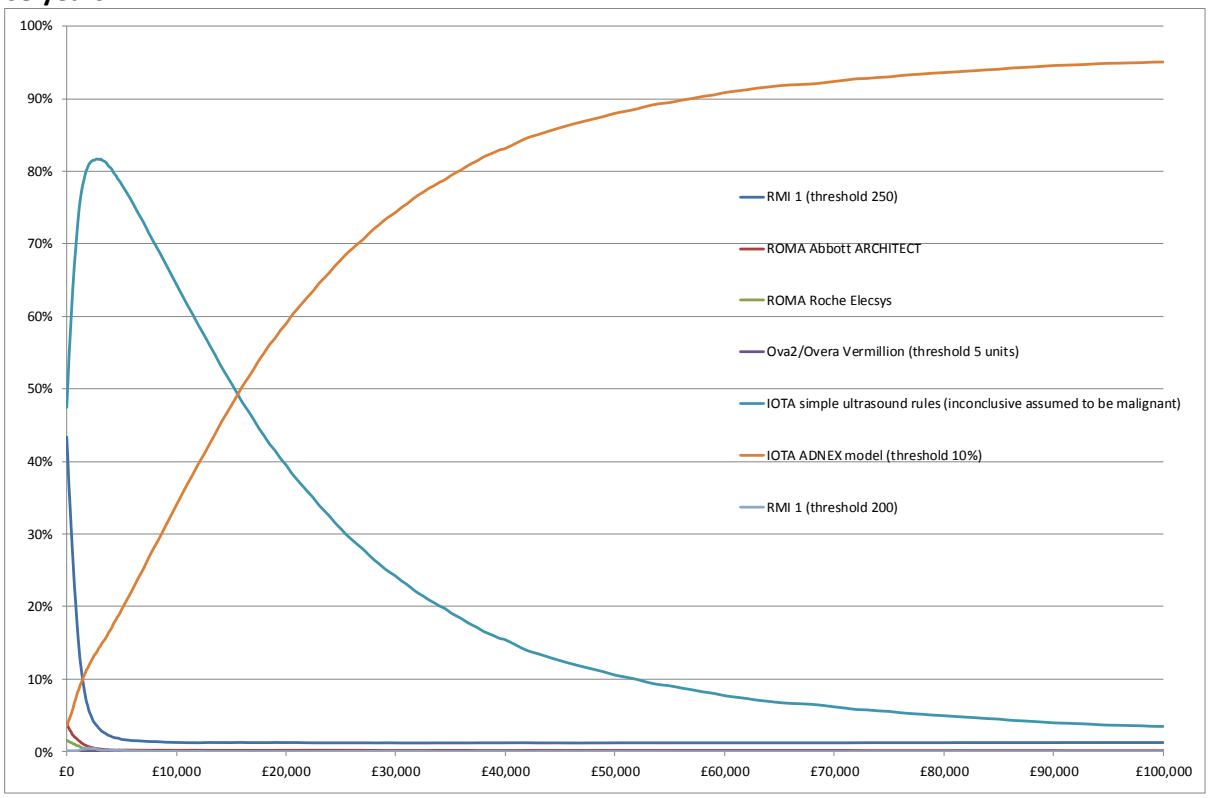


Table 33: Probabilistic results for post-menopausal subgroup analysis, mean age 68: costs, QALYs and incremental analysis

Risk score	Costs (95% CI)	QALYs (95% CI)	Compared to RMI 250			Full incremental
			ΔCosts	ΔQALYs	ΔCosts / ΔQALYs	ΔCosts / ΔQALYs
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£7742 (95% CI: £6338 - £9281)	5.718 (95% CI: 5.061 - 6.178)	-£1	0.028	dominance	cheapest
RMI (threshold 250)	£7743 (95% CI: £6334 - £9289)	5.69 (95% CI: 5.035 - 6.153)	£0	0.000		dominated
RMI (threshold 200)	£7765 (95% CI: £6356 - £9309)	5.703 (95% CI: 5.043 - 6.168)	£22	0.013	£1,746	dominated
IOTA ADNEX model (threshold 10%)	£7774 (95% CI: £6370 - £9318)	5.721 (95% CI: 5.063 - 6.181)	£31	0.031	£1,013	£12,876
ROMA Abbott ARCHITECT (threshold 7.4%/25.3%)	£7788 (95% CI: £6381 - £9329)	5.716 (95% CI: 5.059 - 6.177)	£45	0.026	£1,759	dominated
ROMA Roche Elecsys (threshold 11.4%/29.9%)	£7789 (95% CI: £6377 - £9334)	5.702 (95% CI: 5.044 - 6.168)	£46	0.012	£3,738	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£7842 (95% CI: £6429 - £9396)	5.715 (95% CI: 5.058 - 6.176)	£99	0.025	£3,992	dominated

Figure 15: Cost effectiveness acceptability curve for post-menopausal subgroup analysis, mean age 68 years



Additional subgroup analyses

The results for the subgroup where 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 thresholds of £20,000 and £30,000 per QALY gained, the IOTA ADNEX model (threshold 10%) remained the most cost effective strategy. In contrast, for the subgroup consisting of advanced stage ovarian cancer only, IOTA simple ultrasound rules were cost effective at thresholds of £20,000 and £30,000 per QALY gained. The tabulated results are provided in Appendix 10.

5. DISCUSSION

5.1 Statement of principal findings

5.1.1 Clinical effectiveness

All studies included in our systematic review were diagnostic cohort studies which reported data on the diagnostic accuracy of one or more ovarian cancer risk scores (ROMA score, IOTA simple ultrasound roles, the ADNEX model, or Overa (MIA2G)), or which provided data on the accuracy of the RMI 1 at different decision thresholds (including 250, as specified in current NICE guidelines[1]). With the exception of Overa (MIA2G), studies were identified which provided direct comparisons of the performance of each included risk score versus RMI 1 (threshold 200), i.e. where the performance of the intervention risk score and the performance of the RMI 1 (threshold 200) were assessed in the same patient cohort. No study reported data to allow a direct comparison of all included index tests (risk scores) to each other and the RMI, 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 Roche Elecsys or Abbott ARCHITECT tumour marker assays. None of the included studies used the Fujirebio Lumipulse G automated CEIA system. We included two studies that used a ROMA score based on manual Fujeribio tumour marker EIA assays (Appendix 4, Tables 41 and 42).[95, 99] 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 the NICE clinical guideline CG122,[1] i.e. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma, and borderline ovarian cancer; excluded conditions are pseudomyxoma peritonei, relapsed cancers, germ cell tumour 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 mass. The definition of ovarian cancer varied between studies and did not always include borderline tumours. Additionally, the definition of disease/reference standard positive could include all malignancies or only ovarian malignancies. Although studies which report the performance of risk scores for the specific target condition of ovarian cancer (as described in CG122)[1] could be considered most applicable to the scope of this assessment (and, accordingly, have been rated as having '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 participants 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 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 Abbott ARCHITECT tumour marker assays. The summary sensitivity estimate for the ROMA score (using the manufacturer's recommended cut-off values of 7.4% in pre-menopausal women and 25.3% in post-menopausal women) was highest, 96.4% (95% CI: 93.6 to 98.2%), where analyses excluded participants with borderline tumours and those with malignancies other than epithelial ovarian cancer and lowest, 75.0% (95% CI: 60.4 to 86.4%), where all participants were included in the analysis, regardless of their final histopathological diagnosis and 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 which included all participants in the analysis reported similar sensitivity and specificity estimates for the ROMA score and the RMI 1 (threshold 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.[104] By contrast, where participants 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 (threshold 200), 80.3% (95% CI: 77.5 to 82.9). The only study to report a direct comparison of the ROMA score, using Roche Elecsys tumour marker assays (manufacturer's recommended thresholds of 11.4% in pre-menopausal women and 29.9% in post-menopausal women), with the RMI 1 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 slightly higher than that for the RMI 1, 83.8% (95% CI: 73.4 to 91.3%) versus 78.4 (95% CI: 67.3 to 87.1%) and the specificity estimate for the ROMA score appeared slightly lower than that for the RMI 1, 68.8% (95% CI: 61.6 to 75.4%) versus 79.6% (95% CI: 73.1 to 85.1%), but neither difference was statistically significant.[90] This study may be considered more applicable to clinical practice, if it is considered preferable to manage patients with borderline tumours in non-specialist settings. The summary estimates of sensitivity and specificity for the ROMA score, using Roche Elecsys tumour marker assays at the manufacturer's recommended thresholds, derived from non-comparative accuracy studies where all participants 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 where the manufacturer's recommended cut-offs were used, the performance of the ROMA score did not differ significantly between pre- and post-menopausal women.

When considering the risk scoring methods produced by the IOTA group. Our report focuses on data for the ADNEX model where the validated 10% threshold is used and on data for the IOTA simple ultrasound rules where 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 where patients with an inconclusive IOTA simple ultrasound rules assessment were not classified (excluded from the analyses) are reported in Appendix 4, Table 39, however, these results are considered to be of limited clinical value, since it is unclear what alternative methods might be used to select the most appropriate care pathway for these patients. The majority of these studies included all participants in their 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 IOTA simple ultrasound rules where inconclusive results were assumed to be malignant, 94.2% (95% CI: 93.3 to 95.1%); where subjective assessment was applied to inconclusive and IOTA 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 IOTA simple ultrasound rules where 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%), where subjective assessment was applied to inconclusive and IOTA simple ultrasound rules results. Menopausal status did not significantly affect the performance of either the ADNEX model or the IOTA simple ultrasound rules, however, the specificity estimate was significantly higher in pre-menopausal women than in post-menopausal women, for both instruments. One published study[44] and one unpublished interim report (personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017) provided comparative accuracy data for the ADNEX model, IOTA simple ultrasound rules where 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 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 simple ultrasound rules, respectively. The summary estimate of sensitivity, for 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 simple ultrasound rules estimates. Conversely, the specificity estimate for RMI 1 at a decision threshold of 200 was significantly higher, 89% (95%CI: 87 to 90.7%) than both the ADNEX model and IOTA simple ultrasound rules estimates.

No studies were identified that directly compared Overa (MIA2G) to RMI 1 at either decision threshold (200 or 250). One study[105] reported comparative accuracy data for Overa (MIA2G) versus the ROMA score, using Roche Elecsys tumour marker assays. This study included all participants in the analysis, regardless of their final histopathological diagnosis (i.e. target condition all malignancies including borderline). At a threshold of five units, the sensitivity estimate for Overa (MIA2G) was, 91% (95% CI: 86.8, 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 the ROMA score, 79.2% (95% CI: 73.7 to 83.8%), whilst the specificity of the Overa (MIA2G) score was significantly lower than the ROMA score, 78.9 % (95% CI: 75.8, 81.7).

Summary estimates derived from studies which compared the diagnostic performance of different RMI 1 decision thresholds (between 25 and 500) and which 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 the sensitivity RMI 1, derived from studies included in our systematic review were lower than that reported in a recent systematic review (75% [95% CI: 72 to 74%], based on 14 studies),[150] 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 included studies of women undergoing surgery for adnexal mass and excluded any studies which selectively excluded some histopathological subtypes of ovarian cancer or which classified borderline tumours as benign.[150] As would be expected, the sensitivity estimate for the RMI 1 increased and specificity decreased with decreasing threshold.

For both the IOTA simple ultrasound rules and the ADNEX model, there was evidence that specificity can be significantly decreased in post-menopausal women in comparison to overall populations or premenopausal women. Neither of these risk score 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 outlined

above, 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 which 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); we therefore consider that studies which include all participants in their analysis, irrespective of final histological diagnosis, are more likely to produce estimates of risk score performance which 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 which included only those studies reporting a direct comparison to the RMI 1 (see sections 3.2.3, 3.2.4 and 3.2.6). Therefore, 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 as a separate intervention, for each tumour marker manufacturer (Roche Elecsys and Abbott ARCHITECT; none of the included studies used the Fujirebio Lumipulse G automated CEIA system, 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 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, pre-menopausal women or post-menopausal women). Where no data were available for the RMI 1 with a decision threshold of 250, we used data for a decision threshold of 200; the analysis reported in section 3.2.6 indicated no significant difference in the performance of RMI 1 at these two thresholds.

5.1.2 Cost effectiveness

The review of economic analyses examined studies reporting outcomes of a full cost effectiveness analysis, examining (quality-adjusted) life-years, 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 while the remaining two considered secondary care from the UK and US perspectives. The UK study, indicated that multimodal screening consisting of CA125 followed by TVS could be cost effective compared with ultrasound screening and no screening. The two studies considering MIA, both from the US perspective, provided conflicting results, one indicating 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 ROMA and also indicated that this score would be dominated by other strategies (when considering LYs). Moreover, this study indicated that “refer all” is cost effective for thresholds above $10,644 per LY gained. In conclusion, there is limited and conflicting evidence regarding the cost effectiveness of alternative risk scores,

which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥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, CA125 or ultrasound, compared to the RMI 1 score as used in current practice for patients with suspected ovarian in secondary care, was assessed to guide decisions about referral to a specialist SMDT. The base-case analysis includes seven risk scores:

- RMI 1 (threshold 250)
- ROMA Abbott ARCHITECT
- ROMA Roche Elecsys
- Overa (MIA2G) Vermillion (threshold 5 units)
- IOTA simple ultrasound rules (inconclusive assumed to be malignant)
- IOTA ADNEX model (threshold 10%)
- RMI 1 (threshold 200)

In the base-case analysis, the RMI 1 with a threshold of 250 was least effective (16.926 life years, 13.820 QALYs) and second cheapest (£5,669). The IOTA Simple Rules (inconclusive assumed to be malignant), was cheapest (£5,667) and second most effective (16.954 life years, 13.841 QALYs) and thereby dominating the RMI 1 (at both the 200 and 250 thresholds). The IOTA ADNEX model (threshold 10%), cost £5,699, was most effective (16.957 life years, 13.843 QALYs) and compared with the IOTA simple ultrasound rules resulted in an ICER of £15,304 per QALY gained. The remaining risk scores (ROMA Abbott ARCHITECT, ROMA Roche Elecsys and Overa (MIA2G) Vermillion) were dominated. As a result, incremental analysis indicated that up to thresholds of £15,304 per QALY gained the IOTA simple ultrasound rules is cost effective whereas the IOTA ADNEX model (threshold 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 simple ultrasound rules and IOTA ADNEX model (threshold 10%) this was 39% and 60% respectively (£20,000 threshold) and 23% and 75% respectively (£30,000 threshold). The probabilities for the other risk scores were <1% for these thresholds.

Sensitivity and scenario analyses indicated the hazard ratio for SMDT versus no SMDT referral (for patient with ovarian cancer) was the most influential parameter in the model and that the results are reasonably robust. Most scenario analyses indicated that at thresholds of £20,000 and £30,000 per QALY gained, the IOTA ADNEX model (threshold 10%) remained the cost-effective strategy. In two scenario analyses IOTA simple ultrasound rules (inconclusive assumed to be malignant) was considered 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 £2,890 per QALY gained or higher), the RMI 1 was still dominated.

For the pre-menopausal and post-menopausal subgroups, the IOTA ADNEX model (threshold 10%) was cost effective at thresholds of £20,000 and £30,000 per QALY gained.

5.2 Strengths and limitations of assessment

5.2.1 Clinical effectiveness

We are not aware of any previous systematic review that has considered the performance of both ultrasound-based risk scores such as IOTA simple rules and biomarker-based scores such as ROMA and Overa (MIA2G). The most recent systematic review of the ROMA score completed searching in November 2014.[151] Additionally, previous systematic reviews of the ROMA score have focused on predicting ovarian cancer (no definition reported) or epithelial ovarian cancer, 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.[9, 151, 152] A more recent (searches completed July 2015) systematic review is available for ultrasound-based risk scores,[150] however, previous systematic reviews have tended to focus on comparing these scores to subjective ultrasound evaluation[150, 153] rather than to other types of risk scoring. Risk scoring for ovarian malignancy is a rapidly evolving field and we believe that the full update, comparing all options currently available to the UK NHS, provided by this assessment will be of value to clinicians and decision makers. Additionally, there is currently a large, ongoing Cochrane review, 'Symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer,' which will provide data on testing options that lie outside the scope of this assessment.[154]

Extensive literature searches were conducted in an attempt to maximise 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 which favours treatment. This is not the case for test accuracy studies, which measure agreement between index test and reference standard. It would seem likely that studies finding greater agreement (high estimates of sensitivity and specificity) will be published more often, however, 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 randomised controlled trials and are therefore more easily discarded when results appear unfavourable. The extent to which publication bias occurs in studies of test accuracy remains unclear, however, 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] We did not undertake a statistical assessment of publication bias in this review. 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.

Despite our extensive searches, no studies were identified which assessed the diagnostic performance of the ROMA score the Fujirebio Lumipulse G automated CEIA system. We do not consider that it would be 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, we did not identify any 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.

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

Clear inclusion criteria were specified in the protocol for this review, a copy of which is available on-line (https://www.nice.org.uk/guidance/GID-DG10012/documents/final-protocol). The eligibility of studies for inclusion is therefore transparent. In addition, we have provided specific reasons for exclusion for all of the studies which were considered potentially relevant at initial citation screening and were subsequently excluded on assessment of the full publication (Appendix 5). 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 (MW, SD and 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] QUADAS-2 is structured into four key domains covering participant selection, index test, reference standard, and the flow of patients through the study (including 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 3a and are summarised in section 3.2.2. Those studies which reported development of risk scores, in addition to test accuracy data, were also assessed using the PROBAST tool.[37] PROBAST (Prediction model study Risk Of Bias Assessment Tool) 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. 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.

Studies included in our systematic review used a variety of definitions of disease/reference standard positive. In order to facilitate clinically relevant comparisons, we have chosen to group studies according to whether or not they included borderline tumours in their definition of malignancy and whether patients found to have non-ovarian malignancies were included in the analyses or were excluded post-hoc. However, a detailed breakdown of histopathological diagnoses was not always reported (see Appendix 2, Table 36) and hence 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. histological confirmation of disease status was available. In practice, risk scores may be used, in secondary care, to triage patients to surgery or surveillance/conservative management, as well as to guide decisions about where 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 (25/49) were conducted in Europe,[17, 42-46, 48-50, 52, 58, 60, 63-65, 67, 77, 79-82, 84, 87, 98, 99] however, only six were conducted solely in the UK[45, 60, 63, 67, 79, 80] and a further two were multi-national studies that included a UK centre.[17, 42] 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 on 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 to 2,445), it should be noted that the largest data sets were derived from the various phases of the IOTA study and that these tended to dominate the analyses for the ADNEX model and for IOTA simple rules. Only one report per intervention (IOTA simple rules of the ADNEX model) was included for each phase of the IOTA study.

5.2.2 Cost effectiveness

Our cost effectiveness analysis is the most comprehensive to date in terms of cost 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. 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, or any of the further assessments, or, where necessary, were based on a focused literature review, prioritising the key input parameters (e.g. the hazard ratio for SMDT versus 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 Section 4.2.3). 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.

5.3 Uncertainties

5.3.1 Clinical effectiveness

There remain a number of areas of uncertainty in relation to the performance characteristics of risk scores for ovarian cancer in specific patient subgroups; 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 downstream consequences of using the various risk scoring options available to select the most appropriate care pathway for patients with an adnexal mass (management by a general gynaecologist or referral to an SMDT). The limited data available for the ROMA score do not suggest any substantial performance advantage over current practice (RMI 1), particularly where the more inclusive definition of malignancy is used (target condition all malignant tumours including borderline). Consideration of the data from studies that report accuracy estimates for both the whole study population (target condition all malignant tumours including borderline) and for selected populations where participants found to have borderline tumours and/or those with rare ovarian cancers or non-ovarian primaries were excluded, indicates that patients with borderline tumours and those with rare ovarian cancers or non-ovarian primaries may be disproportionately represented amongst those with false negative, low risk ROMA scores. One comparative accuracy ROMA score study, using Abbot ARCHITECT tumour marker assays, reported test performance estimates for the target condition epithelial ovarian cancer, calculated both with and without the inclusion of those with borderline tumours;[100] these data indicated that around half of false negative risk scores were accounted for by patients with borderline tumours, 3/6 (50%) using the ROMA score and 7/13 (54%) using the RMI 1.[100] Similarly a comparative accuracy ROMA score study, using Roche Elecsys tumour marker assays, reported test performance estimates for the whole study population and for a selected population where eight (3%) patients with non-epithelial ovarian cancer and non-ovarian primaries were excluded from the analysis; patients with malignancies other than epithelial ovarian cancer accounted for four (50%) of the false negative results using the ROMA score and three (37.5%) using the RMI 1.[90] The potential to detect non-epithelial ovarian cancers by including other tests (e.g. alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (beta-hCG), as recommended in CG122,[1] for women under 40 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, using Roche 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;[98] these data indicated that patients with borderline tumours and those with non-ovarian primaries accounted for a high proportion, 12/14 (86%), of the false negative risk scores observed.[98] It should be noted that these observations are based on small numbers of patients. Furthermore, although other risk scores (Overa (MIA2G), IOTA simple ultrasound rules and the ADNEX model) appear to offer increased sensitivity, data were not available to explore the distribution of histological diagnoses amongst those patients with false negative low risk classifications. The downstream consequences of a false negative, low risk

classification are likely to differ between patients with different histological cancer types and between those with borderline tumours and those with higher stage malignancies. A more complete exploration of the types of patients who are likely to be misclassified as low risk, 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 section 5.1.1, indicate that both the ADNEX model and IOTA simple ultrasound rules (where 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 patients with malignancy who are falsely classified as low risk. However, this increased sensitivity is accompanied by a decrease in specificity and hence an increase in the numbers of patients with benign disease who would be unnecessarily referred to a specialist gynaecological oncology MTD, relative to that associated with risk scoring using the RMI 1. This trade off can be illustrated using a hypothetical cohort of 1,000 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 patients with malignancy who would not be referred to an SMDT would be 18, 33 and 154, based on the ADNEX model, IOTA 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. To achieve a similar level of sensitivity to that of the ADNEX model or IOTA simple rules, using the RMI 1, would require a very low decision threshold; for the same sample cohort of 1,000 patients an RMI 1 threshold of 25 would result in 16 patients with malignancy who would not be referred to an SMDT and 335 'unnecessary' referrals of women with benign disease.

It should also be noted that the performance of risk scoring tools which include morphological features seen on ultrasound is likely to be affected by the level of skill and experience of ultrasonographers. This is particularly the case where the method of applying the score includes an un-specified element of subjective judgement (e.g. IOTA simple ultrasound rules with expert subjective assessment for inconclusive results). The effect of ultrasonographer experience on measures of test performance was considered in our systematic review, however, we found very little data to inform this question. The majority of the studies of the ADNEX model and IOTA 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 simple ultrasound rules

and found no significant differences in test performance between EFSUMB level 2/3 examiners and EFSUMB level 1 examiners, however, it should be noted that the information value of this study is limited, as all examiners received a half day of practical training in IOTA simple rules before the study. Perhaps more interestingly, two of the studies evaluating IOTA 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';[63] '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 13, section 3.2.4), providing some indication that IOTA 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 and increases in specialist cancer centre workload arising from the use of triage methods with higher sensitivity, introduction of routine transvaginal ultrasound in secondary care assessment).

The ideal method of comparing the downstream resource use and clinical consequences of using the various risk scoring options available would be an RCT comparing treatment pathways and subsequent clinical outcomes following risk scoring by different methods. We did not identify any randomised or non-randomised controlled trials that met the inclusion criteria for our systematic review. A recently published RCT,[156] conducted in asymptomatic post-menopausal women with an incidentally detected adnexal mass on ultrasound, compared to risk assessment protocols based on the RMI 1 and on IOTA 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 simple ultrasound rules protocol had surgery, 18/68 (28.1%), versus 7/68 (10.3%) relative risk 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 to this assessment, since the prevalence of malignancy was much lower (2.7%) in this study population than in patients referred to secondary care for investigation of an adnexal mass, as seen in the studies included in our systematic review (median 29.9% (range 15 to 48.4%). The question of how different risk scoring strategies affect referral rates and subsequent clinical outcomes, in this population, remains outstanding.

5.3.2 Cost effectiveness

The economic analyses emphasise the importance of prioritising the sensitivity of the risk scores above the specificity. The benefits of referring as many as possible or, if possible, all patients with ovarian cancer to the SMDT, outweigh the additional SMDT costs; even the additional SMDT costs related to unnecessary referrals (i.e. for false positives). More specifically, informal analyses using a 100% sensitivity and a 0% specificity indicate that a "refer all to SMDT" strategy, without risk scores, might be cost effective at thresholds of £20,000 and £30,000 per QALY gained. It is however questionable whether such a strategy would be feasible for clinical practice, considering, amongst others, the potentially limited SMDT capacity. Particularly since clinician opinion indicated that the real impact of false positives is probably not the additional cost but the time/resources it takes away from true positives patients. 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 confirmed diagnosis. Adding this might for instance favor the IOTA Simple Rules (given it is ultrasound only) and/or ROMA (if ultrasound has already been done in primary care). However, no evidence was found to inform this potential difference in time between the strategies neither was evidence found 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 from the ADNEX model (as the test can be used without CA125) and 2) adding the cost of CA125 to IOTA 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. Also, the handling of patients with malignancies other than ovarian cancer by assuming all have colorectal cancer 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 where patients 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 hazard ratios for SMDT versus no SMDT referral. This hazard ratio was obtained from a Cochrane review.[132] Although this Cochrane review 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). Additionally, it is unclear whether this hazard ratio is representative of the difference between SMDT and no SMDT referral in the UK; also because this hazard ratio might be country specific due to differences between health systems. It is however, reassuring, that the review only included studies performed in developed countries (Canada, Netherlands, UK, and US). Nevertheless, given the above considerations, this hazard ratio should be considered an area of uncertainty. Furthermore, scenario analyses indicated that the SMDT (surgery) costs as well as a potential disutility for FP (favouring risk scores with higher specificity and hence informing the trade-off between sensitivity and specificity) are areas of uncertainty. With regards to the SMDT costs, feedback from clinicians highlighted that SMDT costs used in the model (obtained from NHS reference costs) were likely an under-estimate and did not appropriately reflect the high costs associated with extensive surgery, which is performed in a proportion of patients undergoing surgery in this setting. It should therefore be borne in mind that RMI (threshold 250) was cost effective at a threshold of £20,000 per QALY gained while IOTA simple ultrasound rules was cost effective at a threshold of £30,000 per QALY gained in the scenario were higher SMDT costs were used. Unfortunately, there was no evidence to inform this increased cost of surgery.

6. CONCLUSIONS

6.1 Implications for service provision

There is evidence to suggest that using either the ADNEX model or IOTA simple ultrasound rules to assess the risk of malignancy in women with adnexal mass may offer increased sensitivity relative to current practice (the RMI 1 at a decision threshold of 250 or 200), i.e. a higher proportion of those women who have a malignant tumour would be referred to an SMDT. Both the ADNEX model and IOTA 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 an SMDT. 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 ADNEX model or IOTA 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 higher sensitivity than the ROMA score, there were no data to support a direct comparison between Overa (MIA2G) and the RMI.

In the base-case analysis, the IOTA ADNEX model (threshold 10%) was considered 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 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 simple ultrasound rules and the IOTA ADNEX model (threshold 10%); this was 39% and 60% respectively (£20,000 threshold) and 23% and 75% respectively (£30,000 threshold).

For the pre-menopausal and post-menopausal subgroups, the IOTA ADNEX model (threshold 10%) remained cost effective at 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 ADNEX model (threshold 10%), which had the highest sensitivity (96.3%) was considered cost effective.

6.2 Suggested research priorities

Further, 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 to

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 investigation of an adnexal mass.

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 tests.

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.

Studies on the acceptability (likely up-take) of transvaginal ultrasound, for women being assessed in general gynaecology (secondary care) settings, may also be useful.

Further studies are required to explore the distribution of histological diagnoses amongst patients with false negative low risk classifications. A more complete exploration of the types of patients who are likely to be misclassified as low risk, using the various risk scoring options available, as well as an investigation of the downstream clinical consequences for these patients, is required. If one or more of the risk scores evaluated in this assessment is introduced into routine practice, post-implementation 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 randomised controlled trial, where general gynaecology departments are randomised to use different risk scoring methods to inform decisions about referral to an SMDT; outcomes could include rates of referral, staging investigations, surgery in a specialist setting (gynaecological oncologist), post-surgical outcomes, survival measures.

Further diagnostic cohort studies, or subgroup analyses of existing data sets, are needed to fully explore 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 mentioned in section 5.3.2, 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.

Research should also evaluate whether implementation could be better delivered through one-stop clinics, similar to those used to assess post-menopausal bleeding. Such one-stop clinics, where patients 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.

7. 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 NHS Foundation Trust; Dr Jurjees Hasan, Consultant Medical Oncologist, The Christie, Manchester; Dr Michael Weston, Consultant Radiologist, St James's University Hospital, Leeds; Prof. Richard Edmonds, Professor of Gynaecological Oncology, University of Manchester; Dr Sudha Sundar, Senior Lecturer and Consultant in Gynaecological Oncology, City Hospital, Birmingham; Dr Tracie Miles, Clinical Nurse Specialist, Royal United Hospital, Bath.

We are grateful to the following researchers for providing additional information and un-published 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; 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 Sub-group 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 evidence. Bram Ramaekers and Sabine Grimm planned and performed the cost effectiveness analyses and interpreted results. Nigel Armstrong contributed to planning and interpretation of the systematic review and cost effectiveness analyses, acquisition of input data and conducted model peer review. Shelley de Kock devised and performed the literature searches and provided information support to the project. Jos Kleijnen and Manuela Joore provided senior advice and support to the systematic review and cost effectiveness analyses, respectively. All parties were involved in drafting and/or commenting on the report.

8. DATA SHARING STATEMENT

This is a systematic review and therefore all extracted data are included in the report. Further information can be obtained from the corresponding author.

9. REFERENCES

[1] National Collaborating Centre for Cancer. *Ovarian cancer: the recognition and initial management of ovarian cancer [Internet]*. Manchester: NICE, 2011 [accessed 22.9.16] Available from: https://www.nice.org.uk/guidance/cg122/evidence/full-guideline-181688797

[2] Myers ER, Bastian LA, Havrilesky LJ, Kulasingam SL, Terplan MS, Cline KE, et al. Management of adnexal mass. *Evid Rep Technol Assess (Full Rep)* 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(2):87-9.

[4] Office for National Statistics. *Cancer registration statistics, England: 2014. Cancer diagnoses and age-standardised incidence rates for all cancer sites by age, sex and region [Internet]*. Newport: ONS, 2016 [accessed 22.9.16]. 17 p. Available from: http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2014/pdf

[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 [Internet]*, 2015 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/cases_rates_ovary_1.pdf

[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 [Internet]*. London: Cancer Research UK, 2015 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/deaths_crude_ovary_M14.pdf

[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 [Internet]*. London: Cancer Research UK, 2015 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/cases_crude_ovary_1.pdf

[8] Cancer Research UK. *Ovarian cancer (C56): 2014. Proportion of cancers diagnosed at each stage, all ages, England [Internet]*. London: Cancer Research UK, 2016 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/inc_stage_ovarian_0.pdf

[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(6):6127-38.

[10] British Society for Gynaecological Endoscopy, Royal College of Obstetricians & Gynaecologists. *Management of suspected ovarian masses in premenopausal women [Internet]*. London: RCOG, 2011 [accessed 22.9.16] Available from: http://www.ncin.org.uk/view?rid=1740

[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 (R Coll Radiol)* 2016;28(12):760-5.

[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 (Sao Paulo)* 2012;67(5):437-41.

[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(1):40-6.

[14] Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. 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(5):500-5.

[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(6):681-90.

[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(4):424-37.

[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.

[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(1):32-41.

[19] Scottish Intercollegiate Guidelines Network, Healthcare Improvement Scotland. *Management of epithelial ovarian cancer: SIGN135 [Internet]*. Edinburgh: SIGN, 2013 [accessed 12.10.16]. Available from: http://sign.ac.uk/pdf/sign135.pdf

[20] National Institute for Health and Care Excellence. *Suspected cancer: recognition and referral. NICE guideline 12 [Internet]*. Manchester: NICE, 2015 [accessed 4.10.16] Available from: http://nice.org.uk/guidance/ng12

[21] Royal College of Radiologists. *About iRefer [Internet]*. London: Royal College of Radiologists, 2016 [accessed 12.10.16] Available from: https://www.rcr.ac.uk/clinical-radiology/being-consultant/rcr-referral-guidelines/about-irefer

[22] National Institute for Health and Care Excellence. *Guidance on the use of palitaxel in the treatment of ovarian cancer: technology appraisal 55 [Internet]*. London: NICE, 2005 [accessed 12.10.16] Available from: http://nice.org.uk/guidance/ta55

[23] National Institute for Health and Care Excellence. *Bevacizumab in combination with paclitaxel and carboplatin for first-line treatment of advanced ovarian cancer: technology appraisal 284 [Internet]*. London: NICE, 2013 [accessed 12.10.16] Available from: http://nice.org.uk/guidance/ta284

[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.

[25] National Institute for Health and Care Excellence. *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 381*

[Internet]. London: NICE, 2016 [accessed 12.10.16] Available from: http://nice.org.uk/guidance/ta381

[26] National Institute for Health and Care Excellence. *Topotecan, peylated liposomal doxorubicin hydrochloride, paclitaxel, trabecteden and gemcitabine for treatment recurrent ovarian cancer: technology appraisal 389 [Internet]*. London: NICE, 2016 [accessed 12.10.16] Available from: http://nice.org.uk/guidance/ta389

[27] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 23.3.11] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[28] National Institute for Health and Care Excellence. *Diagnostics assessment programme manual [Internet]*. Manchester: NICE, 2011 [accessed 4.10.16] Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf

[29] Cochrane Diagnostic Test Accuracy Working Group. *Handbook for DTA Reviews [Internet]*: The Cochrane Collaboration, 2009 [accessed 23.3.11] Available from: http://srdta.cochrane.org/handbook-dta-reviews

[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(6):602-7.

[31] Canadian Agency for Drugs and Technologies in Health. *CADTH peer review checklist for search strategies [Internet]*. Ottawa: CADTH, 2013 [accessed 17.7.13]. 3p. Available from: http://www.cadth.ca/en/resources/finding-evidence-is

[32] Wright K, McDaid C. Is the retraction of journal articles in electronic journals and databases consistent and timely? A case study [Poster]. *19th Cochrane Colloquium*. Madrid, 19–22 October 2011.

[33] Wright K, McDaid C. Reporting of article retractions in bibliographic databases and online journals. *J Med Libr Assoc* 2011;99(2):164-7.

[34] Royle P, Waugh N. Should systematic reviews include searches for published errata? *Health Info Libr J* 2004;21(1):14-20.

[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) [Oral presentation]. *19th Cochrane Colloquium.* Madrid, 19-22 October 2011.

[36] Whiting PF, Rutjes AWS, 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(8):529-36.

[37] Wolff R, Whiting P, Mallett S, Riley R, Westwood M, Kleijnen J. PROBAST: Prediction model risk of bias assessment tool. [Oral presentation] *Evidence Synthesis Network: Systematic reviews of prognostic studies: New approaches to prognostic reviews and qualitative evidence synthesis*. Manchester, 27 May 2014.

[38] Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PMM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58(10):982-90.

[39] Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;61(11):1095-103.

[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(2):239-51.

[41] Zamora J, Abraira V, Nuriel A, Khan KS, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;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(3):490-5.

[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(10):557-564.

[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(5):542-8.

[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(9):3835-8.

[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(53):145-54.

[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(5):516-23.

[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(4):680-8.

[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(13):5123-6.

[52] Alcazar JL, Pascual MA, Olartecoechea B, Graupera B, Auba M, Ajossa S, et al. IOTA simple rules for discriminating between benign and malignant adnexal masses: prospective external validation. *Ultrasound Obstet Gynecol* 2013;42(4):467-71.

[53] Weinberger V, Minar L. Diagnostics of malign ovarian tumors by ultrosound and CA 125-our experience. *Int J Gynecol Cancer* 2013;Conference End: 20131022. Conference Publication:(var.pagings). 23 (8 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;Conference End: 20131022. Conference Publication:(var.pagings). 23 (8 SUPPL. 1):155-156.

[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.

[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;Conference End: 20130626. Conference Publication: (var.pagings). 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(9):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, Leone F, Savelli L, et al. Lesion size affects the diagnostic performance of the international ovarian tumor analysis (IOTA) logistic regression models, the IOTA simple rules and the risk of malignancy index to estimate the risk of malignancy in adnexal masses. *Int J Gynaecol Obstet* 2012;Conference End: 20121012. Conference Publication: (var.pagings). 119:S741.

[62] 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(3):345-54.

[63] 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(12):2448-54.

[64] Fathallah K, Huchon C, Bats AS, Metzger U, Lefrere-Belda MA, Bensaid C, et al. [External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors]. *Gynecol Obstet Fertil* 2011;39(9):477-81.

[65] Ruiz de Gauna B, Rodriguez D, Olartecoechea B, Auba M, Jurado M, Gomez Roig MD, et al. 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;191:10-4.

[66] 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.

[67] Baker C, Pasipanodya T, Dwivedi R. The management of suspected ovarian masses in premenopausal women in a DGH setting. *BJOG* 2013;Conference End: 20130626. Conference Publication: (var.pagings). 120:372.

[68] 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(5):582-91.

[69] 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(15 Suppl.):5561.

[70] 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(Suppl.):A5556.

[71] 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(1):82.e1-11.

[72] Akturk E, Karaca RE, Alanbay I, Dede M, Karasahin E, Yenen MC, et al. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. *Journal of Gynecologic Oncology* 2011;22(3):177-82.

[73] 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(2):163-7.

[74] 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]. *Chung Kuo I Hsueh Ko Hsueh Yuan Hsueh Pao* 2010;32(3):297-302.

[75] 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(3):186-91.

[76] Manjunath AP, Pratapkumar, Sujatha K, Vani R. Comparison of three risk of malignancy indices in evaluation of pelvic masses. *Gynecol Oncol* 2001;81(2):225-9.

[77] Tingulstad S, Hagen B, Skjeldestad FE, Onsrud M, Kiserud T, Halvorsen T, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the

pre-operative diagnosis of pelvic masses. *British Journal of Obstetrics & Gynaecology* 1996;103(8):826-31.

[78] Asif N, Sattar A, Dawood MM, Rafi T, Aamir M, Anwar M. Pre-operative evaluation of ovarian mass: risk of malignancy index. *JCPSP, Journal of the College of Physicians & Surgeons - Pakistan* 2004;14(3):128-31.

[79] 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. *British Journal of Obstetrics & Gynaecology* 1990;97(10):922-9.

[80] Davies AP, Jacobs I, Woolas R, Fish A, Oram D. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. *British Journal of Obstetrics & Gynaecology* 1993;100(10):927-31.

[81] 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. *British Journal of Obstetrics & Gynaecology* 1999;106(6):524-7.

[82] Presl J, Kucera R, Topolcan O, Novotny Z, Vrzalova J, Fuchsova R, et al. [HE4 a biomarker of ovarian cancer]. *Ceska Gynekol* 2012;77(5):445-9.

[83] 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(2):239-44.

[84] Karlsen MA, Sandhu N, Hogdall 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(2):379-83.

[85] 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.

[86] 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(5):863-70.

[87] 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(9):4137-40.

[88] 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;Conference End: 20110309. Conference Publication: (var.pagings). 120:S68.

[89] 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.

[90] 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.

[91] 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.

[92] 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(2):377-83.

[93] 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;Conference End: 20130626. Conference Publication: (var.pagings). 120:360-361.

[94] 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.

[95] 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.

[96] 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(1-2):32-40.

[97] 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.

[98] Janas L, Glowacka E, Wilczynski JR, Malinowski A, Nowak M. [Evaluation of applicability of HE4 and ROMA in the preoperative diagnosis of adnexal masses]. *Ginekol Pol* 2015;86(3):193-7.

[99] 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(11):1649-56.

[100] 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(5):1949-53.

[101] 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(1):89-97.

[102] 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(2 Pt 1):280-8.

[103] 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(52):e2402.

[104] Al Musalhi K, Al Kindi M, Al Aisary F, Ramadhan F, Al Rawahi T, Al Hatali K, et al. 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(5):336-44.

[105] 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. In: Annual Meeting of American College of Obstetricians and Gynecologists; 2016 May 14-17; Washington DC, 2016.

[106] 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 (WHO) (date unknown) [accessed 9.3.17]. Available from: http://isrctn.com/ISRCTN45238573 Identifier: ISRCTN45238573

[107] 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(8):p21.

[108] Ali M, Jabeen R, Khurshid M. Preoperative diagnosis of malignancy in ovarian masses: a comparison of risk of malignancy indices I and II. *Medical Forum Monthly* 2010;21(11):19-23.

[109] 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 Medical Journal* 2010;10(4):286-9.

[110] Shahzad N, Rashid N, Zahra S, Malik A. Role of malignancy index in prediction of malignancy in ovarian masses preoperative. *Medical Forum Monthly* 2015;26(6):44-7.

[111] 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(4):421-4.

[112] 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]. *Klinicka Biochemie a Metabolismus* 2012;20(4):248-51.

[113] Woman's Health University Hospital, Egypt. IOTA versus pattern recognition method in diagnosis of ovarian masses. In: ClinicalTrials.gov [Internet]. 2000- [cited 9.3.17]. Available from: https://ClinicalTrials.gov/show/NCT02800031 NLM Identifier: NCT02800031

[114] 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.

[115] Sundar S. *Refining Ovarian Cancer Test Accuracy Socres: a test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer protocol. ROCkeTS Study. In: NIHR [Internet]*, (No date) [accessed 9.3.17] Available from: http://www.nets.nihr.ac.uk/__data/assets/pdf_file/0007/130669/PRO-13-13-01.pdf Identifier: ISRCTN17160843

[116] 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.

[117] 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;Conference End: 20131022. Conference Publication:(var.pagings). 23 (8 SUPPL. 1):905.

[118] Keogh F, Tan AL, Eva LJ. HE4 as a tumour marker for the prediction of ovarian carcinoma. *BJOG* 2015;Conference End: 20150415. Conference Publication: (var.pagings). 122:137-138.

[119] 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;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 SUPPL. 1):528-529.

[120] 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;Conference End: 20150625. Conference Publication: (var.pagings). 53:S424.

[121] 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(3):448-52.

[122] Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford: Oxford University Press, 2005.

[123] Ding Y, Hay J. Cost-effectiveness analysis of multimodal screening for ovarian cancer. *Value Health* 2010;Conference End: 20100519. Conference Publication:(var.pagings). 13 (3):A37.

[124] 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(2):321-9.

[125] Havrilesky LJ, Dinan M, Sfakianos GP, Curtis LH, Barnett JC, Van Gorp T, et al. Costs, effectiveness, and workload impact of management strategies for women with an adnexal mass. *J Natl Cancer Inst* 2015;107(1):no pagination.

[126] Drescher CW, Hawley S, Thorpe JD, Marticke S, McIntosh M, Gambhir SS, et al. Impact of screening test performance and cost on mortality reduction and cost-effectiveness of multimodal ovarian cancer screening. *Cancer Prev Res (Phila)* 2012;5(8):1015-24.

[127] 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(1):200.

[128] Cancer Research UK. *Ovarian cancer incidence statistics [Internet]*. London: Cancer Research UK, 2017 [accessed 13.4.17] Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One

[129] 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(6):1165-71.

[130] 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(9332):505-15.

[131] 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: KSR, 2016

[132] Woo YL, Kyrgiou M, Bryant A, Everett T, Dickinson HO. Centralisation of services for gynaecological cancer. *Cochrane Database Syst Rev* 2012, Issue 3. Art. No.: CD007945. DOI: 10.1002/14651858.CD007945.pub2

[133] Colombo N, Peiretti M, Castiglione M. Non-epithelial ovarian cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol* 2009;20 Suppl 4:24-6.

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

[135] National Cancer Registration and Analysis Service. *Colorectal cancer survival by stage - NCIN data briefing [Internet]*. London: NCIN, 2010 [accessed 13.4.17] Available from: http://www.ncin.org.uk/publications/data_briefings/colorectal_cancer_survival_by_stage

[136] Venesmaa P, Ylikorkala O. Morbidity and mortality associated with primary and repeat operations for ovarian cancer. *Obstet Gynecol* 1992;79(2):168-72.

[137] Gerestein CG, Damhuis RAM, Burger CW, Kooi GS. Postoperative mortality after primary cytoreductive surgery for advanced stage epithelial ovarian cancer: A systematic review. *Gynecol Oncol* 2009;114(3):523-527.

[138] Loft A, Andersen TF, Bronnum-Hansen H, Roepstorff C, Madsen M. Early postoperative mortality following hysterectomy. A Danish population based study, 1977-1981. *Br J Obstet Gynaecol* 1991;98(2):147-54.

[139] 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(5):677-84.

[140] Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;13(5):509-18.

[141] Havrilesky LJ, Broadwater G, Davis DM, Nolte KC, Barnett JC, Myers ER, et al. Determination of quality of life-related utilities for health states relevant to ovarian cancer diagnosis and treatment. *Gynecol Oncol* 2009;113(2):216-20.

[142] 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(3):979-85.

[143] Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. *Am J Gastroenterol* 1999;94(6):1650-7.

[144] Department of Health. *NHS reference costs 2015-2016 [Internet]*. London: Department of Health, 2016 [accessed 29.11.16] Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016

[145] Personal Social Services Research Unit. *Unit Costs of Health and Social Care 2016 [Internet]*. Canterbury: University of Kent, 2016 [accessed 29.11.16] Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2016/index.php

[146] Joint Formulary Committee. *British National Formulary. November 2016*. London: BMJ Group and Pharmaceutical Press, 2016 [accessed 29.11.16].

[147] 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(2):199-203.

[148] 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(4):431-9.

[149] 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(8):928-36.

[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.

[151] Dayyani F, Uhlig S, Colson B, Simon K, Rolny V, Morgenstern D, et al. 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(9):1586-93.

[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.

[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(5):503-14.

[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, Issue 12. Art. No.: CD011964. DOI: 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(9):882-93.

[156] Nunes N, Ambler G, Foo X, Naftalin J, Derdelis G, Widschwendter M, et al. 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.

APPENDIX 1: LITERATURE SEARCH STRATEGIES

Clinical Effectiveness searches

MEDLINE (Ovid): 1946 to November Week 2, 2016

Searched: 24.11.16

Records found: 644

1 exp Ovarian Neoplasms/ (79417)

2 Fallopian Tube Neoplasms/ (2722)

3 Uterine Neoplasms/ (40421)

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. (104209)

6 or/1-5 (151393)

7 Peritoneal Neoplasms/ (13589)

8 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (337303)

9 or/7-8 (348068)

10 ovar$.ti,ab,ot. (225015)

11 9 and 10 (23572)

12 6 or 11 (155473)

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. (91426)

15 exp Menopause/ (55959)

16 14 or 15 (102167)

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. (315798)

18 Ultrasonography/ (67496)

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 (335099)
32 30 and 31 (38)
33 (adnex$ adj8 (model$ or score$ or assess$)).ti,ab,ot. (287)
34 (ova2 or overa).ti,ab,ot. (25)
35 Follicle Stimulating Hormone/ (35545)
36 (Follicle stimulat$ 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. (49152)
37 35 or 36 (49152)
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/ (17226)
42 (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1).ti,ab,ot,rn. (34367)
43 41 or 42 (34367)
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 (Ovid): to 22 November 2016
MEDLINE Daily Update (Ovid): to 22 November 2016
MEDLINE Epub Ahead of Print (Ovid): to 23 November 2016.
Searched: 24.11.16
Records found: 83

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. (25858)
9 or/7-8 (25858)

10 ovar$.ti,ab,ot. (18312)
11 9 and 10 (2204)
12 6 or 11 (10115)
13 ((risk adj4 malignan$ 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$).ti,ab,ot. (41009)
18 Ultrasonography/ (0)
19 17 or 18 (41009)
20 (CA125$ 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$)).ti,ab,ot. (144)
26 (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (185)
27 16 and 22 and 26 (22)
28 25 or 27 (149)
29 (IOTA or international ovarian tumo?r analysis).ti,ab,ot. (93)
30 ((Simple adj3 rules) or (simple adj3 descriptors) or SRrisk or b-rules or m-rules).ti,ab,ot. (347)
31 19 or 29 (41084)
32 30 and 31 (11)
33 (adnex$ 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 (Ovid): 1974 to 23 November 2016
Searched: 24.11.16
Records found: 1185

1 exp ovary cancer/ (97370)
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$)).ti,ab,ot. (136192)
6 peritoneum cancer/ (3891)
7 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (405888)
8 or/6-7 (408204)
9 ovar$.ti,ab,ot. (278995)
10 8 and 9 (30110)
11 1 or 2 or 3 or 4 or 5 or 10 (168296)
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/ (136468)
16 (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. or ultrasound/ or sonography/ (591945)
17 (CA125$ 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. (11962)
18 CA 125 antigen/ (13650)
19 17 or 18 (16956)
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 (592674)
32 30 and 31 (66)
33 (adnex$ adj8 (model$ or score$ or assess$)).ti,ab,ot. (466)
34 (ova2 or overa).ti,ab,ot. (78)
35 follitropin/ (56500)
36 (Follicle stimulat$ 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. (67340)
37 or/35-36 (67499)
38 apolipoprotein A1/ (16294)
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 (18737)
41 transferrin/ (27791)
42 (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1).ti,ab,ot,rn. (42260)
43 or/41-42 (42344)
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 (CDSR) (Wiley): Issue 11 of 12, November 2016
Database of Abstracts of Reviews of Effect (DARE) (Wiley): Issue 2 of 4, April 2015
Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): Issue 10 of 12, October 2016
Health Technology Assessment Database (HTA) (Wiley): Issue 4 of 4, October 2016
Searched: 24.11.16
Records found: 43
CDSR: 1
DARE: 5
CENTRAL: 37
HTA: 0
#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 18330

#15 MeSH descriptor: [Menopause] explode all trees 6396

#16 #14 or #15 18330

#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 21215

#18 MeSH descriptor: [Ultrasonography] this term only 956

#19 #17 or #18 21215

#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 21234

#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 (INAHTA) Publications (Internet) http://www.inahta.org/publications/
Searched 25.11.16
Records found: 0
(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)

NIHR Health Technology Assessment Journals Library (Internet)
https://www.journalslibrary.nihr.ac.uk/#/
Searched 25.11.16
Records found: 23

Search terms	Journal reports	Research Projects
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 (ARIF) database (Internet)
http://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/index.aspx
Searched 25.11.16
Records found: 25
Advanced Search
All published libraries

Search terms	Results
Abstract: ovarian AND Abstract: rmi	18

Abstract: ovarian AND Abstract: malignancy index	2
Abstract: ovarian AND Abstract: malignancy indices	0
Abstract: ovarian AND Abstract: ROMA	3
Abstract: ovarian AND Abstract: malignancy algorithm	0
Abstract: ovarian AND Abstract: IOTA	1
Abstract: International ovarian tumor analysis	1
Abstract: ovarian AND Abstract: simple rule	1
Abstract: SRrisk OR Abstract: b-rule OR Abstract: m-rule	0
Abstract: ovarian AND Abstract: adnex	4
Abstract: OVA2 OR Abstract: HE4 OR Abstract: human epididymis protein 4 OR Abstract: human epididymal protein 4	2
Abstract: WAP4 OR Abstract: WAP 4 OR Abstract: WAP four OR Abstract: WAP5	0
Abstract: CA125 OR Abstract: CA-125 OR Abstract: CA 12-5 OR Abstract: antigen 125	4
Total	**36**
Total after removal of duplicates	**25**

PROSPERO (Internet) http://www.crd.york.ac.uk/prospero/

Searched 25.11.16

Records found: 4

#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 (Internet) http://clinicaltrials.gov/ct2/search/advanced
Searched 24.11.16
Records found: 269
Expert search option
(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)

EU Clinical Trials Register (Internet) https://www.clinicaltrialsregister.eu/ctr-search/search
Searched 25.11.16
Records found: 122
(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)

WHO International Clinical Trials Register Portfolio (ICTRP) (Internet)
http://apps.who.int/trialsearch/
Searched 24.11.16
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 Intervention: risk of malignancy index	(2 records for) **2** trials found
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 Title: risk of malignancy index	(5 records for) **5** trials found
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 Intervention: Risk of Ovarian Malignancy Algorithm	(0 records for) **0** trials found
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 Title: Risk of Ovarian Malignancy Algorithm	(1 records for) **1** trial found
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 Intervention: IOTA OR International ovarian tumor analysis OR simple ultrasound rule OR simple rule OR SRrisk OR b-rule OR m-rule	(1 records for) **1** trial found
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	(6 records for) **6** trials found

uterus OR tubal OR peritoneal OR peritoneum 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 Intervention: HE4 OR human epididymis protein 4 OR human epididymal protein 4	(4 records for) **4** trials found
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 Title: HE4 OR human epididymis protein 4 OR human epididymal protein 4	(10 records for) **10** trials found
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 Intervention: CA125 OR CA-125	(23 records for) **20** trials found
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 Title: CA125 OR CA-125	(27 records for) **21** trials found
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**

NB. 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 indete**rmi**nate, hyperthe**rmi**c, Metfo**rmi**n, etc. While 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 (Internet) http://www.rsna.org/Past_Meetings.aspx
Plenary Sessions, Science Sessions
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 dup 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 dup 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 dup removed
Sertoli-leydig	0
Fallopian	1
Oviduct	0
Uterine	16/17 dup removed
Uterus	0
Tub*	0/2 dup removed
Total	**22**

American Society of Clinical Oncology annual conference (Internet)

http://meetinglibrary.asco.org/abstracts

Searched: 2.2.17

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 (Internet)

Searched: 2.2.17

Records found: 108

2016 https://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 http://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 https://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 (Internet)

Searched: 2.2.17

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
Searched: 2.2.17
Records found: 25

2016 – Scientific Sessions and Clinical Trials in Radiology
https://www.myesr.org/congress/about-ecr/past-congresses/ecr-2016

Text words	Hits
Ovar*	5

2015 – Scientific Sessions and Late-Breaking Clinical Trials
https://www.myesr.org/congress/about-ecr/past-congresses/ecr-2015

Text words	Hits
Ovar*	13

2014 – Scientific sessions
https://www.myesr.org/congress/about-ecr/past-congresses/ecr-2014

Text words	Hits
Ovar*	7

Cost Effectiveness searches
MEDLINE (Ovid): 1946 to November Week 2, 2016
Searched: 23.11.16
Records found: 370

1 exp Ovarian Neoplasms/ (79388)
2 Fallopian Tube Neoplasms/ (2721)
3 Uterine Neoplasms/ (40416)
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$)).ti,ab,ot. (104167)
6 or/1-5 (151338)
7 Peritoneal Neoplasms/ (13578)
8 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (337128)
9 or/7-8 (347883)
10 ovar$.ti,ab,ot. (224941)
11 9 and 10 (23562)
12 6 or 11 (155417)

13 (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. (315645)
14 Ultrasonography/ (67487)
15 13 or 14 (333646)
16 (CA125$ 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/ (118213)
21 (Tumo?r marker$ or biomarker$ or bio-marker$ or cancer marker$ or neoplasm marker$).ti,ab,ot. (159924)
22 20 or 21 (246574)
23 15 or 18 or 19 or 22 (580167)
24 12 and 23 (20146)
25 economics/ (28593)
26 exp "costs and cost analysis"/ (216876)
27 economics, dental/ (1917)
28 exp "economics, hospital"/ (23025)
29 economics, medical/ (9388)
30 economics, nursing/ (4000)
31 economics, pharmaceutical/ (2804)
32 (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. (541112)
33 (expenditure$ not energy).ti,ab. (21940)
34 (value adj1 money).ti,ab. (29)
35 budget$.ti,ab. (20740)
36 or/25-35 (682801)
37 ((energy or oxygen) adj cost).ti,ab. (3151)
38 (metabolic adj cost).ti,ab. (1034)
39 ((energy or oxygen) adj expenditure).ti,ab. (20576)
40 or/37-39 (23929)
41 36 not 40 (677656)
42 letter.pt. (943994)
43 editorial.pt. (416892)
44 historical article.pt. (507294)
45 or/42-44 (1844260)
46 41 not 45 (644021)
47 24 and 46 (370)

Economics terms based on Costs filter:

Centre for Reviews and Dissemination. Search strategies: NHS EED MEDLINE using OvidSP (economics filter) [Internet]. York: Centre for Reviews and Dissemination; 2014 [accessed 2.6.14]. Available from: http://www.crd.york.ac.uk/crdweb/searchstrategies.asp[157]

MEDLINE In-Process & Other Non-Indexed Citations (Ovid): to 22 November 2016
MEDLINE Daily Update (Ovid): to 22 November 2016
MEDLINE Epub Ahead of Print (Ovid): to 23 November 2016.
Date searched: 24.11.16
Records found: 31

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$)).ti,ab,ot. (9696)
6 or/1-5 (9801)
7 Peritoneal Neoplasms/ (0)
8 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (25858)
9 or/7-8 (25858)
10 ovar$.ti,ab,ot. (18312)
11 9 and 10 (2204)
12 6 or 11 (10115)
13 (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. (41009)
14 Ultrasonography/ (0)
15 13 or 14 (41009)
16 (CA125$ 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. (32099)
22 20 or 21 (32099)
23 15 or 18 or 19 or 22 (73106)
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$).ti,ab. (91939)
33 (expenditure$ not energy).ti,ab. (2877)
34 (value adj1 money).ti,ab. (5)
35 budget$.ti,ab. (3723)
36 or/25-35 (95781)
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 (95025)
42 letter.pt. (38204)
43 editorial.pt. (27650)
44 historical article.pt. (0)
45 or/42-44 (65854)
46 41 not 45 (94317)
47 24 and 46 (31)

Economics terms based on Costs filter:
Centre for Reviews and Dissemination. Search strategies: NHS EED MEDLINE using OvidSP (economics filter) [Internet]. York: Centre for Reviews and Dissemination; 2014 [accessed 2.6.14]. Available from: http://www.crd.york.ac.uk/crdweb/searchstrategies.asp[157]

EMBASE (Ovid): 1974 to 23 November 2016
Searched: 24.11.16
Records found: 665
1 exp ovary cancer/ (97370)
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. (136192)
6 peritoneum cancer/ (3891)
7 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (405888)

8 or/6-7 (408204)
9 ovar$.ti,ab,ot. (278995)
10 8 and 9 (30110)
11 1 or 2 or 3 or 4 or 5 or 10 (168296)
12 (Ultraso$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph$ or doptone$ or echograph$ or echogram$ or echosound$).ti,ab,ot. or ultrasound/ or sonography/ (591945)
13 (CA125$ 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. (11962)
14 CA 125 antigen/ (13650)
15 13 or 14 (16956)
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/ (62368)
20 (Tumo?r marker$ or biomarker$ or bio-marker$ or cancer marker$ or neoplasm marker$).ti,ab,ot. (260389)
21 19 or 20 (294218)
22 12 or 15 or 18 or 21 (887954)
23 11 and 22 (25903)
24 health-economics/ (37185)
25 exp economic-evaluation/ (262667)
26 exp health-care-cost/ (249052)
27 exp pharmacoeconomics/ (184843)
28 or/24-27 (566720)
29 (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. (778313)
30 (expenditure$ not energy).ti,ab. (30211)
31 (value adj2 money).ti,ab. (1844)
32 budget$.ti,ab. (29697)
33 or/29-32 (806888)
34 28 or 33 (1107372)
35 letter.pt. (963779)
36 editorial.pt. (523590)
37 note.pt. (663117)
38 or/35-37 (2150486)
39 34 not 38 (1007210)
40 (metabolic adj cost).ti,ab. (1124)
41 ((energy or oxygen) adj cost).ti,ab. (3573)
42 ((energy or oxygen) adj expenditure).ti,ab. (25235)
43 or/40-42 (29019)
44 39 not 43 (1001250)
45 exp animal/ (22704681)

46 exp animal-experiment/ (2060346)
47 nonhuman/ (4993357)
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. (5418904)
49 or/45-48 (24242905)
50 exp human/ (18234517)
51 exp human-experiment/ (393716)
52 50 or 51 (18236045)
53 49 not (49 and 52) (6007829)
54 44 not 53 (926312)
55 23 and 54 (665)

Economics terms based on Costs filter:
Centre for Reviews and Dissemination. Search strategies: NHS EED EMBASE using OvidSP (economics filter) [Internet]. York: Centre for Reviews and Dissemination; 2014 [accessed 4.4.16]. Available from: http://www.crd.york.ac.uk/crdweb/searchstrategies.asp[158]

NHS Economic Evaluation Database (NHS EED) (Wiley): Issue 2 of 4, April 2015
Searched: 24.11.16
Records found: 11
#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 21215
#14 MeSH descriptor: [Ultrasonography] this term only 956
#15 #13 or #14 21215
#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 19067

#22 #20 or #21 19071

#23 #15 or #18 or #19 or #22 40201

#24 #12 and #23 708

EconLit (EBSCO): 1966 to 25 November 2016

Date searched: 25.11.16

Records found: 1

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
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

CEA Registry (Internet) www.cearegistry.org

Searched: 25.11.16

Records found: 2

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 (RePEc) (Internet) http://repec.org/
Searched: 25.11.16
Records found: 6

IDEAS search interface

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine |uterus | 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: 1

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine |uterus | 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: 3

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine |uterus | 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: 0

(ovarian | ovary | ovaries | "high-grade serous" | "low-grade serous | sertoli-leydig cell" | fallopian | oviduct | uterine |uterus | 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: 3

Records retrieved in Total: 7
Records retrieved after duplicates: 6

Key:
| OR
+ AND
" " phrase search

Focused outcomes searches

MEDLINE (Ovid): 1946 to January Week 3, 2017
Searched: 31.1.17
Records found: 205
1 Specialization/ (22763)
2 Surgical Oncology/ (9)
3 ((medical or surg$ or gyn?ecolog$ or physician$) adj1 (speciali$ or oncolog$)).ti,ab,ot. (18188)
4 ((special$ or tertiary) adj5 (hospital$ or care$ or healthcare or centre$ or center$ or facility or facilities)).ti,ab,ot. (102854)
5 (central$ adj5 (hospital$ or care$ or healthcare$ or facility or facilities)).ti,ab,ot. (11444)
6 exp Tertiary Healthcare/ (601)
7 or/1-6 (148856)
8 ((general$ or obstetric$ or secondary or regular) adj1 (care or healthcare or surg$ or gyn?ecolog$)).ti,ab,ot. (30791)
9 exp Secondary Care/ (274)
10 or/8-9 (30863)
11 7 and 10 (3252)
12 exp Gynecologic Surgical Procedures/ (74324)
13 (gyn?ecolog$ adj2 surger$).ti,ab,ot. (5371)
14 or/12-13 (77049)
15 exp Ovarian Neoplasms/ (73422)
16 Fallopian Tube Neoplasms/ (2583)
17 Uterine Neoplasms/ (38852)
18 (AOSCa$ 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$)).ti,ab,ot. (96213)
20 or/15-19 (141283)
21 Peritoneal Neoplasms/ (13029)
22 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (298409)
23 or/21-22 (308780)
24 ovar$.ti,ab,ot. (208481)
25 23 and 24 (21354)
26 20 or 25 (145013)

27 exp Cystadenoma/ (5907)
28 (cystadenoma$ or cystoma$ or cyst$ adenoma$).ti,ab,ot. (5325)
29 Fibroma/ (11012)
30 (Fibroma$ or acrochordon$ or fibroepithelial or fibrous tumo?r$).ti,ab,ot. (12675)
31 exp Teratoma/ (19759)
32 (teratoma$ or dermoid$ or dentigerous cyst$ or dysembryoplastic anomal$ or goiter$ or goitre$ or struma$ or sacrococcygeal fistle$ or teratodermoid cyst$ or teratoid tumo?r$).ti,ab,ot. (34781)
33 or/27-32 (68835)
34 Pelvis/ (20123)
35 exp Adnexa Uteri/ (96410)
36 (pelvi$ or ovar$ or adnexa$).ti,ab,ot. (312267)
37 or/34-36 (357008)
38 33 and 37 (8928)
39 ((pelvi$ or adnexa$ or ovar$) adj6 (mass or masses)).ti,ab,ot. (7720)
40 14 or 26 or 38 or 39 (213422)
41 11 and 40 (205)

MEDLINE Epub Ahead of Print (Ovid): to 30 January 2017
MEDLINE In-Process & Other Non-Indexed Citations: to 30 January 2017
MEDLINE Daily Update: to 30 January 2017
Searched: 31.1.17
Records found: 29

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. (20479)
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 (24851)
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$)).ti,ab,ot. (12027)
20 or/15-19 (12208)
21 Peritoneal Neoplasms/ (41)
22 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (28926)
23 or/21-22 (28956)
24 ovar$.ti,ab,ot. (20892)
25 23 and 24 (2883)
26 20 or 25 (12554)
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$).ti,ab,ot. (33945)
37 or/34-36 (33961)
38 33 and 37 (792)
39 ((pelvi$ or adnexa$ or ovar$) adj6 (mass or masses)).ti,ab,ot. (1272)
40 14 or 26 or 38 or 39 (14145)
41 11 and 40 (29)

EMBASE (Ovid): 1974 to 30 January 2017
Searched: 31.1.17
Records found: 524

1 medical specialist/ (102067)
2 ((medical or surg$ or gyn?ecolog$ or physician$) adj1 (speciali$ or oncolog$)).ti,ab,ot. (40274)
3 ((special$ or tertiary) adj5 (hospital$ or care$ or healthcare or centre$ or center$ or facility or facilities)).ti,ab,ot. (189219)
4 (central$ adj5 (hospital$ or care$ or healthcare$ or facility or facilities)).ti,ab,ot. (18866)
5 exp tertiary healthcare/ (71024)
6 or/1-5 (326375)
7 ((general$ or obstetric$ or secondary or regular) adj1 (care or healthcare or surg$ or gyn?ecolog$)).ti,ab,ot. (47054)
8 exp secondary healthcare/ (4685)
9 or/7-8 (48456)
10 6 and 9 (7200)
11 (gyn?ecolog$ adj2 surger$).ti,ab,ot. (8507)
12 exp gynecologic surgery/ (132958)

13 or/11-12 (135722)
14 exp ovary cancer/ (99193)
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$)).ti,ab,ot. (138404)
19 exp peritoneum cancer/ (13274)
20 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (412063)
21 or/19-20 (422066)
22 ovar$.ti,ab,ot. (283143)
23 21 and 22 (31658)
24 14 or 15 or 16 or 17 or 18 or 23 (171059)
25 cystadenoma/ (7719)
26 (cystadenoma$ or cystoma$ or cyst$ adenoma$).ti,ab,ot. (7325)
27 fibroma/ (11802)
28 (Fibroma$ or acrochordon$ or fibroepithelial or fibrous tumo?r$).ti,ab,ot. (16228)
29 teratoma/ (25747)
30 ovary teratoma/ (2771)
31 (teratoma$ or dermoid$ or dentigerous cyst$ or dysembryoplastic anomal$ or goiter$ or goitre$ or struma$ or sacrococcygeal fistle$ or teratodermoid cyst$ or teratoid tumo?r$).ti,ab,ot. (43888)
32 or/25-30 (58645)
33 pelvis/ (67918)
34 ovary/ (67924)
35 (pelvi$ or ovar$ or adnexa$).ti,ab,ot. (445224)
36 or/33-35 (462883)
37 32 and 36 (10871)
38 ((pelvi$ or adnexa$ or ovar$) adj6 (mass or masses)).ti,ab,ot. (12936)
39 13 or 24 or 37 or 38 (295584)
40 10 and 39 (524)

Cochrane Database of Systematic Reviews (CDSR) (Wiley): Issue 1, January 2017
Database of Abstracts of Reviews of Effect (DARE) (Wiley): Issue 2, April 2015
Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): Issue 11, November 2016
Health Technology Assessment Database (HTA) (Wiley): Issue 4 of 4, October 2016
NHS Economic Evaluation Database (NHS EED) (Wiley): Issue 2 of 4, April 2015
Searched: 31.1.17
Records found: 23
CDSR: 6

DARE: 0

CENTRAL: 17

NHS EED: 0

HTA: 0

#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 12086

#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 10032

#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 17164
#36 #33 or #34 or #35 17379
#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 12832
#40 #10 and #39 23

APPENDIX 2: FULL STUDY DETAILS

Table 34: Study details and baseline participant characteristics

Study Details	Selection criteria	Participant details	Test(s)
Aarenstrup 2012[84] **Country:** Denmark **Funding:** Industry (Abbott provided assay reagents) **Recruitment:** Sep-04 to Jan-10	**Inclusion criteria:** Patients admitted to the gynaecology clinic for surgery due to a pelvic mass or pelvic pains potentially caused by a malignant disease or endometriosis **Exclusion criteria:** Pre-operative known relapse of a previous cancer; active cancer other than ovarian **Study Setting:** Secondary care (general gynaecology) **Point in care pathway at which the test is given:** Following referral to a gynaecology clinic	**All** **N Tested:** 1218 **Age**: Median 51, range (16-90) **% premenopausal : % postmenopausal:** 49 : 51 **Definition of postmenopausal:** Previous hysterectomy and age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** NR **Age**: NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 28.7, range (3-3586) **Median HE4 (pmol/L), range:** 53.4, (19-1426) **Malignant** **N Tested:** NR **Age**: NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 647 , range (10-10000) **Median HE4 (pmol/L), range:** 436, (16-15000)	ROMA Abbott; RMI 1
Abdalla 2013[48] **Country:** Poland **Funding:** NR	**Inclusion criteria:** Patients admitted with adnexal mass **Exclusion criteria:** Ultsrasound examination >90 days before surgery; no CA125 level	**All** **N Tested:** 87 **Age**: Mean 44.5 SD (16.6), range (17-79) **% premenopausal : % postmenopausal:** 60.9 : 39.1 **Definition of postmenopausal:** Amenorrhea ≥1 year and age ≥50 in patients with a history of hysterectomy	IOTA simple ultrasound rules; RMI 1

Study Details	Selection criteria	Participant details	Test(s)
Recruitment Start - End: Jan-11 to Dec-11	**Study setting:** Mixed **Point in care pathway index test is given:** Following referral to the Department of Clinical Obstetrics, Women's Diseases and Gynaecological Oncology, in a University Hospital	**CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	
Aktürk 2011[72] **Country:** Turkey **Funding:** NR **Recruitment Start - End:** Oct-08 to Feb-10	**Inclusion criteria:** Women with pelvic masses scheduled for laparotomy or laparoscopy **Exclusion criteria:** NR **Study setting:** Secondary care (Department of Gynecology and Obstetrics) **Point in care pathway index test is given:** Following referral to secondary care (department of obstetrics and gynaecology)	**<u>Benign</u>** **N Tested:** 80 **Age**: NR **% premenopausal : % postmenopausal:** 80 : 20 **Definition of postmenopausal:** >1 year of amenorrhea or an age of more than 50 years in women who have had a hysterectomy. **CA125 (U/ml):** Mean 28 SD (23.8), range (3-120) **Median HE4 (pmol/L), range:** NR **<u>Malignant</u>** **N Tested:** 20 **Age**: NR **% premenopausal : % postmenopausal:** 45 : 55 **Definition of postmenopausal:** >1 year of amenorrhea or an age of more than 50 years in women who have had a hysterectomy. **CA125 (U/ml):** Mean 329.2 SD (648), range (12-2821) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison
Al Musalhi 2016[104] **Country:** Oman **Funding:** Other (Unfunded) **Recruitment Start - End:**	**Inclusion criteria:** Patients attending a gynaecology department for investigation of an ovarian mass **Exclusion criteria:** NR **Study setting:**	**<u>Benign</u>** **N Tested:** 165 **Age**: Median 33, range (13-80) **% premenopausal : % postmenopausal:** 85 : 15 **Definition of postmenopausal:** Previous hysterectomy and age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 23, range (1-978) **Median HE4 (pmol/L), range:** 43 (18-2677)	ROMA Abbott; RMI 1

Study Details	Selection criteria	Participant details	Test(s)
Mar-14 to Apr-15	Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to a gynaecology department	 **Malignant** **N Tested:** 48 **Age**: Median 55, range (21-83) **% premenopausal : % postmenopausal:** 44 : 56 **Definition of postmenopausal:** Previous hysterectomy and age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 261, range (7-14507) **Median HE4 (pmol/L), range:** 207 (27-5932)	
Alcazar 2013[52] **Country:** Spain **Funding:** NR **Recruitment Start - End:** Jan-11 to Jun-12	**Inclusion criteria:** Patients with an adnexal mass, referred to one of two Spanish university centers (Clinica Universidad de Navarra, Pamplona, Spain and Institut Dexeus, Barcelona, Spain) **Exclusion criteria:** Pregnancy; spontaneous resolution of the mass by the time of a 2–3-month follow-up scan; surgery not performed because of physician's and/or patient's decision at follow-up; surgery performed in another centre **Study setting:** Secondary care (Department of Gynecology and Obstetrics) **Point in care pathway index test is given:** Following referral to secondary care	**All** **N Tested:** 340 **Age**: Mean 42.1 SD (13.2), range (13-79) **% premenopausal : % postmenopausal:** 77.1 : 22.9 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules
Asif 2004[78] **Country:** Pakistan **Funding:** NR	**Inclusion criteria:** Consecutive women admitted to a department of gynecology and obstetrics (Military hospital & Combined Military hospital Rawalpindi) for elective surgical exploration and resection of proven ovarian mass	**Malignant** **N Tested:** 55 **Age**: Mean 45 SD (11). **% premenopausal : % postmenopausal:** 40 : 60 **Definition of postmenopausal:** one year or more of amenorrhoea **CA125 (U/ml):** Mean 1107 SD (NR), range (15-1107)	RMI 1 threshold comparison

Study Details	Selection criteria	Participant details	Test(s)
Recruitment Start - End: Jan-01 to Jan-02	**Exclusion criteria:** NR **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to secondary care	**Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** 45 **Age**: Mean 37 SD (14). **% premenopausal : % postmenopausal:** 75 : 25 **Definition of postmenopausal:** one year or more of amenorrhoea **CA125 (U/ml):** Mean 26.5, range (2-210) **Median HE4 (pmol/L), range:** NR	
Baker 2013[67] **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 index test is given:** Following referral to a District General Hospital	**All** **N Tested:** 48 **Age**: NR **% premenopausal : % postmenopausal:** 100 : 0 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules
Chan 2013[83] **Country:** Hong Kong; Taiwan; Republic of Korea; Japan; Thailand; Philippines **Funding:** Industry (Abbott diagnostics) **Recruitment Start - End:** NR 2009 to NR 2010	**Inclusion criteria:** Consecutive women (>18 years old) diagnosed with an adnexal mass by ultrasound, CT, PET or MRI **Exclusion criteria:** Previous history of ovarian, primary peritoneal or any known malignancy; previous bilateral oophorectomy **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to one of six obstetrics and	**All** **N Tested:** 414 **Age**: NR **% premenopausal : % postmenopausal:** 73.9 : 26.1 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Abbott

Study Details	Selection criteria	Participant details	Test(s)
	gynaecology departments		
Clemente 2015[91] **Country:** Philippines **Funding:** NR **Recruitment Start - End:** Oct-10 to Dec-13	**Inclusion criteria:** Women with an adnexal mass who underwent surgery **Exclusion criteria:** NR **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to a tertiary care hospital (unclear whether referral was to a specialist gynaecological oncology department)	**<u>All</u>** **N Tested:** 62 **Age**: Median Range: (22-79) **% premenopausal : % postmenopausal:** 77.4 : 22.6 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Abbott
Coleman 2016[71] **Country:** USA **Funding:** Industry (Vermilllion Inc.) **Recruitment Start - End:** Aug-10 to Dec-11	**Inclusion criteria:** Women ≥ 18 years with a documented pelvic mass who were scheduled for surgical intervention within three months of imaging, and who agreed to phlebotomy **Exclusion criteria:** Diagnosis of malignancy in the previous five years (except of nonmelanoma skin cancers) or enrollment by a gynecologic oncologist **Study setting:** Secondary care **Point in care pathway index test is given:** Following referral to secondary care	**<u>All</u>** **N Tested:** 493 **Age**: Median 48 Range: (18-87) **% premenopausal : % postmenopausal:** 56.0 : 44.0 **Definition of postmenopausal:** Absence of menses for ≥12 months or age ≥50 years **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	Overa (MIA2G)
Davies 1993[80] **Country:** UK **Funding:**	**Inclusion criteria:** Retrospective review of women admitted consecutively to a gynaecology department for surgical ivestigation of an adnexal mass	**<u>Malignant</u>** **N Tested:** 37 **Age**: NR **% premenopausal : % postmenopausal:** 83.8 : 16.2	RMI 1 threshold comparison

Study Details	Selection criteria	Participant details	Test(s)
NR **Recruitment Start - End:** NR	**Exclusion criteria:** NR **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to secondary care	**Definition of postmenopausal:** one year or more of amenorrhoea or age over 50 years for women who had previously undergone a hysterectomy **CA125 (U/ml):** Median 173, range (5-1405) **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** 87 **Age**: NR **% premenopausal : % postmenopausal:** 53 : 47 **Definition of postmenopausal:** one year or more of amenorrhoea or age over 50 years for women who had previously undergone a hysterectomy **CA125 (U/ml):** Median 18, range (5-760) **Median HE4 (pmol/L), range:** NR	
Di Legge 2012[62] **Country:** Sweden; Belgium; Italy; Poland; UK; Czech Republic; China; Canada **Funding:** Government (Swedish Research Council; Research Foundation of Flanders) **Recruitment Start - End:** NR	**Inclusion criteria:** Patients with adnexal mass recruited from 11 oncology referral centes, 5 general hospitals and 3 referral centres for ultrasonography **Exclusion criteria:** Surgical removal of the mass >120 days after ultrasound; pregnancy; inability to tollerate transvaginal ultrasonography **Study setting:** Mixed **Point in care pathway index test is given:** Following referral to secondary or tertiary care	**Tumour size <4 cm** **N Tested:** 396 **Age**: Median 42, range (15-87) **% premenopausal : % postmenopausal:** 71 : 29 **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 21, range (3-9814) **Median HE4 (pmol/L), range:** NR **Tumour size 4 to 9.9 cm** **N Tested:** 1457 **Age**: 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), range:** NR **Tumour size ≥10 cm** **N Tested:** 592	IOTA simple ultrasound rules; RMI 1

Study Details	Selection criteria	Participant details	Test(s)
		Age: Median 53, range (15-94) **% premenopausal : % postmenopausal:** 45 : 55 **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 58, range (2-40140) **Median HE4 (pmol/L), range:** NR	
Fathallah 2011[64] **Country:** France **Funding:** NR **Recruitment Start - End:** Jan-02 to Dec-05	**Inclusion criteria:** Patients 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 index test is given:** Following referral to secondary or tertiary care	**<u>All</u>** **N Tested:** 109 masses **Age**: Mean 45.5, range (21-76) **% premenopausal : % postmenopausal:** 35.8 : **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Malignant</u>** **N Tested:** NR **Age**: Mean 51, range (22-75) **% premenopausal : % postmenopausal** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** NR **Age**: Mean 41.5, range (21-76) **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules
IOTA 5 2017 [redacted]	[redacted]	[redacted]	[redacted]

Study Details	Selection criteria	Participant details	Test(s)
[redacted]	[redacted]		
Jacobs 1990[79] **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 index test is given:** Following referral to secondary care	**Malignant** **N Tested:** 42 **Age**: Mean 59 SD (11.8) **% premenopausal : % postmenopausal:** 19.5 : 80.5 **Definition of postmenopausal:** one year or more of amenorrhoea or age over 50 years for women who had previously undergone a hysterectomy **CA125 (U/ml):** Median 122 SD (NR). 95% CI: (6.1-3394.8) **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** 101 **Age**: Mean 48.8 SD (14.3) **% premenopausal : % postmenopausal:** 52 : 48 **Definition of postmenopausal:** one year or more of amenorrhoea or age over 50 years for women who had previously undergone a hysterectomy **CA125 (U/ml):** Median 17.5 SD (NR). 95% CI: (4.3-70.2) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison
Janas 2015[98] **Country:** Poland **Funding:** Government (MNISW)	**Inclusion criteria:** Women referred for surgery for adnexal mass **Exclusion criteria:** NR **Study setting:**	**All** **N Tested:** 259 **Age**: NR **% premenopausal : % postmenopausal:** 51 : 49 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Roche

Study Details	Selection criteria	Participant details	Test(s)
Recruitment Start - End: NR 2011 to NR 2014	Mixed **Point in care pathway index test is given:** Following referral to gynaecology or gynaecological oncology clinic		
Joyeux 2016[43] **Country:** France **Funding:** NR **Recruitment Start - End:** Jan-13 to Dec-15	**Inclusion criteria:** Women aged 14 to 100 years, received or referred with an adnexal mass (detected on ultrasound) requiring surgery **Exclusion criteria:** The absence of transvaginal ultrasound, pregnancy, an echographic aspect of functional ovarian cyst and the lack of a CA 125 level **Study setting:** Secondary care (Deptepartment Gynecology and Obstetrics or Gynecological Surgery) **Point in care pathway index test is given:** Following referral to secondary care (department of obstetrics and gynaecology)	**Malignant** **N Tested:** **Age**: Mean 57.5 SD (3.7) **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** Mean 653.6 SD (321) **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** **Age**: Mean 50.3 SD (16) **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** Mean 21.4 SD (34.9) **Median HE4 (pmol/L), range:** NR	ADNEX
Knafel 2015[49] **Country:** Poland **Funding:** NR **Recruitment Start - End:** Jan-11 to Oct-12	**Inclusion criteria:** Patients, age ≥18 years, with an adnexal tumour requiring surgery **Exclusion criteria:** Pregnancy; lack of a histopathology result; surgery performed more than 90 days after diagnosis **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to a university hospital, department	**All** **N Tested:** 226 **Age**: Mean 47 SD (NR) **% premenopausal : % postmenopausal:** 63.3 : 36.7 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules

Study Details	Selection criteria	Participant details	Test(s)
	of oncology and gynaecology		
Langhe 2013[95] **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 index test is given:** Following referral to hospital	**<u>All</u>** **N Tested:** 223 **Age**: Median 56 SD (NR) **% premenopausal : % postmenopausal:** 35 : 65 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range** NR	ROMA Fujirebio
Li 2016[97] **Country:** China **Funding:** Government (Specialised Research Fund for the Doctoral program of Higher Education of China; Science and Technology Department, Guangdong province; Natural Science Foundation, Guangdong province; Science and Technology Department of Guangzhou City) **Recruitment Start - End:** Sep-12 to Apr-14	**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 index test is given:** Following referral to a university hospital	**<u>All</u>** **N Tested:** 917 **Age**: NR **% premenopausal : % postmenopausal:** 81.2 : 18.8 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Malignant</u>** **N Tested:** 190 **Age**: Median 50 SD (NR), range (18-82) **% premenopausal : % postmenopausal:** 56.8 : 43.2 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 727 **Age**: NR **% premenopausal : % postmenopausal:** 87.6 : 12.4 **Definition of postmenopausal:** NR	ROMA Abbott

Study Details	Selection criteria	Participant details	Test(s)
		CA125 (U/ml): NR **Median HE4 (pmol/L), range:** NR	
Lou 2010[74] **Country:** China **Funding:** NR **Recruitment Start - End:** Jun-08 to Dec-08	**Inclusion criteria:** Women with an adnexal mass **Exclusion criteria:** NR **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to an obstetrics and gynaecology department	**<u>All</u>** **N Tested:** 223 **Age**: NR **% premenopausal : % postmenopausal:** 74 : 26 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Malignant</u>** **N Tested:** 61 **Age**: NR **% premenopausal : % postmenopausal:** 47.5 : 52.5 **Definition of postmenopausal:** NR **CA125 (U/ml):** Mean 145.9 SD (NR) **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 162 **Age**: NR **% premenopausal : % postmenopausal:** 84.0 : 16.0 **Definition of postmenopausal:** NR **CA125 (U/ml):** Mean 16.1 SD (NR) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison
Manjunath 2001[76] **Country:** India **Funding:** NR **Recruitment Start - End:** Jan-97 to Aug-99	**Inclusion criteria:** Retrospective study of women admitted for surgical exploration of pelvic masses **Exclusion criteria:** NR **Study setting:** Secondary care (general gynaecology)	**<u>Malignant</u>** **N Tested:** 93 **Age**: NR **% premenopausal : % postmenopausal:** 51 : 48 **Definition of postmenopausal:** one year or more of amenorrhoea or age over 50 years for women who had previously undergone a hysterectomy **CA125 (U/ml):** Mean 1215 SD (3315.8), range (1-24607) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison

Study Details	Selection criteria	Participant details	Test(s)
	Point in care pathway index test is given: Following referral to secondary care	**<u>Benign</u>** **N Tested:** 55 **Age**: NR **% premenopausal : % postmenopausal:** 65 : 34 **Definition of postmenopausal:** one year or more of amenorrhoea or age over 50 years for women who had previously undergone a hysterectomy **CA125 (U/ml):** Mean 27.7 SD (41.9), range (2-250) **Median HE4 (pmol/L), range:** NR	
Meys 2016[44] **Country:** The Netherlands **Funding:** Government (Academic fund of the University of Maastricht) **Recruitment Start - End:** Jul-11 to Jul-15	**Inclusion criteria:** Consecutive patients with adnexal pathology **Exclusion criteria:** No pathology result obtained; pathology result known before the ultrasound scan; pathology >120 days after ultrasound; previous oophorectomy **Study setting:** Secondary **Point in care pathway index test is given:** Following referral to an obstetrics and gynaecology department	**<u>Metastases</u>** **N Tested:** 14 **Age**: Median 64.6 SD (NR), range: (20-87.1) **% premenopausal : % postmenopausal:** 21.4 : 78.6 **Definition of postmenopausal:** Previous hysterectomy, age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 78.6 SD (NR) IQR (27.5-260.8) **Median HE4 (pmol/L), range:** NR **<u>Stage II-IV</u>** **N Tested:** 56 **Age**: Median 67.7 SD (NR), range (32.3-87) **% premenopausal : % postmenopausal:** 12.5 : 87.5 **Definition of postmenopausal:** Previous hysterectomy, age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 456 SD (NR) IQR (170.8-1175) **Median HE4 (pmol/L), range:** NR **<u>Stage I</u>** **N Tested:** 18 **Age**: Median 63.1 SD (NR), range (50.3-68.5) **% premenopausal : % postmenopausal:** 33.3 : 66.7 **Definition of postmenopausal:** Previous hysterectomy, age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 109.5 SD (NR) IQR (16.8-	ADNEX; IOTA simple ultrasound rules; RMI 1

Study Details	Selection criteria	Participant details	Test(s)
		361.5) **Median HE4 (pmol/L), range:** NR **<u>Borderline</u>** **N Tested:** 27 **Age**: Median 50.6 SD (NR), range: (36.9-65.8) **% premenopausal : % postmenopausal:** 55.6 : 44.4 **Definition of postmenopausal:** Previous hysterectomy, age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 61.9 SD (NR) IQR (27.5-295) **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 211 **Age**: Median 53.2 SD (NR), range (16.1-87.2) **% premenopausal : % postmenopausal:** 45.9 : 54 **Definition of postmenopausal:** Previous hysterectomy, age ≥50, or amenorrhrea ≥1 year **CA125 (U/ml):** Median 26 SD (NR) IQR (16.5-27) **Median HE4 (pmol/L), range:** NR	
Moffatt 2016[45] **Country:** UK **Funding:** NR **Recruitment Start - End:** Jan-14 to Sep-15	**Inclusion criteria:** Patients with excised adnexal masses which had been sent for histological analysis **Exclusion criteria:** Ectopic pregnancy; no ultrasound available; no CA125 level **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to secondary care	**<u>All</u>** **N Tested:** 81 **Age**: NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ADNEX
Moore 2011[102]	**Inclusion criteria:**	**<u>All</u>**	ROMA Abbott

Study Details	Selection criteria	Participant details	Test(s)
Country: USA **Funding:** Mixed (Fujirebio and the National Cancer Institute) **Recruitment Start - End:** Oct-09 to Aug-10	Women (≥18 years) presenting to a generalist (general gynaecologist, inturnist, 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 index test is given:** Following referral to a general or specialist hospital	**N Tested:** 472 **Age**: Mean 50.3 SD (NR), range (18-89) **% premenopausal : % postmenopausal:** 54 : 46 **Definition of postmenopausal:** Age ≥55 years or FSH >22 IU/L **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Malignant</u>** **N Tested:** NR **Age**: NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 266.8, range (31.2-13250) **Median HE4 (pmol/L), range:** 366.8, (11.9-1073.9) **<u>Benign</u>** **N Tested:** NR **Age** NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 58.1, range (27.1-403.2) **Median HE4 (pmol/L), range:** 19.9, (3.6-1085.1)	
Morgante 1999[81] **Country:** Italy **Funding:** NR **Recruitment Start - End:** Jan-95 to Dec-97	**Inclusion criteria:** Women over 30 years of age admitted consecutively for surgical excision of ovarian masses **Exclusion criteria:** NR **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:**	**<u>Malignant</u>** **N Tested:** 31 **Age**: NR **% premenopausal : % postmenopausal:** 29 : 68 **Definition of postmenopausal:** >1 year amenorrhea or >50 years old with a hysterectomy **CA125 (U/ml):** Mean 354 SD (NR) IQR (102-290) **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 93	RMI 1 threshold comparison

Study Details	Selection criteria	Participant details	Test(s)
	Following referral to secondary care	**Age**: NR **% premenopausal : % postmenopausal:** 65 : 35 **Definition of postmenopausal:** >1 year amenorrhea or >50 years old with a hysterectomy **CA125 (U/ml):** Mean 29.6 SD (NR) IQR (10-22) **Median HE4 (pmol/L), range:** NR	
Murala 2014[60] **Country:** UK **Funding:** NR **Recruitment Start - End:** Sep-12 to Sep-13	**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 index test is given:** Following referral to secondary care	**All** **N Tested:** 51 **Age**: NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple rules; RMI 1
Novotny 2012[87] **Country:** Czech republic **Funding:** Government (Ministry of Health, 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 index test is given:** Following referral to a gynaecology and obstetrics department	**Malignant** **N Tested:** 21 **Age**: Median 63 SD (NR), range: (47-82) **% premenopausal : % postmenopausal:** 0 : 100 **Definition of postmenopausal:** Age ≥55 years or FSH >22 IU/L **CA125 (U/ml):** Median 295 SD (NR), range (32.8-44850) **Median HE4 (pmol/L), range:** 312, (17.1-1842) **Benign** **N Tested:** 256 **Age**: Median 64 SD (NR), range (48-93) **% premenopausal : % postmenopausal:** 0 : 100 **Definition of postmenopausal:** Age ≥55 years or FSH	ROMA Abbott

Study Details	Selection criteria	Participant details	Test(s)
		>22 IU/L **CA125 (U/ml):** Median 16.2 SD (NR), range (3.6-2331) **Median HE4 (pmol/L), range:** 39.5, (26.7-3590)	
Piovanao 2016[58] **Country:** Italy **Funding:** NR **Recruitment Start - End:** Feb-13 to Jan-15	**Inclusion criteria:** Consecutive women (≥18 years), with an adnexal mass, who were candidiates for surgery **Exclusion criteria:** NR **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to hospital	**All** **N Tested:** 391 **Age**: Median 47 SD (NR), range: (18-86) **% premenopausal : % postmenopausal:** 56.5 : 43.5 **Definition of postmenopausal:** Amenorrhea for at least 12 months or over 50 years of age and hysterectomy before menopause **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules
Presl 2012[82] **Country:** Czech Republic **Funding:** Government **Recruitment Start - End:** Jun-10 to Jan-11	**Inclusion criteria:** Patients with abnormalities in the pelvis **Exclusion criteria:** NR **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to a university hospital department of obstetrics and gynaecology	**All** **N Tested:** 552 **Age**: NR **% premenopausal : % postmenopausal:** 53.6 : 46.4 **Definition of postmenopausal:** FSH≥22 IU/L **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Abbott
Ruiz de Gauna 2015[65] **Country:** Spain	**Inclusion criteria:** Women diagnosed with persistent adnexal mass evaluated in one of two Spanish	**All** **N Tested:** 247 **Age**: Mean 43.6 SD (14.1), range (14-83) **% premenopausal : % postmenopausal:** 72.1 : 27.9	IOTA simple ultrasound rules

Study Details	Selection criteria	Participant details	Test(s)
Funding: NR **Recruitment Start - End:** Jun-12 to Dec-13	**Exclusion criteria:** Pregnant women; masses with spontaneous resolution; masses removed surgically in another center from recruitment **Study setting:** Mixed **Point in care pathway index test is given:** Following referral to secondary care	**Definition of postmenopausal:** **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	
Sayasneh 2013b[63] **Country:** UK **Funding:** Government (NHS; UK NIHR; Imperial College London) **Recruitment Start - End:** Sep-10 to Sep-12	**Inclusion criteria:** Patients with at least one adnexal mass, who underwent transvaginal ultrasound examination at one of the participating centres **Exclusion criteria:** Surgical removal of the mass >120 days after ultrasound; refusal to undergo transvaginal ultrasonography; pregnancy; examined by a consultant with a specialist interest in gynaecological malignancy; cytology rather than histology used to establish diagnosis **Study setting:** Mixed **Point in care pathway index test is given:** Following referral to secondary or tertiary care	**All** **N Tested:** 255 **Age**: Mean 46 SD (NR), 95% CI: (34-57) **% premenopausal : % postmenopausal:** 64.7 : 35.3 **Definition of postmenopausal:** ≥50 years who had undergone hysterectomy **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules; RMI 1
Sayasneh 2016[46] **Country:** UK; Spain **Funding:** Charity (National Institute for Health Research; FWO Grants;	**Inclusion criteria:** Patients presenting ≥ adnexal mass who underwent transvaginal ultrasonography **Exclusion criteria:** Pregnancy, patients examined by a consultant, refusal of transvaginal ultrasonography, cytology rather than	**All** **N Tested:** 610 **Age**: Median 47 **% premenopausal : % postmenopausal:** 58 : 42 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ADNEX

Study Details	Selection criteria	Participant details	Test(s)
KU Leuven Grant) **Recruitment Start - End:** Sep-10 to Feb-15	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 index test is given:** Referral to tertiary care		
Shulman 2016[105] **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 index test is given:** Following referral to secondary care	**<u>All</u>** **N Tested:** 993 **Age**: Mean 50.3 SD (NR), range: (18-92) **% premenopausal : % postmenopausal:** 51 : 49 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	Overa (MIA2G); ROMA Roche
Silvestre 2015[55] **Country:** Brazil **Funding:** NR **Recruitment Start - End:** Sep-08 to Dec-10	**Inclusion criteria:** Women that were consecutively scheduled for surgery to remove adnexal masses **Exclusion criteria:** NR **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to secondary care department of obstetrics and gynaecology	**<u>Malignant</u>** **N Tested:** 32 **Age**: Median 52 SD (NR), range (20-78) **% premenopausal : % postmenopausal:** NR : NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 43 **Age**: Median 42 SD (NR), range (18-82) **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR	IOTA simple ultrasound rules

Study Details	Selection criteria	Participant details	Test(s)
		Median HE4 (pmol/L), range: NR	
Szubert 2016[42] **Country:** Poland; Spain **Funding:** NR **Recruitment Start - End:** Dec-12 to Apr-15	**Inclusion criteria:** Patients requiring surgery due to ovarian tumor, 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 (Dept. Gynecology and Obstetrics or Gynecological Surgery) **Point in care pathway index test is given:** Following referral to secondary care (division of gynaecological surgery or department of obstetrics and gynaecology)	**Poland** **N Tested:** 204 **Age**: Median 46 SD (NR), range (15-84) **% premenopausal : % postmenopausal:** 65.7 : 32.4 **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 40 SD (NR), range (4-4909) **Median HE4 (pmol/L), range:** NR **Spain** **N Tested:** 123 **Age**: 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) **Median HE4 (pmol/L), range:** NR	ADNEX
Tantipalakorn 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:** Apr-07 to Mar-12	**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, such ovarian cancers scheduled for second look operation, or endometrioma diagnosed by previous laparoscopy, etc; patients undergoing surgery beyond 24 hours after ultrasound examination **Study setting:** Secondary care **Point in care pathway index test is given:** Following referral to secondary care	**All** **N Tested:** 319 masses **Age**: Mean 42.4 SD (16.2), range (13-82) **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules

Study Details	Selection criteria	Participant details	Test(s)
Testa 2014[50] **Country:** Sweden; Belgium; Italy; Poland; Spain; Czech Republic **Funding:** Government (Swedish Research Council; UK NIHR) **Recruitment Start - End:** Oct-09 to May-12	**Inclusion criteria:** Patients with at least one adnexal mass (ovarian, para-ovarian or tubal), who underwent transvaginal ultrasound 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; incomplete final histology **Study setting:** Mixed **Point in care pathway index test is given:** Following referral to secondary or tertiary care	**<u>All</u>** **N Tested:** 2403 **Age** NR **% premenopausal : % postmenopausal:** 56.3 : 43.7 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Malignant</u>** **N Tested:** 980 **Age**: Median 57 SD (NR) IQR (46-66) **% premenopausal : % postmenopausal:** 38.6 : 61.4 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range** NR **<u>Benign</u>** **N Tested:** 1423 **Age**: Median 44 SD (NR) IQR (33-56) **% premenopausal : % postmenopausal:** 68.6 : 31.4 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules; RMI 1
Timmerman 2010[66] **Country:** Sweden; Belgium; Italy; Poland; UK; Czech Republic; China; Canada **Funding:** Government (Swedish Research Council; Research Council KU Leuven) **Recruitment Start - End:**	**Inclusion criteria:** Patients with at least one adnexal mass, who underwent transvaginal ultrasound 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 transvaginal ultrasonography; pregnancy **Study setting:** Mixed	**<u>All</u>** **N Tested:** 1938 **Age**: Mean 46 SD (NR), range (11-94) **% premenopausal : % postmenopausal:** 62 : 38 **Definition of postmenopausal:** Age ≥50 or previous hsyterectomy **CA125 (U/ml):** NR **Median HE4 (pmol/L), range** NR	IOTA simple rules; RMI 1

Study Details	Selection criteria	Participant details	Test(s)
NR 2005 to NR 2007	**Point in care pathway index test is given:** Following referral to secondary or tertiary care		
Tingulstad 1996[77] **Country:** Norway **Funding:** NR **Recruitment Start - End:** Feb-92 to Feb-94	**Inclusion criteria:** Patients with a pelvic mass, who were scheduled for laparotomy and who were at least 30 years old **Exclusion criteria:** NR **Study setting:** Secondary care (Dept. Gynecology and Obstetrics) **Point in care pathway index test is given:** Following referral to secondary care	**Malignant** **N Tested:** 56 **Age**: NR **% premenopausal : % postmenopausal:** 20 : 80 **Definition of postmenopausal:** >1 year amenorrhea or >50 years old with a hysterectomy **CA125 (U/ml):** Median 180 SD (NR), range (7-18400) **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** 117 **Age**: NR **% premenopausal : % postmenopausal:** 61 : 39 **Definition of postmenopausal:** >1 year amenorrhea or >50 years old with a hysterectomy **CA125 (U/ml):** Median 12 SD (NR), range (5-538) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison
Tinnangwattana 2015[47] **Country:** Thailand **Funding:** Government (Office of the Highern Education Commission) **Recruitment Start - End:** Mar-14 to Dec-14	**Inclusion criteria:** Patients scheduled for surgery because of an adnexal mass either detected by pelvic examination or previous ultrasound examination **Exclusion criteria:** Known diagnoses; surgery>24 hours after ultrasound examination **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to secondary care department of obstetrics and gynaecology	**All** **N Tested:** 100 **Age**: Mean 44.21 SD (12.9), range (19-75) **% premenopausal : % postmenopausal:** 73 : 27 **Definition of postmenopausal:** Postmenopausal period **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules

Study Details	Selection criteria	Participant details	Test(s)
Ulusoy 2007[75] **Country:** Turkey **Funding:** NR **Recruitment Start - End:** Sep-02 to Nov-04	**Inclusion criteria:** Consecutive women undergoing surgery for an adnexal mass **Exclusion criteria:** Known ovarian malignancy; pregnancy **Study setting:** Mixed **Point in care pathway index test is given:** Following referral to secondary or tertiary care (Dept. Gynecology and Obstetrics and Gynaecological Oncology Clinic)	**Malignant** **N Tested:** 106 **Age**: Mean 47 SD (NR) **% premenopausal : % postmenopausal:** 53.8 : 46.2 **Definition of postmenopausal:** >1 year of amenorrhea or an age of more than 50 years in women who have had a hysterectomy **CA125 (U/ml):** Median 152.75 SD (NR), range (1-5000) **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** 190 **Age**: Mean 42 SD (NR) **% premenopausal : % postmenopausal:** 67.9 : 31.2 **Definition of postmenopausal:** >1 year of amenorrhea or an age of more than 50 years in women who have had a hysterectomy **CA125 (U/ml):** Median 31.42 SD (NR), range (3-1153) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison
Van Calster 2014[17] **Country:** Belgium; Sweden; Italy; Czech Republic; Poland; UK **Funding:** Government (Flemish government: Research Foundation–Flanders (FWO)) **Recruitment Start - End:** NR 1999 to NR 2012	**Inclusion criteria:** Consecutive patients with ≥1 adnexal mass (judged not to be a physiological cyst), examined with transvaginal ultrasound and selected for surgical intervention **Exclusion criteria:** Refusal of transvaginal ultrasonography; pregnancy; surgical removal of the mass >120 days after ultrasound **Study setting:** Mixed (oncology centres, general hospitals and gynaecology units) **Point in care pathway index test is given:**	**All** **N Tested:** 2403 **Age**: NR **% premenopausal : % postmenopausal:** 56.3 : 43.7 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **Comments:** for IOTA phase 3 validation data set **Malignant** **N Tested:** 980 **Age**: Median 57 SD (NR) IQR (46-66) **% premenopausal : % postmenopausal:** 38.6 : 61.4 **Definition of postmenopausal:** NR	ADNEX

Study Details	Selection criteria	Participant details	Test(s)
	Following referral to secondary or tertiary care	**CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 1423 **Age**: Median 44 SD (NR) IQR (33-56) **% premenopausal : % postmenopausal:** 68.6 : 31.4 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	
Van Gorp 2012[99] **Country:** Belgium **Funding:** Government (Belgian Federation against Cancer; Research Foundation Flanders) **Recruitment Start - End:** Aug-05 to Mar-09	**Inclusion criteria:** Patients 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 index test is given:** Following referral to a university hospital for resection of a pelvic mass	**<u>Malignant</u>** **N Tested:** 150 **Age**: Mean 57.7 SD (NR) 95% CI: (55.7-59.8) **% premenopausal : % postmenopausal:** 26 : 74 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 224 **Age**: Mean 46.2 SD (NR) 95% CI: (44.1-48.3) **% premenopausal : % postmenopausal:** 62.1 : 37.9 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Fujirebio; RMI 1
Weinberger 2013[53] **Country:** NR **Funding:** Other (NR) **Recruitment Start - End:** NR 2010 to NR 2012	**Inclusion criteria:** Patients with suspiscious adnexal mass **Exclusion criteria:** NR **Study setting:** Unclear **Point in care pathway index test is given:**	**<u>All</u>** **N Tested:** 347 **Age**: NR **% premenopausal : % postmenopausal:** NR **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	IOTA simple ultrasound rules

Study Details	Selection criteria	Participant details	Test(s)
	Following referral (unclear whether secondary or tertiary care)		
Winarto 2014[100] **Country:** Indonesia **Funding:** NR **Recruitment Start - End:** Nov-10 to May-11	**Inclusion criteria:** Patients diagnosed with ovarian tumour, by physical examination and transvaginal ultrasound **Exclusion criteria:** Un-resectable tumour; non-epithelial histopathological results; history of oophorectomy; ovarian cancer treatment; pregnancy **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to hospital	**Malignant** **N Tested:** **Age**: Mean 44 SD (NR) **% premenopausal : % postmenopausal:** 50 : 25 **Definition of postmenopausal:** NR **CA125 (U/ml):** Median 357.5 SD (NR), range (13.1-9872.3) **Median HE4 (pmol/L), range:** 495.5, (436.3-15000) **Benign** **N Tested:** **Age**: 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) **Median HE4 (pmol/L), range:** 52.3, (29.5-26.1)	ROMA Abbott; RMI 1
Xu 2016[96] **Country:** China **Funding:** Government (Guangdong Natural Science Foundation; Guangdong Province Science and Technology Project Plan and Social Development Foundation; Medical Science and Technology Research Foundation of Guangdong Province) **Recruitment Start - End:**	**Inclusion criteria:** Retrospective study of patients with an ovarian mass **Exclusion criteria:** Missing tumour marker data; patients with non-epithelial ovarian cancer **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to secondary care	**Malignant** **N Tested:** 239 **Age**: Mean 57 SD (NR) **% premenopausal : % postmenopausal:** 54 : 46 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **Borderline** **N Tested:** 45 **Age**: Mean 40 SD (NR) **% premenopausal : % postmenopausal:** 80 : 20 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Roche

Study Details	Selection criteria	Participant details	Test(s)
Jul-13 to Nov-14		**Benign** **N Tested:** 311 **Age**: Mean 42 SD (NR) **% premenopausal : % postmenopausal:** 85 : 15 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	
Yamamoto 2009[73] **Country:** Japan **Funding:** NR **Recruitment Start - End:** Jan-02 to Apr-05	**Inclusion criteria:** Women with a pelvic mass scheduled for laparotomy and laparoscopy at the Department of Obstetrics and Gynecology, Kochi Medical School **Exclusion criteria:** NR **Study setting:** Secondary care (Dept. Gynecology and Obstetrics) **Point in care pathway index test is given:** Following referral to secondary care (department of obstetrics and gynaecology)	**Malignant** **N Tested:** 40 **Age**: NR **% premenopausal : % postmenopausal:** 37.5 : 62.5 **Definition of postmenopausal:** >1 year of amenorrhea or an age of more than 50 years in women who have had a hysterectomy **CA125 (U/ml):** Median 124 SD (NR), range (11.4-4340) **Median HE4 (pmol/L), range:** NR **Benign** **N Tested:** 213 **Age**: NR **% premenopausal : % postmenopausal:** 81.7 : 18.3 **Definition of postmenopausal:** >1 year of amenorrhea or an age of more than 50 years in women who have had a hysterectomy **CA125 (U/ml):** Median 35.2 SD (NR), range (5-616) **Median HE4 (pmol/L), range:** NR	RMI 1 threshold comparison
Yanaranop 2016[90] **Country:** Thailand **Funding:** NR **Recruitment Start - End:**	**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; cancelled	**All** **N Tested:** 260 **Age**: Mean 48.2 SD (14.2) **% premenopausal : % postmenopausal:** 56.9 : 43.1 **Definition of postmenopausal:** Age >55, age <45 with amenorrhrea >1 year, or FSH <25 IU/L **CA125 (U/ml):** NR **Median HE4 (pmol/L),** range NR	ROMA Roche; RMI 1

Study Details	Selection criteria	Participant details	Test(s)
Jan-12 to Dec-12	surgery **Study setting:** Secondary care (general gynaecology) **Point in care pathway index test is given:** Following referral to an obstetrics and gynaecology department	 **Malignant** **N Tested:** 74 **Age**: NR **% premenopausal : % postmenopausal:** 37.8 : 62.2 **Definition of postmenopausal:** Age >55, age <45 with amenorrhrea >1 year, or FSH <25 IU/L **CA125 (U/ml):** Median 274.1 SD (NR) **Median HE4 (pmol/L),** range 165.1, (NR) **Benign** **N Tested:** 186 **Age**: NR **% premenopausal : % postmenopausal:** 64.5 : 35.5 **Definition of postmenopausal:** Age >55, age <45 with amenorrhrea >1 year, or FSH <25 IU/L **CA125 (U/ml):** Median 32.9 SD (NR) IQR (NR) **Median HE4 (pmol/L),** range 57.3, (NR)	
Zhang 2015a[69] **Country:** USA **Funding:** NR **Recruitment Start - End:** NR	**Inclusion criteria:** Patients with a documented pelvic mass scheduled for surgery **Exclusion criteria:** NR **Study setting:** Unclear **Point in care pathway index test is given:** Following referral (unclear whether secondary or tertiary care)	**All** **N Tested:** 305 **Age**: NR **% premenopausal : % postmenopausal:** 50.5 : 49.5 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **Malignant** **N Tested:** 264 **Age**: NR **% premenopausal : % postmenopausal:** 36 : 64 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **Benign**	Overa (MIA2G)

Study Details	Selection criteria	Participant details	Test(s)
		N Tested: 348 **Age**: NR **% premenopausal : % postmenopausal:** 81.9 : 18.1 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	
Zhang 2015b[103] **Country:** China **Funding:** Government (National High Technology Research and Deveopment Program; China Postdoctoral Science Special Foundation; National Science and Technology Infrastructure; National Science Foundation of China) **Recruitment Start - End:** Oct-12 to Feb-13	**Inclusion criteria:** Patients with a pelvic mass, suspected to be of ovarian origin, who were to undego surgery **Exclusion criteria:** Age <18 years; missing clinical examination results; blood sample <0.5 ml, stored or transporte at >0°C, lipemic or haemolytic appearance; pregnancy; family history of ovarian cancer; receiving chemotherapy, radiotherapy and other treatments **Study setting:** Unclear **Point in care pathway index test is given:** Following referral to one of nine centres	**<u>Malignant</u>** **N Tested:** 264 **Age**: NR **% premenopausal : % postmenopausal:** 36 : 64 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR **<u>Benign</u>** **N Tested:** 348 **Age**: NR **% premenopausal : % postmenopausal:** 81.9 : 18.1 **Definition of postmenopausal:** NR **CA125 (U/ml):** NR **Median HE4 (pmol/L), range:** NR	ROMA Roche

Table 35: Index test details
ROMA Abbott

Study Details	ROMA Abbott test details		
	Analyser **Manufacturer of CA125 and HE4 Assays**	**Sample collection** **storage**	**Time from test to surgery**
Aarenstrup 2012[84]	ARCHITECT I 2000sr Abbott	Blood samples were collected within 2 weeks prior to surgery Samples were centrifuged within 6 h of collection. After centrifugation, serum samples were stored at -80°C until analysis	NR
Al Musalhi 2016[104]	ARCHITECT I2000 Abbott	Samples collected using serum separator tubes and centrifuged immediately Serum samples were stored at -20°C	NR
Chan 2013[83]	ARCHITECT Abbott	Blood samples were collected after pelvic mass was confirmed and surgery schedlued, to minimise the time between testimng and surgery. Samples were centrifuged and serum separated within 4 hours of collection Serum samples were stored at -20°C	NR
Chan 2013[83]	ARCHITECT Abbott	Blood samples were collected after pelvic mass was confirmed and surgery schedlued, to minimise the time between testimng and surgery Samples were centrifuged and serum separated within 4 hours of collection Serum samples were stored at -20°C	NR
Clemente 2015[91]	NR Abbott	NR	NR
Li 2016[97]	ARCHITECT Abbott	Blood samples were collected on the day of surgery, before anesthesis. Samples were centrifuged and serum separated Serum samples were stored at -80°C	<1 day
Moore 2011[102]	ARCHITECT i2000 Abbott	Blood samples were collected within 30 days prior to surgery and before induction of anesthesia Samples were collected into a serum separator tube and centrifuged after clotting Serum samples were stored at -20°C	30 days or less

Study Details	ROMA Abbott test details		
	Analyser **Manufacturer of CA125 and HE4 Assays**	**Sample collection** **storage**	**Time from test to surgery**
Novotny 2012[87]	ARCHITECT 1000i Abbott	Blood samples were collected prior to surgery or treatment and centrifuged Serum samples were stored at -80°C	NR
Presl 2012[82]	ARCHITECT 1000 Abbott	Blood samples were centrifuged immediately or within 24 hours of collection Serum samples were stored at -80°C	NR
Winarto 2014[100]	NR Abbott	NA NA	Blood samples collected 1 day before surgery, time from ultrasound to surgery unclear

ROMA ROCHE

Study Details	ROMA Roche test details		
	Analyser **Manufacturer of CA125 and HE4 Assays**	**Sample collection** **storage**	**Time from test to surgery**
Janas 2015[98]	NR Roche	NR	NR
Shulman 2016[105]	NR	NR	NR
Xu 2016[96]	Cobas E170 Roche	Blood samples collected before surgery and centrifuged within 3 h Serum samples were stored at -80°C	NR
Yanaranop 2016[90]	Cobas 6000 Roche	Samples collected within 48 h prior to surgery and centrifuged immediately Serum samples were stored at -20°C	Within 6 weeks before surgery
Zhang 2015b[103]	Cobas 601 Roche	Blood samples collected into a tube containing clot activator and centrifuged Serum samples were stored at -80°C	NR

ROMA Fujirebo

Study Details	ROMA Fujirebo test details		
	Analyser **Manufacturer of CA125 and HE4 Assay**	**Sample collection,** **storage**	**Time from test to surgery**
Langhe 2013[95]	NR Fujirebio	NR	Collected before surgery
Van Gorp 2012[99]	NR Fujirebio	Blood samples were collected intclotting tubes, immediately before surgery After centrifugation, serum samples were stored at -80°C until analysis	Time from ultrasound to surgery not reported

ADNEX

Study Details	ADNEX test details			
	Analyser Manufacturer of CA125 Assay	**Ultrasound details**	**Sample collection storage**	**Time from test to surgery**
IOTA5 2017*	[redacted]	[redacted]	[redacted]	[redacted]
Joyeux 2016[43]	NR	Transvaginal ultrasound could be complemented by another imaging technique (abdominal ultrasound, CT scan, MRI)	NR	NR
Meys 2016[44]	NR	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	NA	120 days or less
Meys 2016[44]	NR	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	NA	120 days or less
Moffatt 2016[45]	NR	NR	NR	NR
Sayasneh 2016[46]	NR	Transvaginal ultrasound examinations performed by EFSUMB level 2 ultrasound examiners (nonconsultant gynaecology specialist, gynaecology trainees doctors and gynaecology sonographers) The ultrasound examiners were blind to the results of the reference test Transvaginal ultrasonography was performed using the standardised approach previously published by the IOTA group[14] Transabdominal ultrasonography was undertaken for a large mass	NR	NR

Study Details	ADNEX test details			
	Analyser Manufacturer of CA125 Assay	**Ultrasound details**	**Sample collection storage**	**Time from test to surgery**
Szubert 2016[42]	Unclear Immunoenzymatic test (ST AIA-PACK OVCATosoH Bioscience) and Cobas-Core CA-125-II (Roche) Roche and TosoH Bioscience	Poland: Aloka Alpha 10 (3.75–7.5 MHz) endovaginal probe and Aloka 3500 (7.5 MHz) endovaginal probe (Hitach Aloka) Spain: Transvaginal or transrectal ultrasound Voluson E8 (RIC5-9MHz) endovaginal probe (GE Healthcare) A transabdominal probe was used for large tumors Tumors were ultrasonographically assessed according to the 2000 IOTA criteria[14]	Poland: Assessed 1 to 5 days before surgery Spain: Assessed 1 to 5 days before ultrasound No further details reported	1 to 5 days
Van Calster 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 transvaginal ultrasound examination (additional transabdominal sonography for women with large masses) All ultrasound examinations were performed by one of three experienced practitioners (8 to 20 years experience in gynaecological sonography and, EFSMUB level 2 (poland) and level 3 (Spain))	NR	NR

*:personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

IOTA simple ultrasound rules

Study details	Ultrasound details	Time from test to surgery
Abdalla 2013[48]	Transvaginal ultrasound with transabdominal ultrasound for tumours larger than 5 cm and extended beyond the pelvis minor. Morphology, echostructure and vascularization were assesed by Doppler examination. Ultrasound examinations were performed by the attending physician (various levels of experience) rior to referral to the hospital	90 days or less
Alcazar 2013[52]	Transvaginal color Doppler ultrasound (5–9 MHz transducers), Voluson E8 or 730 machine (GE Healthcare, 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 2013[67]	Retospective review of ultrasound scan reports	NR
DiLegge 2012[62]	High frequency transvaginal probe with transabdominal ultasonography for large masses that could not be entirely visualised using a transvaginal probe	120 days or less
Fathallah 2011[64]	Endovaginal	NR
IOTA5 2017*	[redacted]	[redacted]
Knafel 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 a half day of practical training in IOTA simple ultrasound rules before the study	90 days or less
Meys 2016[44]	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	120 days or less
Murala 2014[60]	Scan images were analyses by non-expert gynaecology trainees and masses were classified as benign, malignant or inconclusive	NR
Piovano 2016[58]	Gray-scale and Doppler transvaginal ultrasound, performed by a trainee who had undergone IOTA simple rules traininf and was supervised by an experienced examiner. A transabdominal probe was used for large masses that culd not be entirely visualised transvaginally. All inconclusive messes were re-evaluated by a cunsultant expert (EFSUMB level)	30 days or less
Ruiz de Gauna 2015[65]	Transvaginal color Doppler ultrasound (5–9 MHz transducers), Voluson E8 machine (GE Healthcare, USA). Transabdominal scanning was also performed in large masses. In Centre A ultrasound was performed by an expert and in center B was performed by a trainee. In both centres, masses classified as inconclusive by IOTA simple 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 three weeks after ultrasound examination

Study details	Ultrasound details	Time from test to surgery
Sayasneh 2013b[63]	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 or less
Silvestre 2015[55]	The descriptions of the masses were interpreted based on the IOTA simple rules[15] to characterise whether the features were malignant (M) or benign (B). Vascular power Doppler score is included in IOTA simple rules as one variable: a score of 1 is given when no blood flow is found in the tumor, a score of 2 when only minimal flow is detected, a score of three when moderate flow is present and a score of four when the tumor presents marked blood flow	7 days
Tantipalakorn 2014[51]	Transabdominal (3.5-5 MHz curvilinear transducer) or transvaginal (real time 5-7.5 MHz) or both, connected to Aloka model SSD alpha-10 (Tokyo, Japan). IOTA simple 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 inconclusivie	All participants underwent ultrasound examination within 24 hours of operation
Testa 2014[50]	Standardised transvaginal ultrasound by examiners experienced in gynaecological ultrasound (Level III Education, Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology); grey scale and Doppler imaging; where there was more than one adnexal mass, the mass with the most complex morphology was assessed and analysed	120 days or less
Timmerman 2010[66]	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 five 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	120 days or less
Timmerman 2010[66]	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 five 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	120 days or less

Study details	Ultrasound details	Time from test to surgery
Tinnangwattana 2015[47]	All examinations were done with either transabdominal or transvaginal approach as suitable, using real-time 5-7.5 MHz transvaginal or 2.5-5 MHz transabdominal curvilinear transducer connected to a machine Hitachi- Aloka model ProSound37	24 hours or less
Weinberger 2013[53]	Retsopective analysis by an experienced sonographer	NR

*:personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

OVERA (MIA2G)

Study Details	Overa (MIA2G) test details		
	Analyser Manufacturer of Assays	**Sample collection storage**	**Time from test to surgery**
Coleman 2016[71]	Roche cobas 6000 clinical analyzer (c501 and e601 modules) Roche cobas assays for apolipoprotein A-1, TRF (immunoturbidimetric assays), CA125-II, HE4, and FSH (electrochemiluminescent detection)	A preoperative blood sample of 80mL was processed within 1 to 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 to 11)
Shulman 2016[105]	NR	NR	NR
Zhang 2015[69]	NR	NR	NR

RMI 1

Study Details	RMI test details			
	Analyser **Manufacturer of CA125 Assay**	**Ultrasound details**	**Sample collection** **storage**	**Time from test to surgery**
Aarenstrup 2012[84]	ARCHITECT I 2000sr Abbott	No details reported	Blood samples were collected within 2 weeks prior to surgery Samples were centrifuged within 6 h of collection. After centrifugation, serum samples were stored at -80°C until analysis	Time from ultrasound to surgery not reported
Abdalla 2013[48]	NR	Transvaginal ultrasound with transabdominal ultrasound for tumours larger than 5 cm and extended beyond the pelvis minor Morphology, echostructure and vascularization 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 or less
Aktürk 2011[72]	ECLIA Roche	Siemens transvaginal 7.5-MHz transducer, with transabdominal ultrasound if the mass was too large for complete visualisation transvaginally	Serum samples were collected pre-operatively NR	NR
Al Musalhi 2016[104]	ARCHITECT I2000 Abbott	Pelvic ultrasonography by specialist gynaecologists	Samples collected using serum separator tubes and centrifuged immediately Serum samples were stored at -20°C	NR

Study Details	RMI test details			
	Analyser **Manufacturer of CA125 Assay**	**Ultrasound details**	**Sample collection** **storage**	**Time from test to surgery**
Asif 2004[78]	Immulite-Automated Analyser DPC-U5A (CA125) Immulite	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 = 2 to 5 positive factors	Venous blood collected in a plain tube avoiding hemodialysis. Serum was isolated by centrifugation. Serum was stored at -20°C	NR
Davies 1993[80]	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 sample were drawn from each patient before surgery. Blood was allowed to clot at room temperature then centrifuged at 3000 rpm for 10 mins and serum was separated and stored at -20 deg C	NR
DiLegge 2012[62]	Centocor or Cis-Bio or Abbott Axsym system or Immuno-l-analyser or Vidas Centocor or Cis-Bio or Abbott 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 or less
IOTA5 2017*	[redacted]	[redacted]	[redacted]	[redacted]
Jacobs 1990[79]	Radioimmunoassay Abbott Laboratories	One point score was assigned for ultrasound investigation for each of the following: multilocular cyst, evidence of solid areas, evidence of metastases,	Peripheral venous blood sample were drawn from each patient before surgery.	NR

Study Details	RMI test details			
	Analyser **Manufacturer of CA125 Assay**	**Ultrasound details**	**Sample collection** **storage**	**Time from test to surgery**
	CA125	presence of ascites and bilateral lesions.	Blood was allowed to clot at room temperature then centrifuged at 3000 rpm for 10 mins and serum was separated and stored at -20 deg C	
Lou 2010[74]	NR	NR	NR	NR
Manjunath 2001[76]	microparticle enzyme immunoassay Abbott 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 was assigned for the following ultrasound features suggestive of malignancy: the presence of a multilocular cystic lesion, solid areas, bilateral lesions, ascites, and intraabdominal metastases, scored as one point each.	Serum samples were collected preoperatively No details of storage reported	NR
Meys 2016[44]	NR	Transvaginal or transrectal ultrasound with transabdominal ultrasound for larger masses	NR	120 days or less
Morgante 1999NR	NA Centocor	Siemens Somoline SL2 with transabdominal probe (3.5MHz) 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; intra-abdominal metastates	Peripheral venous blood sample were drawn from each patient before surgery. Serum stored at -15 °C	NR
Sayasneh 2013b[63]	Abbott Architect CA125 II immunoassay kit, Advia Centaur XP	Transabdominal ultasonography was used for large masses that could not be entirely visualised using a transvaginal probe	NR	NR

Study Details	RMI test details			
	Analyser **Manufacturer of CA125 Assay**	**Ultrasound details**	**Sample collection** **storage**	**Time from test to surgery**
	Immunoassay System, UniCel DxI Immunoassay System Beckman; Abbott; Siemens			
Testa 2014[50]	NR	Standardised transvaginal ultrasound by examiners experienced in gynaecological ultrasound (EFSUMB Level 3) Grey scale and Doppler imaging Where there was more than one adnexal mass, the mass with the most complex morphology was assessed and analysed	NR	120 days or less
Timmerman 2010[66]	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 five 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
Tingulstad 1996[77]	NR	Transvaginal ultrasound with transabdominal ultrasound as needed	NR	NR

Study Details	RMI test details			
	Analyser **Manufacturer of CA125 Assay**	**Ultrasound details**	**Sample collection** **storage**	**Time from test to surgery**
	Abbott			
Ulusoy 2007[75]	Roche–Hitachi Modular E170 Immunologic Analyzer System NR	Ultrasound examinations performed with Toshiba Sonolayer SSA-270A and/or a Siemens Sonoline G50 (abdominal convex transducers and/or endovaginal probes). Gynecological oncologists evaluated all cases.	NR	NR
Van Gorp 2012[99]	NR Fujirebio	Standardised transvaginal ultrasound 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 intclotting tubes, immediately before surgery After centrifugation, serum samples were stored at -80°C until analysis	Time from ultrasound to surgery not reported
Winarto 2014[100]	NR Abbott	Transvaginal ultrasound	NR	Samples collected 1 day before surgery, time from ultrasound to surgery unclear
Yamamoto 2009[73]	ECLusys CA125 II assay Roche	Transvaginally with a 6.0 MHz transducer (an abdominal scan was also conducted when indicated)	Blood samples taken pre-operatively No further details reported	NR
Yanaranop 2016[90]	Cobas 6000 Roche	Pelvic (transabdominal or transvaginal) ultrasound using a Voluson E8 (GE Medical Systems)	Samples collected within 48 h prior to surgery and centrifuged immediately	within 6 weeks to surgery

Study Details	RMI test details			
	Analyser **Manufacturer of CA125 Assay**	**Ultrasound details**	**Sample collection** **storage**	**Time from test to surgery**
		Examiner blinded to clinical information and serum biomarkers Morphological features noted (wall structure and thickness, echogenicity, multiocularity, solid areas, bilaterality, ascites, intra-abdominal metastases)	Serum samples were stored at -20°C	

*:personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

Table 36: Study level data for the histological details of malignant tumour diagnoses

Study ID	Test	Histology Details For Malignancy (n)
Aarenstrup 2012[84]	ROMA Abbott; RMI 1	No histology details
Adballa 2013[48]	IOTA simple ultrasound rules; RMI 1	Malignancy included: serous cystadenocarcinoma (7), metastatic tumours from the gastrointestinal tract (3), borderline tumours (3), mucinous cystadenocarcinoma (1), fallopian tube carcinoma (1), mixed carcinoma(1) and undifferentiated carcinoma (1)
Aktürk 2011[72]	RMI 1 threshold comparison	Malignancy included: primary ovarian cancer (19) and metastases (1)
Al Musalhi[104]	ROMA Abbott; RMI	Malignancy included: serous adenocarcinoma (20), mucinous adenocarcinoma (1), endometrial adenocarcinoma (3), undifferentiated (1), borderline epithelial (7), granulosa (5), yolk sac cancer (1), teratoma (2), secondaries (7) and lymphoma (1)
Alcazar 2013[52]	IOTA simple ultrasound rules	Malignancy included: invasive epithelial ovarian cancer (29), borderline ovarian cancer (16), other malignancy (7)
Asif 2004[78]	RMI 1 threshold comparison	No histology details
Baker 2013[67]	IOTA simple ultrasound rules	No histology details
Chan 2013[83]	ROMA Abbott	Malignancy included epithelial ovarian carcinoma only (65): Serous (30), Mucinous (14), Endometrioid (7), Clear cell (8), Mixed (5), Poorly differentiated (1) Subgroups looked at Stage I, II, II, IV and borderline
Clemente 2015[91]	ROMA Abbott	No histology details

Study ID	Test	Histology Details For Malignancy (n)
Coleman 2016[71]	Overa (MIA2G)	Malignancy included: epithelial ovarian cancer (60), non-epithelial ovarian cancer (5), borderline ovarian cancer (17), metastases to the ovaries (6) and other non-ovarian malignancies (4)
Davies 1993[80]	RMI 1 threshold comparison	Malignancy included: epithelial ovarian cancer (28), borderline ovarian cancer (7), other malignancies (2)
Di Legge 2012[62]	IOTA simple ultrasound rules; RMI 1	Malignancy included primary invasive (476), borderline (128) and metastatic tumours (78)
Fathallah 2011[64]	IOTA simple ultrasound rules	Malignancy included: primary ovarian cancer (8); borderline ovarian cancer (6)
IOTA5 2017*	ADNEX; IOTA simple ultrasound rules; RMI 1	[redacted]
Jacobs 1990[79]	RMI 1 threshold comparison	Malignancy included: primary invasive epithelial ovarian malignancies (36), dysgerminoma (1), metastatic bowel adenocarcinoma (1), borderline malignancy (4)
Janas 2015[98]	ROMA Roche	Malignancy included: primary ovarian cancer (44), metastases to the ovary (14), borderline tumours (8) Subgroups looked at ovarian alone
Joyeux 2016[43]	ADNEX	Malignancy included: primary ovarian cancer (25) and borderline ovarian cancer (5)
Knafel 2015[49]	IOTA simple ultrasound rules	Malignancy included ovarian carcinoma (60: serous, clear cell, endometrioid, mucinous, undifferentiated, carcinosarcoma), borderline (7), sex cord stromal tumours (2), germ cell tumours (5), metastases (8) Note that 15/82 malignancies were not classed as ovarian
Langhe 2013[95]	ROMA Fujirebio	Malignant included 53 borderline tumours No further details reported
Li 2016[97]	ROMA Abbott	Ovarian malignancy included serous (80), mucinous (42), endometrioid (40), clear cell (21), and undifferentiated (7)
Lou 2010[74]	RMI 1 threshold comparison	Ovarian malignancy included epithelial ovarian cancer (50), non-epithelial ovarian cancer (8) and metastatic carcinoma (3)
Manjunath 2001[76]	RMI 1 threshold comparison	Malignancy included: primary ovarian malignancies (88), germ cell tumours (3), metastases (2)
Meys 2016[44]	ADNEX; IOTA simple ultrasound rules; RMI 1	Malignancy included epithelial ovarian cancer (70), borderline (27), granulosa cell carcinoma (3), yolk sac tumour (1), metastatic tumour (10), non-primary ovarian carcinoma (4)
Moffatt 2016[45]	ADNEX	No histology details
Moore 2011[102]	ROMA Abbott	Epithelial ovarian cancer (43) and low malignant potential tumours (14), (non-epithelial tumours and other gynaecological cancers, other cancers and metastatic cancers excluded) Subgroups looked at stage I, II, III, IV
Morgante 1999[81]	RMI 1 threshold comparison	Malignancy included (31): serous cystadenocarcinoma (14), mucinous cystadenocarcinoma (6), borderline (2), clear cell carcinoma (2) undifferentiated carcinoma (2), granulosa cell carcinoma (1) Kruckemberg (1),

Study ID	Test	Histology Details For Malignancy (n)
		immature teratoma (1), endometrioid adenocarcinoma (1), metastatic carcinoma (1)
Murala 2014[60]	IOTA simple ultrasound rules; RMI 1	No histology details
Novotny 2012[87]	ROMA Abbott	No histology details
Piovano 2016[58]	IOTA simple ultrasound rules	Ovarian malignancy included epithelial ovarian cancer (45), non-epithelial ovarian cancer (8), borderline (22) and metastatic carcinoma (9)
Presl 2012[82]	ROMA Abbott	No histology details
Ruiz de Gauna 2015[65]	IOTA simple ultrasound rules	Malignancy included primary ovarian cancer, borderline ovarian cancer and metastases; no numbers reported
Sayasneh 2013b[63]	IOTA simple ultrasound rules; RMI 1	Malignancy included, borderline (18), serous cyst/adenocarcinoma (26), mucinous cyst/adenocarcinoma (7), endometrioid carcinoma (6), Clear cell carcinoma (5), granulosa cell tumour (1), transitional cell tumour (1), signet ring cell adenocarcinoma (1), peritoneal serous adenocarcinoma (1), gastrointestinal adenocarcinomas (5), malignant mixed Mullerian tumour (1), large cell neuroendocrine carcinoma (1) and endocrine tumour (1)
Sayasneh 2016[46]	ADNEX	Malignancy included: primary ovarian cancer (116), borderline ovarian cancer (42) and metastatic ovarian cancer (24)
Shulman 2016[105]	Overa (MIA2G); ROMA Roche	Malignancy included: epithelial ovarian cancer (150), non-epithelial ovarian cancer (16), borderline ovarian cancer (42), metastases (23) and non-ovarian malignancies (14)
Silvestre 2015[55]	IOTA simple ultrasound rules	Malignancy included: primary ovarian malignancies (15), borderline ovarian malignancy (5), metastases (5), other malignancies (7)
Szubert 2016[42]	ADNEX	Spain. Malignancy included: primary ovarian cancer (26), borderline ovarian cancer (3), metastatic ovarian cancer (5)
		Poland. Malignancy included: primary ovarian cancer (53), borderline ovarian cancer (12), metastatic ovarian cancer (5)
Tantipalakorn 2014[51]	IOTA simple ultrasound rules	Malignancy included: primary ovarian cancers (62), borderline ovarian cancer (12), germ cell tumours (9), sex cord stromal tumour (6), metastatic adenocarcinoma (10) and other malignant tumours (8)
Testa 2014[50]	IOTA simple ultrasound rules; RMI 1	Malignancy included primary invasive ovarian cancer (633), borderline ovarian tumours (153), metastatic (126), rare primary invasive (68)
Timmerman 2010[66]	IOTA simple rules; RMI 1	Malignancy included: Borderline (111), primary invasive, stages I-IV and rare (373) and metastatic (58)
Tingulstad 1996[77]	RMI 1 threshold comparison	Malignancy included ovarian cancer (51), neurosarcoma (1), leiomyosarcoma (1), lymphoma (1), Kruckenberg tumour (1), rectal cancer (1)
Tinnangwattana 2015[47]	IOTA simple ultrasound rules	Malignancy included: primary ovarian malignancies (13), Borderline (8), metastases (5), other malignancies (3)
Ulusoy 2007[75]	RMI 1 threshold comparison	Malignancy included: primary ovarian cancers (84), borderline ovarian cancers (15), metastases (7)

Study ID	Test	Histology Details For Malignancy (n)
Van Calster 2014[17]	ADNEX	Malignancy included: primary ovarian cancer (701), borderline ovarian cancer (153), metastases (126). IOTA phase 3 - validation data set
Van Gorp 2012[99]	ROMA Fujirebio; RMI 1	Malignancy included epithelial ovarian cancer stages I-IV (120), non-epithelial ovarian cancer (4), metastatic (25)
Weinberger 2013[53]	IOTA simple ultrasound rules	Malignancy included all invasive ovarian cancers and borderline tumours. No further details reported
Winarto 2014[100]	ROMA Abbott; RMI 1	Malignancy included serous cystadenocarcinoma (19), endometrioid (14), mucinous (8), clear cell (7), carcinosarcoma (2), borderline (17)
Xu 2016[96]	ROMA Roche	Malignancy was epithelial ovarian cancer (210): Endometrioid 80, serous (59), papillary serous (15), mucinous (6), seromucinous (2), clear cell (12), adenocarcinoma (36)
Yamamoto 2009[73]	RMI 1 threshold comparison	Malignancy included: primary ovarian cancer (29), borderline ovarian cancer (8), tubal cancer (3)
Yanaranop 2016[90]	ROMA Roche; RMI 1	Malignancy included epithelial ovarian carcinoma (66) and non-epithelial ovarian cancer (8) Subgroups looked at: epithelial and stage I, II, III, IV
Zhang 2015[69]	Overa (MIA2G)	72 malignant cases includes 19 stage I/II and 13 LMP
Zhang 2015b[103]	ROMA Roche	Malignancy was described as epithelial ovarian cancer (264); serous (170), mucinous (20), endometrioid (25), other kinds (13), unknown (36) Subgroups were looked at: stage I, II, III, IV

*:personal communication: e-mail via. Frances Nixon, Technical Advisor, NICE Diagnostic Assessment Programme to Marie Westwood, Project lead, KSR Ltd, 01/03/2017

APPENDIX 3: EXAMPLE ASSESSMENTS OF STUDY QUALITY

Example QUADAS-2 assessment

van Calster 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 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 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 dataset. No details regarding who performed tests and whether they were blind, not 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 pre-specified?		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 2x2 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 or less.		
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**

Example PROBAST assessment

Van Calster 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 more than 120 days after the ultrasound examination. IOTA was established to develop and validate diagnostic models for adnexal masses, based on large multi-centre datasets, using a standardised ultrasound examination protocol, terms and definitions. The ADNEX model was developed using data collected in IOTA phases 1, 1b and 2 (1999 to 2007) and validated using data collected in phase 3 (2009 to 2012); 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.			

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.			

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 pathologists were aware of other predictor information. The final diagnosis was divided into five types: benign, borderline, stage I invasive, stage II-IV invasive, and secondary metastatic cancer.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Yes	Yes
3.2 Was a pre-specified 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 pathologists were blinded to CA125 results.			
B. Applicability			
At what time point was the outcome determined: All surgery was performed within 120 days of ultrasound exanimation. *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:			

DOMAIN 4: Analysis

Risk of Bias			
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor: The development dataset included 3506 women and the validation dataset included 2403 women. There were 10 candidate predictors. The development dataset included 949 (27%) women with malignancies (including borderline tumours) and the validation dataset included 980 (41%) women with malignancies (including borderline tumours).			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): To avoid over fitting, ten candidate predictors were selected; selection was based on topic expertise and stability of predictors across centres. Further, 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: CA125 data were missing for 31% of participants. Predictive mean matching regression, using variables that were related to either the level of CA-125 itself, or to the unavailability of CA-125 (i.e. a binary indicator indicating for each woman whether CA-125 was missing or not) 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:**	**Unclear**	**Low**

	(low/ high/ unclear)		
Rationale of bias rating: Some aspects of model development were not fully reported.			

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: ADDITIONAL RESULTS

Table 37: Additional accuracy data for the ROMA score (accuracy using thresholds other than those recommended by manufacturers)

TEST	Study ID	Subgroup	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
ROMA Abbott	***Target condition: Ovarian malignancies (undefined – not clear whether borderline tumours were included)***									
	[C]Clemente 2015[91]	All	NR	4	5	14	39	62	44.4 (13.7, 78.8)	73.6 (59.7, 84.7)
		Pre-menopausal	NR	2	3	11	32	48	40.0 (5.3, 85.3)	74.4 (58.8, 86.5)
		Post-menopausal	NR	2	2	3	7	14	50.0 (6.8, 93.2)	70.0 (34.8, 93.9)
	[C]Novotny 2012[87]	Post-menopausal	37.7%	18	3	13	243	277	85.7 (63.7, 97.0)	94.9 (91.5, 97.3)
	Target condition: Epithelial ovarian malignancies including borderline									
	Moore 2011[102]	All	13.1%/27.7%	59	8	96	287	450	88.1 (77.8, 94.7)	74.9 (70.3, 79.2)
		Pre-menopausal	13.1%	13	3	60	173	249	81.3 (54.4, 96.0)	74.2 (68.1, 79.7)
		Post-menopausal	27.7%	46	5	36	114	201	90.2 (78.6, 96.7)	76.0 (68.4, 82.6)
	Target condition: Epithelial ovarian malignancies (stage III/IV) – borderline and stage I/II tumours excluded									
	Moore 2011[102]	All	13.1%/27.7%	34	0	96	287	417	100 (89.7, 100)	74.9 (70.3, 79.2)
		Pre-menopausal	13.10%	5	0	60	173	238	100 (47.8, 100)	74.2 (68.1, 79.7)
		Post-menopausal	27.7%	29	0	36	114	179	100 (88.1, 100)	76.0 (68.4, 82.6)
	Target condition: Epithelial ovarian malignancies (stage I/II) – borderline and stage III/IV tumours excluded									

TEST	Study ID	Subgroup	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
	Moore 2011[102]	All	13.1%/27.7%	9	3	96	287	395	75.0 (42.8, 94.5)	74.9 (70.3, 79.2)
		Pre-menopausal	13.1%	3	0	60	173	236	100 (29.2, 100)	74.2 (68.1, 79.7)
		Post-menopausal	27.7%	6	3	36	114	159	66.7 (29.9, 92.5)	25.3 (21.4, 29.6)
ROMA Roche	***Target condition: Epithelial ovarian malignancies excluding borderline***									
	Xu 2016[96]	Pre-menopausal	13.40	58	49	23	241	371	54.2 (44.3, 63.9)	91.3 (87.2, 94.4)
		Post-menopausal	18.70	69	34	6	41	150	67.0 (57.0, 75.9)	87.2 (74.3, 95.2)
	Target condition: Epithelial ovarian malignancies (stage III/IV) – borderline and stage I/II tumours excluded									
	[C]Zhang 2015b[103]	All	11.4%/29.9%	143	16	72	276	507	89.9 (84.2, 94.1)	79.3 (74.7, 83.4)
		Pre-menopausal	11.4%	40	10	58	227	335	80.0 (66.3, 90.0)	79.6 (74.5, 84.2)
		Post-menopausal	29.9%	103	6	14	49	172	94.5 (88.4, 98.0)	77.8 (65.5, 87.3)
	Target condition: Epithelial ovarian malignancies (stage I/II) – borderline and stage III/IV tumours excluded									
	[C]Zhang 2015b[103]	All	11.4%/29.9%	49	15	72	276	412	76.6 (64.3, 86.2)	79.3 (74.7, 83.4)
		Pre-menopausal	11.4%	21	9	58	227	315	70.0 (50.6, 85.3)	79.6 (74.5, 84.2)
		Post-menopausal	29.9%	28	6	14	49	97	82.4 (65.5, 93.2)	77.8 (65.5, 87.3)

c: 2x2 data were calculated (other studies reported 2x2 data); *:calculated confidence intervals

Table 38: Accuracy of the ADNEX model at thresholds other than 10%

Study ID	Threshold %	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
All malignant tumours including borderline									
Sayasneh 2016[46]	1	182	0	377	51	610	calculated	100 (97.4, 100)	11.9 (9.1, 15.5)
	3	182	0	297	131	610	calculated	100 (97.4, 100)	30.6 (26.3, 35.3)
	5	180	2	200	228	610	calculated	99 (94.9, 99.8)	53.2 (48.2, 58.1)
	15	172	10	106	322	610	calculated	94.4 (90, 97)	75.2 (70.7, 79.2)
	20	165	17	89	339	610	calculated	90.6 (85.2, 94.1)	79.3 (75.1, 83)
	30	157	25	69	359	610	calculated	86.3 (80.4, 90.6)	83.9 (80.1, 87.2)
Van Calster 2014[17]	3	969	11	760	663	2403	calculated	98.9 (98, 99.4)	46.6 (44, 49.2)
	5	964	16	578	845	2403	calculated	98.4 (97.4, 99.1)	59.4 (56.8, 62)
	15	913	67	324	1099	2403	calculated	93.2 (92.5, 95.6)	77.2 (74.9, 79.3)
Ovarian malignancies including borderline									
Joyeux 2016[43]	3	30	0	134	120	284	calculated	100 (88.4, 100)	47.2 (41, 53.6)
	5	29	1	78	176	284	calculated	96.6 (82.8, 99.9)	69.2 (63.2, 74.9)
	15	26	4	38	216	284	calculated	86.6 (69.3, 96.2)	85 (80, 89.2)

Table 39: Accuracy of IOTA simple rules, where inconclusive results were not classified

Threshold	Study ID	Subgroup	Index Test variations	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline											
Malignant (inconclusive were excluded)	Alcazar 2013[52]	All		29	4	6	231	270	reported	87.9 (72.4, 95.2)	97.5 (94.6, 98.8)
	Ruiz de Gauna 2015[65]	Centre A		27	0	4	62	93	reported	100 (87.5, 100)	93.9 (85.4, 97.6)
		Centre B		11	2	4	92	109	reported	84.6 (57.8, 95.7)	95.8 (89.8, 98.4)
	Silvestre 2015[55]	All		32	0	8	26	66	reported	100 (89.1, 100)	76.5 (58.8, 89.3)
	Tantipalakorn 2014[51]	All		88	19	10	202	319 (masses)	reported	82.2 (75, 89.5)	95.3 (92.4, 98.1)
	Tinnangwattana 2015[47]	All		25	3	11	55	94	reported	89.3 (77.8, 100)	83.3 (74.3, 92.3)
Malignant (inconclusive were classified by SA) + ROMA >11.4%/29.9%	Piovano 2016[58]	All	+ ROMA	76	8	61	246	391	calculated	90.5 (82.1, 95.8)	80.1 (75.2, 84.4)
		Post-menopausal	+ ROMA	58	5	25	82	170	calculated	92.0 (85.0, 99.0)	77.0 (69.0, 85.0)
		Pre-menopausal	+ ROMA	18	3	36	164	221	calculated	86.0 (71.0, 100)	82.0 (77.0, 87.0)
Malignant (inconclusive were classified by SA) + HE4 ≥70/140		All	+ HE4	73	11	42	265	391	calculated	86.9 (77.8, 93.3)	86.3 (82.0, 90.0)
		Post-menopausal	+ HE4	55	8	18	89	170	calculated	87.0 (79.0, 96.0)	83.0 (76.0, 90.0)
		Pre-menopausal	+ HE4	18	3	24	176	221	calculated	86.0 (71.0, 100)	88.0 (83.0, 92.0)
Malignant on IOTA simple rules (inconclusives were classified by SA) + CA125 ≥35		All	+ CA125	76	8	98	209	391	calculated	90.5 (82.1, 95.8)	68.1 (62.5, 73.3)
		Post-menopausal	+ CA125	58	5	26	81	170	calculated	92.0 (85.0, 99.0)	76.0 (68.0, 84.0)
		Pre-menopausal	+ CA125	18	3	72	128	221	calculated	86.0 (71.0, 100)	64.0 (57.0, 61.0)
Malignant (inconclusives were classified by expert SA; final	Ruiz de Gauna 2015[65]	Centre A		31	0	9	74	114	reported	100 (88.8, 100)	89.2 (80.4, 94.9)
		Centre B		13	2	13	105	133	reported	86.7 (59.5, 98.3)	89 (81.9, 94.0)

Threshold	Study ID	Subgroup	Index Test variations	TP	FN	FP	TN	Total	2x2 Data	Sensitivity % (95% CI)	Specificity % (95% CI)
ratings of unclassifiable were treated as malignant)											
Target condition: Ovarian malignancies including borderline											
Malignant (inconclusive were excluded)	Fathallah 2011[64]	All		8	3	3	95	109	reported	73.0 (45.0, 100)	97.0 (94.0, 100)
Malignant (handling of inconclusive was unclear)	Weinberger 2013[53]	All		118	7	16	206	347	calculated	94.0 (88.8, 97.7)	93.0 (88.6, 95.8)
Target condition: Ovarian malignancies excluding borderline											
Malignant (handling of inconclusive was unclear)	Weinberger 2013[53]	All		99	2	16	222	323	calculated	98.0 (93.0, 99.8)	93.0 (89.3, 96.1)
Target condition: Ovarian malignancies (undefined – not clear whether borderline tumours were included)											
Malignant (inconclusive were excluded)	Baker 2013[67]	Pre-menopausal		2	0	5	21	28	calculated	100 (15.8, 100)	80.8 (60.6, 93.4)
Target condition: Ovarian borderline tumours											
Malignant (handling of inconclusive was unclear)	Weinberger 2013[53]	All		19	5	16	222	262	calculated	79.2 (57.8, 92.9)	93.3 (89.3, 96.1)

Table 40: Accuracy of RMI 1 at decision thresholds other than 200 and 250

Study ID	CA125 assay	Ultrasound details	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
All malignant tumours including borderline										
Davies 1993[80]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer	25	36	2	39	48	124	94.7 (82.3, 99.4)	55.2 (44.1, 65.9)
Jacobs 1990[79]	RIA (Abbott)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer		41	0	37	61	139	100 (91.4, 100)	62.2 (51.9, 71.8)
Morgante 1999[81]	RIA, Centocor	A Siemens Sonoline SL2 was used with a 3.5 MHz transabdominal sectorial probe and a 5.0-7.5 MHz transvaginal probe		30	1	37	56	124	96.8 (83.3, 99.9)	60.2 (49.5, 70.2)
Tingulstad 1996[77]	IMX Abbott	Transvaginal with transabdominal as needed		51	5	37	80	173	91.1 (80.4, 97.0)	68.4 (59.1, 76.7)
Ulusoy 2007[75]	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, 97.9)	28.4 (22.1, 35.4)
Summary estimates									**94.9 (91.5, 97.2)**	**51.1 (47.0, 55.2)**
Davies 1993[80]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer	50	36	2	28	59	124	94.7 (82.3, 99.4)	67.8 (56.9, 77.4)
Jacobs 1990[79]	RIA (Abbott)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer		39	2	23	75	139	95.1 (83.5, 99.4)	76.5 (66.9, 84.5)
Lou 2010[74]	NR	NR		49	12	36	126	223	80.3 (68.2, 89.4)	77.8 (70.6, 83.9)
Morgante 1999[81]	RIA, Centocor	A Siemens Sonoline SL2 was used with a 3.5 MHz transabdominal sectorial probe and a 5.0-7.5 MHz transvaginal probe		29	2	23	70	124	93.5 (78.6, 99.2)	75.3 (65.2, 83.6)
Tingulstad 1996[77]	IMX Abbott	Transvaginal wit transabdominal as needed		49	7	22	95	173	87.5 (75.9, 94.8)	81.2 (72.9, 87.8)

Study ID	CA125 assay	Ultrasound details	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Ulusoy 2007[75]	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, 95.4)	44.2 (37, 51.6)
Summary estimates									**89.5 (85.7, 92.6)**	**68.1 (64.7, 71.5)**
Davies 1993[80]	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, 97.0)	79.3 (69.3, 87.3)
Jacobs 1990[79]	RIA (Abbott)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer		38	3	15	83	139	92.7 (80.2, 98.5)	84.7 (76, 91.2)
Morgante 1999[81]	RIA, Centocor	A Siemens Sonoline SL2 was used with a 3.5 MHz transabdominal sectorial probe and a 5.0-7.5 MHz transvaginal probe		25	6	19	74	124	80.6 (62.5, 92.5)	79.6 (69.9, 87.2)
Tingulstad 1996[77]	IMX Abbott	Transvaginal wit transabdominal as needed		44	12	14	107	173	78.6 (65.6, 88.4)	88.4 (81.3, 93.5)
Summary estimates									**84.8 (78.5, 89.9)**	**83.5 (79.4, 87.0)**
Davies 1993[80]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer	100	32	5	13	74	124	86.5 (71.2, 95.5)	85.1 (75.8, 91.8)
Jacobs 1990[79]	RIA (Abbott)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer		35	6	12	86	139	85.4 (70.8, 94.4)	87.8 (79.6, 93.5)
Lou 2010[74]	NR	NR		45	16	18	144	223	73.8 (60.9, 84.2)	88.9 (83.0, 93.3)
Morgante 1999[81]	RIA, Centocor	A Siemens Sonoline SL2 was used with a 3.5 MHz transabdominal sectorial probe and a 5.0-7.5 MHz transvaginal probe		24	7	9	84	124	77.4 (58.9, 90.4)	90.3 (82.4, 95.5)
Tingulstad 1996[77]	IMX Abbott	Transvaginal wit transabdominal as needed		44	12	13	108	173	78.6 (65.6, 88.4)	89.3 (82.3, 94.2)
Summary estimates									**79.6 (73.8, 84.7)**	**88.4 (85.5, 90.9)**

Study ID	CA125 assay	Ultrasound details	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Morgante 1999[81]	RIA, Centocor	A Siemens Sonoline SL2 was used with a 3.5 MHz transabdominal sectorial probe and a 5.0-7.5 MHz transvaginal probe	120 or 125	23	8	7	86	124	74.2 (55.4, 88.1)	92.5 (85.1, 96.9)
Tingulstad 1996[77]	IMX Abbott	Transvaginal wit transabdominal as needed		44	12	12	109	173	78.6 (65.6, 88.4)	90.1 (83.3, 94.8)
Ulusoy 2007[75]	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, 88.1)	68.9 (61.8, 75.4)
Summary estimates									**79.3 (72.9, 84.8)**	**80.7 (76.5, 84.4)**
Davies 1993[80]	RIA (CIS Bioindustries)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer	150	30	7	13	74	124	81.1 (64.8, 92.0)	85.1 (75.8, 91.8)
Jacobs 1990[79]	RIA (Abbott)	Diasonics DS 1 sector scanner with a 3.5 MHz transducer		35	6	6	92	139	85.4 (70.8, 94.4)	93.9 (87.2, 97.7)
Lou 2010[74]	NR	NR		37	24	8	154	223	60.7 (47.3, 72.9)	95.1 (90.5, 97.8)
Morgante 1999[81]	RIA, Centocor	A Siemens Sonoline SL2 was used with a 3.5 MHz transabdominal sectorial probe and a 5.0-7.5 MHz transvaginal probe		20	11	6	87	124	64.5 (45.4, 80.8)	93.5 (86.5, 97.6)
Tingulstad 1996[77]	IMX Abbott	Transvaginal wit transabdominal as needed		43	13	7	110	173	76.8 (63.6, 87.0)	94.0 (88.1, 97.6)
Ulusoy 2007[75]	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, 84.1)	77.9 (71.3, 83.6)

Study ID	CA125 assay	Ultrasound details	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Summary estimates									**74.1 (69.0, 78.7)**	**89.0 (86.6, 91.2)**
Lou 2010[74]	NR	NR	300	33	28	2	160	223	54.1 (40.8, 66.9)	98.8 (95.6, 99.9)
Ulusoy 2007[75]	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	500	57	49	12	178	296	53.8 (43.8, 63.5)	93.7 (89.2, 96.7)
All malignant tumours excluding borderline										
Aktürk 2011[72]	ECLIA, Roche	Siemens transvaginal 7.5-MHz transducer	50	17	3	33	47	100	85 (62.1, 96.8)	58.8 (47.2, 69.6)
			100	15	5	13	67	100	75 (50.9, 91.3)	83.8 (73.8, 91.1)
			150	15	5	12	68	100	75 (50.9, 91.3)	85.0 (75.3, 92.0)
			300	9	11	2	78	100	45 (23.1, 68.5)	97.5 (91.3, 99.7)
			350	9	11	2	78	100	45 (23.1, 68.5)	97.5 (91.3, 99.7)
			400	6	14	2	78	100	30 (11.9, 54.3)	97.5 (91.3, 99.7)
Manjunath 2001[76]	Micro-particle EIA, Abbott	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 large	25	85	8	27	28	148	91.4 (83.8, 96.2)	50.9 (37.1, 64.6)
			50	75	18	21	34	148	80.6 (71.1, 88.1)	61.8 (47.7, 74.6)
			80	74	19	18	37	148	79.6 (69.9, 87.2)	67.3 (53.3, 79.3)
			100	74	19	14	41	148	79.6 (69.9, 87.2)	74.5 (61.0, 85.3)
			125	73	20	11	44	148	78.5 (68.8, 86.3)	80.0 (67.0, 89.6)
			150	72	21	9	46	148	77.4 (67.6, 85.4)	83.6 (71.2, 92.2)
			300	60	33	3	52	148	64.5 (53.9, 74.2)	94.5 (84.9, 98.9)
			350	58	35	3	52	148	62.4 (51.7, 72.2)	94.5 (84.9, 98.9)
			400	57	36	3	52	148	61.3 (50.6, 71.2)	94.5 (84.9, 98.9)
All malignant tumours (undefined – not clear whether borderline tumours were included)										
Asif 2004[78]	IA, Immulite	NR	25	54	1	15	30	100	98.2 (90.3, 100)	66.7 (51.0, 80.0)
			50	53	2	10	35	100	96.4 (87.5, 99.6)	77.8 (62.9, 88.8)
			75	52	3	8	37	100	94.5 (84.9, 98.9)	82.2 (67.9, 92.0)

Study ID	CA125 assay	Ultrasound details	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
			100	49	6	7	38	100	89.1 (77.8, 95.9)	84.4 (70.5, 93.5)
			125	48	7	5	40	100	87.3 (75.5, 94.7)	88.9 (75.9, 96.3)
			150	47	8	4	41	100	85.5 (73.3, 93.5)	91.1 (78.8, 97.5)
			175	47	8	4	41	100	85.5 (73.3, 93.5	91.1 (78.8, 97.5)
			190	47	8	4	41	100	85.5 (73.3, 93.5	91.1 (78.8, 97.5)
			300	40	15	0	45	100	72.7 (59.0, 83.9)	100 (92.1, 100)
Ovarian malignancies including borderline										
Yamamoto 2009[73]	ECLusys CA125 II	Transvaginal (6.0 MHz transducer), with transabdominal as indicated	100	39	1	65	148	253	97.5 (86.8, 99.9)	69.5 (62.8, 75.6)
			150	34	6	36	177	253	85.0 (70.2, 94.3)	83.1 (77.4, 87.9)
			300	27	13	18	195	253	67.5 (50.9, 81.4)	91.5 (87.0, 94.9)

Table 41: Comparative accuracy of the ROMA score using Fujirebio tumour marker assays versus the RMI 1

Study ID	Subgroup	ROMA Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)	RMI	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline																	
Van Gorp 2012[99]	All	12.5%/14.4%	127	23	52	172	374	84.7 (77.9, 90)	76.8 (70.7, 82.2)	200	114	36	17	207	374	76.0 (68.4, 82.6)	92.4 (88.1, 95.5)
	Pre-menopausal	12.5%	26	13	17	122	178	66.7 (49.8, 80.9)	87.8 (81.1, 92.7)	200	25	14	6	133	178	64.1 (47.2, 78.8)	95.7 (90.8, 98.4)
	Post-menopausal	14.4%	101	10	35	50	196	91 (84.1, 95.6)	58.8 (47.6, 69.4)	200	89	22	11	74	196	80.2 (71.5, 87.1)	87.1 (78, 93.4)

Table 42: Accuracy of the ROMA score using Fujirebio tumour marker assays at the manufacturer's recommended thresholds

Study ID	Subgroup	Thresh-old	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline									
Langhe 2013[95 C]	All	12.5%/14.4%	129	47	31	170	377	73.3 (66.1, 79.7)	84.6 (78.8, 89.3)
Van Gorp 2012[99]	All	12.5%/14.4%	127	23	52	172	374	84.7 (77.9, 90.0)	76.8 (70.7, 82.1)
Summary estimates								**78.5 (73.7, 82.9)**	**80.5 (76.4, 84.1)**
Langhe 2013[95 C]	Pre-menopausal	12.5%	23	22	6	81	132	51.1 (35.8, 66.3)	93.1 (85.6, 97.4)
Van Gorp 2012[99]	Pre-menopausal	12.5%	26	13	17	122	178	66.7 (49.8, 80.9)	87.8 (81.1, 92.7)
Summary estimates								**58.3 (47.1, 69.0)**	**89.8 (85.1, 93.4)**
Langhe 2013[95 C]	Post-menopausal	14.4%	105	26	25	89	245	80.2 (72.3, 86.6)	78.1 (69.4, 85.3)
Van Gorp 2012[99]	Post-menopausal	14.4%	101	10	35	50	196	91 (84.1, 95.6)	58.8 (47.6, 69.4)
Summary estimates								**85.1 (80.0, 89.4)**	**69.8 (63.0, 76.1)**

[C]: 2x2 data were calculated; [*]:calculated confidence intervals

Table 43: Accuracy of the ROMA score using Abbott ARCHITECT tumour marker assays at the manufacturer's recommended thresholds (unclear whether borderline tumours were included in the analysis)

Study ID	Subgroup	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: Ovarian malignancies (undefined – not clear whether borderline tumours were included)									
Li 2016[97]	All	7.4%/25.3%	166	24	141	586	917	87.4 (81.8, 91.7)	80.1 (77.0, 83.0)
[C]Presl 2012[82]	All	7.3%/26.3%	25	5	72	450	552	83.3 (65.3, 94.4)	86.2 (82.9, 89.0)
Summary estimates								**86.8 (81.6, 91.0)**	**82.7 (80.5, 84.8)**
Li 2016[97]	Pre-menopausal	7%	96	12	136	501	745	88.9 (81.4, 94.1)	78.6 (75.3, 81.8))
[C]Presl 2012[82]	Pre-menopausal	7.30%	5	4	44	243	296	55.6 (21.2, 86.3)	84.7 (80.0, 88.6)
Summary estimates								**86.3 (78.7, 92.0)**	**80.5 (77.8, 83.0)**
Li 2016[97]	Post-menopausal	25.3%	70	12	5	85	172	85.4 (75.8, 92.2)	94.4 (87.5, 98.2)
[C]Novotny 2012[87]	Post-menopausal	26.3%	20	1	31	225	277	95.2 (76.2, 99.9)	87.9 (83.3, 91.6)
[C]Presl 2012[82]	Post-menopausal	26.3%	20	1	28	207	256	95.2 (76.2, 99.9)	88.1 (83.2, 91.9)
Summary estimates								**88.7 (81.8, 93.7)**	**89.0 (86.2, 91.4)**

[C]: 2x2 data were calculated (other studies reported 2x2 data); [*]:calculated confidence intervals

Table 44: Accuracy of the ROMA score using Roche Elecsys tumour marker assays at the manufacturer's recommended thresholds (unclear whether borderline tumours were included in the analysis)

Study ID	Subgroup	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: Ovarian malignancies (undefined – not clear whether borderline tumours were included)									
[C]Zhang 2015b[103]	All	11.4%/29.9%	224	40	73	275	612	84.8 (79.9, 88.9)	79.0 (74.4, 83.2)
	Pre-menopausal	11.4%	70	25	59	226	380	73.7 (63.6, 82.2)	79.3 (74.1, 83.9)
	Post-menopausal	29.9%	154	15	14	49	232	91.1 (85.8, 94.9)	77.8 (65.5, 87.3)

Table 45: Accuracy of the ADNEX model (unclear whether borderline tumours were included in the analysis)

Study ID	Threshold	TP	FN	FP	TN	Total	Sensitivity % (95% CI)	Specificity % (95% CI)
Moffatt 2016[45]	10%	4	2	29	46	81	66.7 (22.3, 95.7)	61.3 (64.4, 81.6)

Table 46: Additional accuracy data for the RMI score (unclear whether borderline tumours were included in the analysis)

Threshold		200						250					
study ID	Total N	TP	FN	FP	TN	Sens % (95% CI)	Spec % (95% CI)	TP	FN	FP	TN	Sens % (95% CI)	Spec % (95% CI)
All malignant tumours (undefined – not clear whether borderline tumours were included)													
Asif 2004[78]	100	47	8	3	42	85 (NR)	93 (NR)	40	15	2	43	72 (NR)	95 (NR)

Table 47: Accuracy of the ROMA score using Roche tumour marker assays at the manufacturer's recommended thresholds(using unclear inclusion of borderline tumours)

study ID	Subgroup	Thresh-old	TP	FN	FP	TN	Total N	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: Ovarian malignancies (undefined – not clear whether borderline tumours were included)									
[C]Zhang 2015b[103]	All	11.4%/29.9%	224	40	73	275	612	84.8 (79.9, 88.9)	79.0 (74.4, 83.2)
	Pre-menopausal	11.4%	70	25	59	226	380	73.7 (63.6, 82.2)	79.3 (74.1, 83.9)
	Post-menopausal	29.9%	154	15	14	49	232	91.1 (85.8, 94.9)	77.8 (65.5, 87.3)

[C]: 2x2 data were calculated (other studies reported 2x2 data); [*]:calculated confidence intervals

APPENDIX 5: 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), ADNEX model (IOTA), Overa (MIA2G), RMI (using decision thresholds other than 250)

Reference Standard: Histological examination of a surgically resected of biopsy sample; studies which used follow-up as the reference standard for some or all test negative patients were also eligible for inclusion

Outcome: Sufficient data to construct 2x2 table of test performance, or clinical outcomes

The table below summarises studies which 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 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 which were ("Unclear"). Articles which did not report primary research were not assessed further. Any criteria which are not applicable to a study are marked N/A.

Details of excluded studies with rationale for exclusion

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Abbott Laboratories (Singapore). Evaluation of HE4 and CA125 serum markers to improve the risk determination of ovarian cancer in Malaysian women. ISRCTN45238573. In: WHO International Clinical Trials Registry Platform (ICTRP) [Internet]. Geneva: World Health Organization (WHO). 2014 [dated accessed 24.11.16] Available from: http://isrctn.com/ISRCTN45238573	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	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention IOTA regression model (not simple rules of ADNEX)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
presurgical assesment of adnexal masses. *Int J Gynecol Cancer* 2015;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 SUPPL. 1):379							
Abdulrahman Jnr GO, 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;Conference End: 20121016. Conference Publication: (var.pagings). 22:E411	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200, in a tertiary care setting
Abdulrahman GO, McKnight L, Lutchman Singh K. The risk of malignancy index (RMI) in women with adnexal masses in Wales. *Taiwanese Journal of Obstetrics and Gynecology* 2014;53(3):376-381	Yes	Yes	No	No	No	Yes	No relevant intervention Accuracy of RMI 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(2):178-80	Yes	Yes	Unclear	No	Yes	No	No relevant outcomes Insuffient 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(6):478-82	Yes	No	Yes	No	Unclear	Yes	Case-control study (benign vs borderline) No relevant intervention RMI 4
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(3):397-402	Yes	No	Yes	No	Yes	Yes	No relevant intervention IOTA simple rules in combination with other rules not included in this assessment Selected population (not classifiable using IOTA simple

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
							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(6):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(1):109-12	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 200
Anton C, Carvalho FM, Oliveira EI, Maciel G, Baracat EC, Carvalho JP. Comparison of four methods for classification of ovarian masses using ca 125, HE4, risk of malignancy index, and roma. *Int J Gynecol Cancer* 2011;Conference End: 20110914. Conference Publication:(var.pagings). 21 (12 SUPPL. 3):S658	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (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 (Sao Paulo)* 2012;67(5):437-41	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (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;Conference End: 20110914. Conference Publication:(var.pagings). 21 (12 SUPPL. 3):S820	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Accuracy of RMI 200
Araujo KG, Jales RM, Pereira PN, Yoshida A, de Angelo Andrade L, Sarian LO, et al. Performance	Yes	Yes	No	Yes	Yes	Yes	Tertiary care gynecologic oncology center

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
of the IOTA ADNEX model in the preoperative discrimination of adnexal masses in a gynecologic oncology center. *Ultrasound Obstet Gynecol* 2016;19:19							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(6):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(7):1727-30	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Threshold optimisation study for RMI, data for cut-off of 238
Ashrafganjooei T. Risk of malignancy index in evaluation of pelvic masses. *Int J Gynecol Cancer* 2011;Conference End: 20110914. Conference Publication:(var.pagings). 21 (12 SUPPL. 3):S673	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;Conference End: 20110403. Conference Publication:(var.pagings). 21 (11 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(1):75-80	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Validation of regression models (not iinterventions 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;Conference End: 20140626. Conference Publication: (var.pagings). 52:S365	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
Bailey J, Tailor A, Naik R, Lopes A, Godfrey K, Hatem HM, et al. Risk of malignancy index for	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI 200 in a

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
referral of ovarian cancer cases to a tertiary center: does it identify the correct cases? *Int J Gynecol Cancer* 2006;16 Suppl 1:30-4							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(9):1410-4	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI 200 in a tretiary care setting
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;Conference End: 20141111. Conference Publication:(var.pagings). 24 (9 SUPPL. 4):366-367	Yes	Yes	Unclear	Yes	Yes	No	No relevant outcomes Insuffient information to calculate sensitivity and specificity reported
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	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
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;Conference End: 20161031. 26:21-22	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
Blontzos N, Vorgias G, Papatheodorou D, Vylliotou V, Novkovic N, Diakosavas M, et al. The clinical value of adding HE4 and ROMA index to	Yes	Yes	Unclear	No	Yes	No	No relevant intervention ROMA assays and threshold not reported

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
CA-125 in the preoperative workout of adnexal masses. *Int J Gynecol Cancer* 2016;Conference End: 20161031. 26:172							
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(6):581.e1-581.e8	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;Conference End: 20091009. Conference Publication: (var.pagings). 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 Tumors in Women With Adnexal Masses. *J Ultrasound Med* 2016;35(1):143-52	Yes	Yes	No	Yes	Yes	Yes	Accuracy of RMI 1 in a tertiary care setting
Chia YN, Marsden DE, Robertson G, Hacker NF. Triage of ovarian masses. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2008;48(3):322-8	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI 200 in a tretiary 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. *Journal of Obstetrics & Gynaecology of India* 2015;65(6):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	Yes	Yes	Unclear	No	Unclear	Accuracy	No relevant intervention ROMA using different

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
ovarian tumor patients better when being based on specific age ranges instead of the premenopausal and postmenopausal status? *Tumour Biol* 2016;37(7):8879-87							manufacturers' assays for CA125 and HE4 (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(5):3101-3114	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (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. *Journal of Obstetrics and Gynaecology Canada* 2009;31(5):440-445	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI 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(1):100-6	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA data set used to evaluate performance of different refression models
Dasari P, Catherine Leela Pannirselvan P, Sridhar MG. Ultrasonographic scoring and risk of malignancy index in preoperative prediction of ovarian malignancy. *J Gynecol Surg* 2013;29(2):61-64	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 2
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;Conference End: 20110914.	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200, in a tertiary care setting

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Conference Publication:(var.pagings). 21 (12 SUPPL. 3):S1269							
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;Conference End: 20121016. Conference Publication: (var.pagings). 22:E317	Yes	No	Unclear	No	Unclear	Yes	Case-control study No relevant intervention ROMA assays and threshold not reported
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(1):131-8	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI 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(4):2177-83	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 200
Evelyne M, Jeroen K, Roy K, Arnold-Jan K, Brigitte S, Ben Van C, et al. Subjective assessment of grey scale and color Doppler ultrasound features versus the International Ovarian Tumor 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 registratuion 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(6):e17185	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (not a valid CE marked intervention)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
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;Conference End: 20160622. 123:83-84	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. Available from: https://ClinicalTrials.gov/show/NCT01466049	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. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2008;48(2):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.	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*	Yes	No	No	Yes	N/A	Yes	Not a clinical study in patients with suspected ovarian cancer IOTA training study, using video clips

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Biol 2013;171(1):157-60							
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;Conference End: 20110403. Conference Publication:(var.pagings). 21 (11 SUPPL. 2):117	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI (unspecified threshold)
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(7):820-2	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 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(2):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(6):567-74	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Regression model, including multiple biomarkers + 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;Conference End: 20130629. Conference Publication:(var.pagings). 131 (1):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	Yes	Yes	No	No	Unclear	No	No relevant intervention Tertiary care, ROMA asays and

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
value of HE4, CA125 and the ROMA index in ovarian cancer patients from a tertiary center. *Int J Gynecol Cancer* 2012;Conference End: 20111130. Conference Publication: (var.pagings). 22:S42							threshold not reported, RMI thresold not reported
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;Conference End: 20131022. Conference Publication:(var.pagings). 23 (8 SUPPL. 1):905	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
Imperial, Angeli N, Rivera, Wilhelmina. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *J Obstet Gynaecol Res* 2015;Conference End: 20150606. Conference Publication: (var.pagings). 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;Conference End: 20150415. Conference Publication: (var.pagings). 122:137	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200 and optimised threshold (273)
Imperial NA, Rivera W. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. *Int J Gynaecol Obstet* 2015;Conference End: 20151009. Conference Publication: (var.pagings). 131:E412	Yes	Yes	No	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200
Irshad F, Irshad M, Naz M, Asim Ikram M. Accuracy of "risk of malignancy index" in the preoperative diagnosis of Zovarian malignancy in	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 250

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
post menopausal women. *Rawal Medical Journal* 2013;38(3):266-270							
Jabeen R, Khan SA, Naveed S. Risk of Malignancy index in the preoperative evaluation of patients with ovarian masses. *Rawal Medical Journal* 2015;40(1):78-80	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200
Jacob F, Meier M, Caduff R, Goldstein D, Pochechueva T, Hacker N, et al. No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. *Gynecol Oncol* 2011;121(3):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(4):392	No	Yes					Not a primary study
Javdekar R, Maitra N. Risk of Malignancy Index (RMI) in Evaluation of Adnexal Mass. *Journal of Obstetrics & Gynaecology of India* 2015;65(2):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;Conference End: 20121016. Conference Publication: (var.pagings). 22:E149-E150	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
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	No	Yes					Not a primary study

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
biomarkers or the ROMA algorithm. *Gynecol Oncol* 2013;130(2):398-9							
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(1):20	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (not a valid CE marked intervention)
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. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2010;50(1):77-80	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 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(2):238-44	Yes	Yes	Yes	No	Yes	No	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (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(7):1505-10	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 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*	Yes	Unclear	No	No	No	No	No relevant intervention Risk model development (Copenhagen Index) using data from existing studies and stored blood samples

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Oncol 2015;138(3):640-6							
Keogh F, Tan AL, Eva LJ. HE4 as a tumour marker for the prediction of ovarian carcinoma. *BJOG* 2015;Conference End: 20150415. Conference Publication: (var.pagings). 122:137-138	Yes	Yes	Unclear	No	Yes	No	No relevant intervention ROMA assays and threshold not reported
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;Conference End: 20151011. 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 Tumor Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. *Ultrasound Obstet Gynecol* 2015;45(2):238	No	Yes					Not a primary study
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(2):400	No	Yes					Not a primary study
Kondalsamy-Chennakesavan S, Hackethal A, Bowtell D, Australian Ovarian Cancer Study G, 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(3):467-71	Yes	No	Unclear	Yes	Yes	Yes	Duagnostic case-control study
Lasho MA, Algeciras-Schimnich A. Determination of ROMA score performance using the roche	Yes	Yes	Unclear	Yes	Unclear	Yes	Refernce standard unspecified

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
elecsys HE4 and CA 125 immunoassays. *Clin Chem* 2014;Conference End: 20140730. Conference Publication:(var.pagings). 60 (10 SUPPL. 1):S12-S13							
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:S22-30	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]. *Giornale Italiano di Ostetricia e Ginecologia* 2012;34(3):409-414	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 (not a valid CE marked intervention)
Loh AHP, Ong CL, Lam SL, Chua JHY, Chui CH. Risk of malignancy index for preoperative evaluation of pediatric ovarian tumors. *Pediatr Blood Cancer* 2010;Conference End: 20101024. Conference Publication:(var.pagings). 55 (5):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(2):248-52	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (not a valid CE marked intervention)
Lokich E, Palisoul M, Romano N, Stuckey AR,	Yes	Yes	No	No	Unclear	Yes	No relevant intervention

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Robison KM, DiSilvestro PA, et al. ROMA guided conservative management for women diagnosed with an ovarian cyst or pelvic mass. *Gynecol Oncol* 2015;Conference End: 20150331. Conference Publication: (var.pagings). 137:21							Tertiay 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;Conference End: 20130629. Conference Publication:(var.pagings). 131 (1):259	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Multivariate index assay (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(3):162-4	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200
Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. *Chin Med J* 2003;116(3):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;Conference End: 20140330. Conference Publication: (var.pagings). 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;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 SUPPL. 1):223	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Accuracy of RMI 200

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
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(5):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;Conference End: 20150625. Conference Publication: (var.pagings). 53:S424	Yes	Yes	Unclear	Yes	Unclear	No	No relevant outcomes Insufficient data to determine accuracy measures
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;Conference End: 20110914. Conference Publication:(var.pagings). 21 (12 SUPPL. 3):S385	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. *Journal of the Turkishgerman Gynecological Association* 2010;11(1):22-6	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 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;Conference End: 20150415. Conference	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Publication: (var.pagings). 122:144-145							
Mohammed ABF, Ahuga VK, Taha M. Validation of the Risk of Malignancy Index in primary evaluation of ovarian masses. *Middle East Fertility Society Journal* 2014;19(4):324-328	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(2):162-7	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Validation study of 21 published models(not iinterventions 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;Conference End: 20140318. Conference Publication: (var.pagings). 35:S9	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
Molina R, Escudero JM, Fuste P. HE-4 utility to increase efficiency in patients with abdominal masses. *Tumor Biol* 2014;Conference End: 20140318. Conference Publication: (var.pagings). 35:S6	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
Moolthiya W, Yuenyao P. The risk of malignancy index (RMI) in diagnosis of ovarian malignancy. *Asian Pac J Cancer Prev* 2009;10(5):865-8	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 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;Conference End: 20090208. Conference Publication:(var.pagings). 112 (2	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention RMI (unspecified threshold)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
SUPPL. 1):S25-S26							
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(3):228.e1-6	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (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;Conference End: 20131022. Conference Publication:(var.pagings). 23 (8 SUPPL. 1):64	Yes	Yes	No	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
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(3):547-51	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (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(12):892-9	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention ROMA using different manufacturers' assays for CA125 and HE4 (not a valid CE marked intervention) IOTA data are for models other than simple rules or ADNEX
Nahar S, Shamsuddin L, Faruqui M, Ara G. Sonographic prediction of ovarian malignancy in adnexal mass. *Bangladesh Journal of Obstetrics and Gynecology* 2012;27(2):67-71	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention
Numanoglu C, Kuru O, Sakinci M, Akbayir O, Ulker	Yes	Yes	Unclear	No	Yes	Yes	RMI 200, data for a small

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
V. Ovarian fibroma/fibrothecoma: retrospective cohort study shows limited value of risk of malignancy index score. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2013;53(3):287-92							subgroup of patients with fibroma/fibrothecoma
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(3):136-9	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Data for various RMI thresholds (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(1):345-9	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 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 ovairan tumor. *Int J Gynecol Cancer* 2015;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 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. *Journal of Gynecologic Oncology* 2011;22(4):244-52	Yes	Yes	No	Yes	Yes	Yes	Tertiary care setting gynecologic 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 Tumor Analysis (IOTA) Group: a prospective study of combined	Yes	No	Yes	Yes	Yes	No	No relevant outcomes Selected population (un-classifiable using IOTA simple descriptors)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
use for the description of adnexal masses. *Eur J Obstet Gynecol Reprod Biol* 2015;195:7-11							
Pineda L, Salcedo E, Vilhena C, Juez L, Alcazar JL. Interobserver agreement in assigning IOTA color score to adnexal masses using three-dimensional volumes or digital videoclips: potential implications for training. *Ultrasound Obstet Gynecol* 2014;44(3):361-4	Yes	No	No	No	N/A	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 using different manufacturers' assays for CA125 and HE4 (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;Conference End: 20141111. Conference Publication:(var.pagings). 24 (9 SUPPL. 4):211-212	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 2
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;Conference End: 20141111. Conference Publication:(var.pagings). 24 (9 SUPPL. 4):486-487	Yes	Yes	Unclear	No	No	Yes	No relevant intervention ROMA assays and threshold not reported
Raza A, Mould T, Wilson M, Burnell M, Bernhardt L. Increasing the effectiveness of referral of ovarian masses from cancer unit to cancer center	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 450

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
by using a higher referral value of the risk of malignancy index. *Int J Gynecol Cancer* 2010;20(4):552-4							
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;Conference End: 20150415. Conference Publication: (var.pagings). 122:150	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Assay details not reported
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. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2015;55(5):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;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 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	Yes	Yes	Unclear	Yes	Unclear	No	No relevant outcomes

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
study. *Int J Gynecol Cancer* 2015;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 SUPPL. 1):528-529							
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(1):84-8	Yes	Yes	Yes	No	Unclear	Yes	No relevant intervention
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(1):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(2):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(1):140-6	Yes	No	Unclear	No	Yes	Yes	No relevant intervention No data for IOTA simple rules alone Selected population (un-classifiable using IOTA 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	Yes	Yes	No	No	N/A	Yes	No relevant intervention IOTA regression model (not simple rules or ADNEX)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
phase 4 IOTA study. *BJOG* 2013;Conference End: 20130626. Conference Publication: (var.pagings). 120:357-358							
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(5):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;Conference End: 20151027. Conference Publication:(var.pagings). 25 (9 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 modeling versus magnetic resonance imaging for discriminating between benign and malignant adnexal masses: A prospective study. *Int J Gynecol Cancer* 2016;Conference End: 20161031. 26:820	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA regression model (not simple rules or ADNEX)
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(18):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*	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Optimised RMI threshold (163.85)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
2014;Conference End: 20141111. Conference Publication:(var.pagings). 24 (9 SUPPL. 4):348							
Sladkevicius P, Valentin L. Intra- and interobserver agreement when describing adnexal masses using the International Ovarian Tumor Analysis terms and definitions: a study on three-dimensional ultrasound volumes. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2013;41(3):318-327	Yes	Yes	Unclear	No	Yes	No	No relevant intervention IOTA models (not simple rules or ADNEX)
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;Conference End: 20121016. Conference Publication: (var.pagings). 22:E967-E968	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention Unspecified RMI
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 improves the diagnosis of patients with a pelvic mass. *Int J Gynecol Cancer* 2016;Conference End: 20161031. 26:169	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 200 + HE4
Sumpaico WW. Comparison of roma to rmi for ovarian carcinoma in asia. *Int J Gynaecol Obstet* 2012;Conference End: 20121012. Conference Publication: (var.pagings). 119:S248-S249	Yes	Yes	Unclear	No	Unclear	No	No relevant intervention ROMA assays and threshold not reported
Tanriverdi HA, Sade H, Akbulut V, Barut A, Bayar U. Clinical and ultrasonographic evaluation of pelvic masses. [Turkish]. *Journal of the Turkish German Gynecology Association* 2007;8(1):67-70	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 200
Terzic M, Dotlic J, Ladjevic IL, Atanackovic J, Ladjevic N. Evaluation of the risk malignancy	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 200

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
index diagnostic value in patients with adnexal masses. *Vojnosanit Pregl* 2011;68(7):589-93							
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 tumors. *Taiwan J Obstet Gynecol* 2013;52(2):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;Conference End: 20141128. Conference Publication: (var.pagings). 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(1):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(34):8794-801	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA models (not simple rules or ADNEX)
Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, et al. Inclusion of CA-125 does not improve mathematical models	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention IOTA model (not simple rules or ADNEX)

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
developed to distinguish between benign and malignant adnexal tumors. *J Clin Oncol* 2007;25(27):4194-200							
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(2):226-34	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Validation of other IOTA models (not simple rules or ADNEX)
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(4):424-37	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Development and validation of the IOTA Simple 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(3):448-52	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention RMI 2
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;Conference End: 20161118. 23(7 Supplement 1):S217-S218	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. *Sao Paulo Med J* 2002;120(3):72-6	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Not RMI 1

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
Trevino-Baez JD, Cantu-Cruz JA, Medina-Mercado J, Abundis A. [Diagnostic accuracy of malignancy risk index II in post-menopausal women with adnexal tumours]. *Cir Cir* 2016;84(2):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. 2016. Available from: https://ClinicalTrials.gov/show/NCT01698632	No	Yes	Unclear	No	N/A	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(4):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 gray-scale and Doppler ultrasound findings: logistic regression models do not help. *Ultrasound Obstet Gynecol* 2011;38(4):456-65	Yes	No	Unclear	No	Yes	Yes	No relevant intervention Accuracy of RMI 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(5):570-81	Yes	Yes	Unclear	No	Yes	No	No relevant intervention Development of an IOTA model to predict malignancy in uniocular 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	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
serum CA-125. *J Natl Cancer Inst* 2007;99(22):1706-14							
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(6):662-71	Yes	Yes	Yes	No	Yes	No	No relevant intervention IOTA model (not simple rules or ADNEX)
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(3):384-8	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 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(15 Pt 1):4440-7	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention Accuracy of RMI 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 tumor analysis study. *Clin Cancer Res* 2009;15(2):684-91	Yes	Yes	Unclear	No	Yes	Yes	No relevant intervention Validation of IOTA models (not simple rules or ADNEX)
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*	Yes	Yes	Unclear	No	Yes	No	No relevant intervention Accuracy of RMI 200

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
2012;18(3):815-25							
Villiotou V, Vorgias G, Lekka I, Karampelas A, Dertimas V. Evaluation of HE4, CA 125 and ROMA predictive index in patients with gynecological diseases. *Clin Chem Lab Med* 2014;Conference End: 20140626. Conference Publication: (var.pagings). 52:S479	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA assays and threshold not reported
Wang LM, Song H, Song X, Zhou XB. An improved risk of malignancy index in diagnosis of adnexal mass. *Chin Med J* 2012;125(3):533-5	Yes	Yes	Yes	No	Yes	No	No relevant intervention Unspecified RMI
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(1):46-53	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention
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;Conference End: 20110914. Conference Publication:(var.pagings). 21 (12 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. *Clinical Ovarian and other Gynecologic Cancer* 2014;7(1-2):8-12	Yes	Yes	Yes	No	Yes	Yes	No relevant intervention Accuracy of RMI 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	Yes	No	Unclear	No	Yes	Yes	Not patients with suspected ovarian cancer

Study Details	Primary study	Population	Setting	Index Test	Reference Standard	Outcome	Reason for exclusion
adnexal masses in Asian and Pacific populations? *Asian Pac J Cancer Prev* 2013;14(9):5455-9							
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(3):320-4	Yes	Yes	No	No	Yes	Yes	No relevant intervention Accuracy of RMI 200 in a tretiary 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(3):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 Tumor Analysis terminology: reproducibility study involving seven observers. *Ultrasound Obstet Gynecol* 2014;44(1):100-8	No	No	No	No	N/A	No	Not a clinical study in patients with suspected ovarian cancer IOTA 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(8):513-17	Yes	Yes	Unclear	No	Unclear	Yes	No relevant intervention ROMA assays and threshold not reported

APPENDIX 6: COSTS CALCULATIONS FOR RISK SCORES

Risk assessment tool cost calculations

Test	Test cost per kit (£)	Sum of HE4 test-related costs (capital, other, personnel as per below) (£)	Ultrasound (£)	CA125 (£)	Total cost for risk score (£)
ROMA, Abbott ARCHITECT*	21.33	6.64	76.75	25.58	130.31
ROMA, Roche Elecsys*	15.95	7.81	76.75	25.58	126.09
Vermillion Overa (MIA2G)*	99.00	-	76.75	-	175.80
IOTA simple ultrasound rules	-	-	76.75	-	76.75
IOTA ADNEX model	-	-	76.75	25.58	102.34
RMI 1	-	-	76.75	25.58	102.34

*Manufacturers stated that final costs may be subject to volume-based discounts

Risk assessment tool components – cost breakdown

Risk assessment tool component	Cost pertest kit / ultrasound (£)
Serum CA125	25.58
Transvaginal ultrasound	76.75
Abbott ARCHITECT HE4	21.33
Roche Elecsys HE4	15.95
Vermillion Overa (MIA2G)	99.0

Capital cost calculation items for Abbott and Roche 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	£6,785.46	£1.92
Resale value	0	-	-
Lifetime of analyser equipment	10 years	-	-
Number of tests per year on one analyser (full capacity	3,542	-	-

Other cost items for Abbott and Roche HE4 tests	Cost item (£)	per year (£)	Cost per test (£)
Quality control Abbott (1.5 times per year)	87.52	131.28	0.04
Quality control Roche (12 times per year)	354.37	4,252.44	1.20
Maintenance (per year, but not in the first year), taken from Fujirebio and assumed to be the same for Abbott and Roche, in the absence of other information	3,819.51	3,437.56	0.97
Calibration (six time per year)	566.98	3,401.88	0.96

Abbott and Roche, taken from Roche, assumed to be the same for Abbott in the absence of other information			
Shipment (per month), taken from Fujirebio and assumed to be the same for Roche, in the absence of other information	0.26	3.12	0.001

Personnel cost items for Abbott and Roche HE4 tests	**Personnel cost items**	**Personnel cost 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 RMI 1 thresholds

Threshold	Sensitivity	Se	Specificity	Se	Source (systematic review, appendix 4d)
250	64.4%	1.4%	91.8%	0.7%	Summary estimate derived from all studies, 6 published studies[74, 75, 77, 79-81] and 1 un-published study that reported data for RMI 1 (threshold 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, 63, 74, 75, 77, 79-81, 99, 104] and 1 un-published study that reported data for RMI 1 (threshold 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[75, 77, 79-81]
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[74, 75, 77, 79-81]
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[74, 77, 79-81]
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[75, 77, 79-81]
300	53.8%	5.0%	98.8%	1.1%	Estimate from one RMI 1 threshold comparison study that reported data for this threshold[74]

Pre-menopausal subgroup

Test	Sensitivity	Se	Specificity	Se	Source (systematic review, 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 Abbott ARCHITECT	52.4%	11.4%	90.1%	2.7%	Sensitivity and specificity estimates taken from the only study to report sub-group data for the target condition 'all malignant tumours'[104] (see Table 8)
ROMA Roche Elecsys	90.0%	11.3%	82.0%	3.6%	Sensitivity and specificity estimates taken from the only study to report sub-group data for the target condition 'all malignant tumours'[98] (see Table 11)
Overa (MIA2G) Vermillion	90.3%	5.5%	71.4%	2.9%	Sensitivity and specificity estimates taken from the only study to report sub-group data for the target condition 'all malignant tumours'[71] (see Table

					18)
IOTA simple ultrasound rules (inconclusive results treated as malignant)	94.5%	1.1%	79.3%	1.1%	Summary estimate derived from the 4 studies that reported sub-group data for the target condition 'all malignant tumours'[44, 49, 50, 63] (see Table13)
IOTA ADNEX model	97.0%[a]	2.9%	71.0%	4.8%	Sensitivity and specificity estimates taken from the only study to report sub-group data for the target condition 'all malignant tumours'[44] (see Table 12)
RMI 1 threshold 200	53.3%	2.3%	93.5%	0.7%	Summary estimate derived from all 5 studies that reported sub-group data for the target condition 'all malignant tumours'[44, 50, 63, 99, 104]

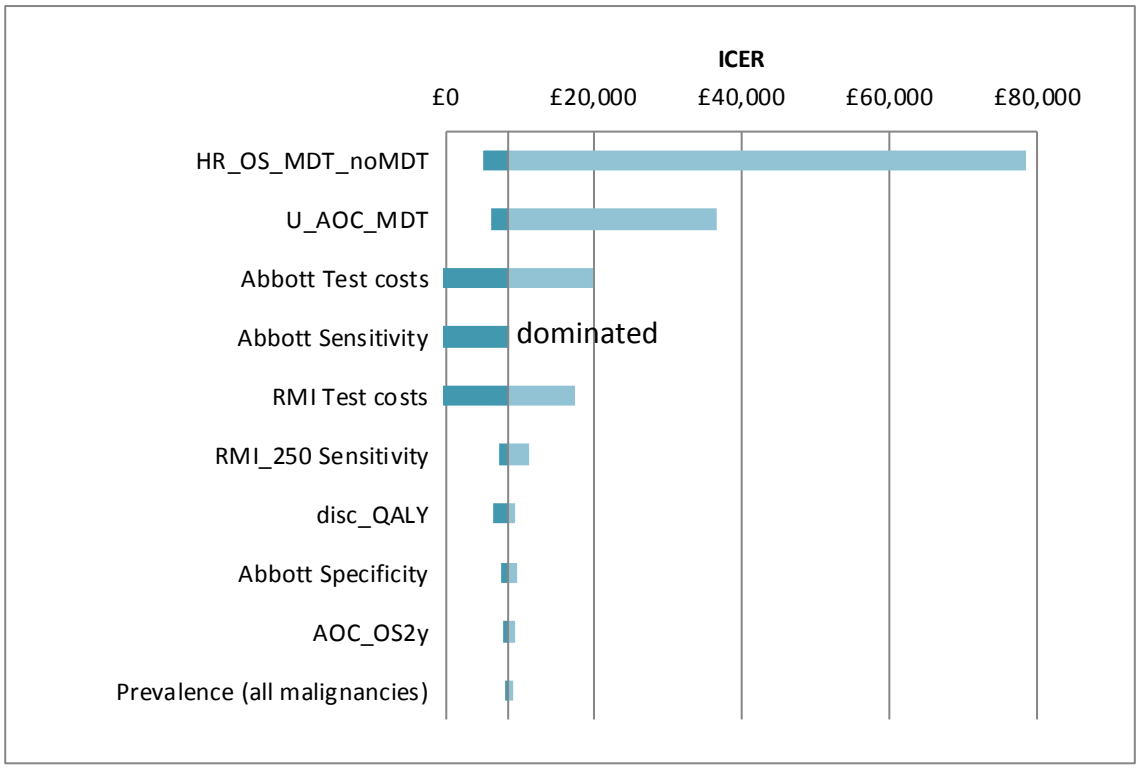
[a] A weakly informative prior (alpha = 1; beta =1) was used to handle 'zero events/counts issue' in the data obtained from Meys 2016 (indicating 100% sensitivity).

Post-menopausal subgroup

Test	Sensitivity	Se	Specificity	Se	Source (systematic review, 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 Abbott ARCHITECT	92.6%	6.0%	79.2%	9.0%	Sensitivity and specificity estimates taken form the only study to report sub-group data for the target condition 'all malignant tumours'[104] (see Table 8)
ROMA Roche Elecsys	78.6%	5.8%	76.1%	6.1%	Sensitivity and specificity estimates taken form the only study to report sub-group data for the target condition 'all malignant tumours'[98] (see Table 11)
Overa Vermillion (MIA2G)	91.8%	3.6%	65.4%	3.8%	Sensitivity and specificity estimates taken form the only study to report sub-group data for the target condition 'all malignant tumours'[71] (see Table 18)
IOTA simple ultrasound rules (inconclusive results treated as malignant)	95.4%	0.8%	67.3%	1.9%	Summary estimate derived from the 4 studies that reported sub-group data for the target condition 'all malignant tumours'[44, 49, 50, 63] (see Table13)
IOTA ADNEX model	98.0%	2.3%	54.0%	4.8%	Sensitivity and specificity estimates taken form the only study to report sub-group 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 5 studies that reported sub-group data for RMI 1 (threshold 200) and the target condition 'all malignant tumours'[44, 50, 63, 99, 104]

APPENDIX 8: DETERMINISTIC ONE-WAY SENSITIVITY ANALYSES

Top 10 influential parameters in the comparison of the RMI 1 (threshold 250) versus ROMA Abbott ARCHITECT



Key for all figures:

HR_OS_SMDT_noSMDT: the overall survival hazard ratio for patients referred to SMDT compared to those not referred to SMDT

U_AOC_SMDT / U_EOC_SMDT: the utility associated with advanced / early ovarian cancer when patient had been referred to SMDT

disc_QALY / disc_Costs: the discount rate used for QALYs / costs

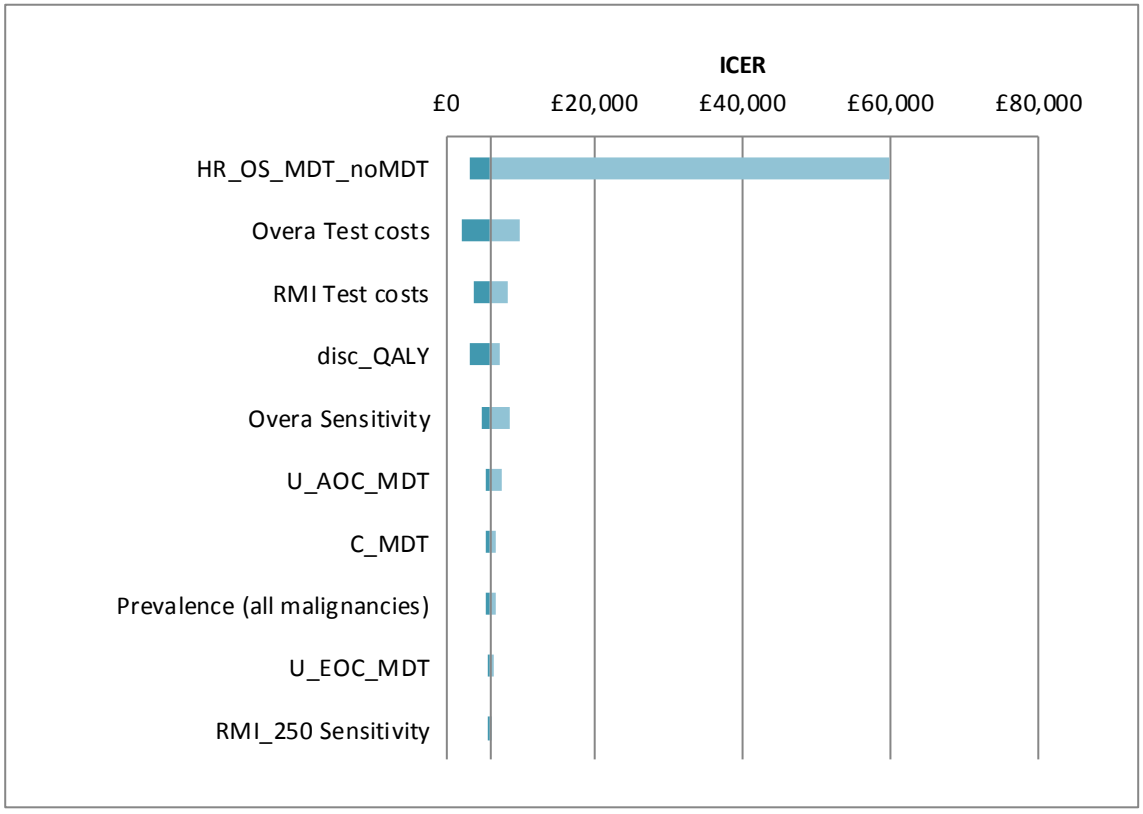
AOC_OS2y: the overall survival estimate for patients with advanced ovarian cancer at 2 years

EOC_OS10y: the overall survival estimate for patients with early ovarian cancer at 10 years

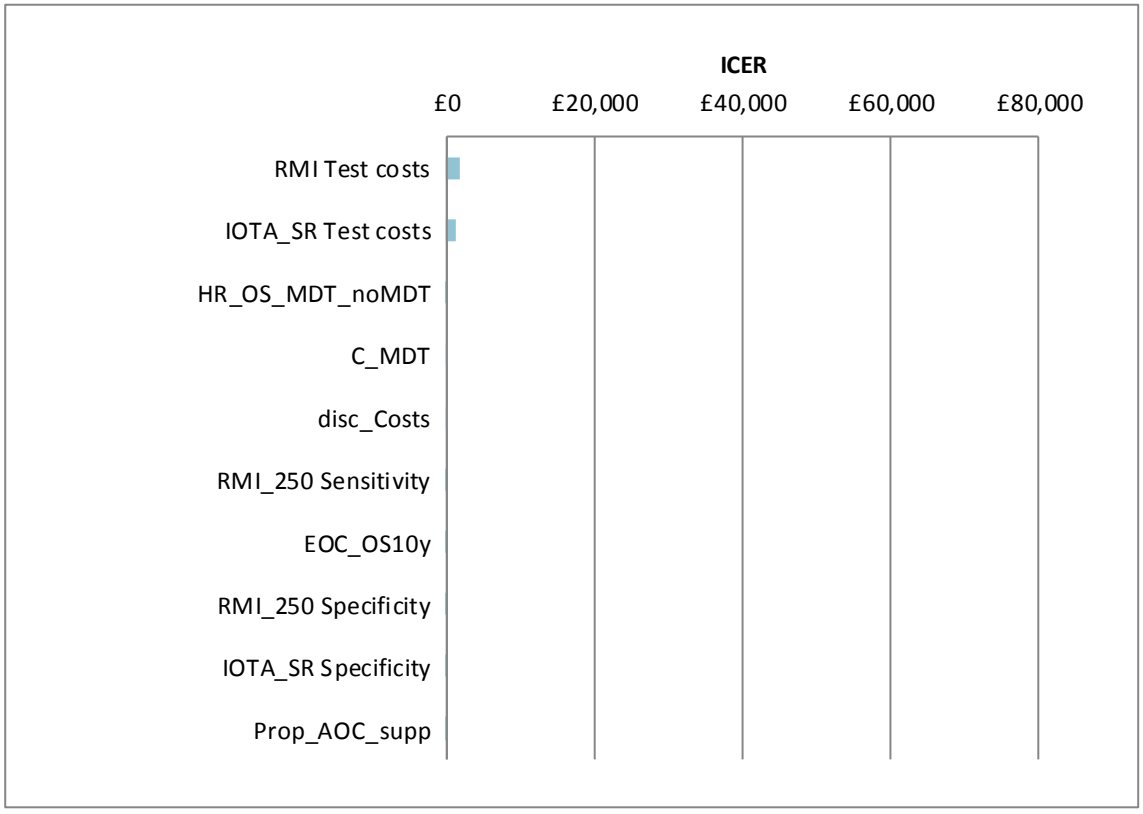
C_SMDT: Costs associated with SMDT

Prop_AOC_supp: Proportion of patients with advanced ovarian cancer receiving only supportive care (and no surgery)

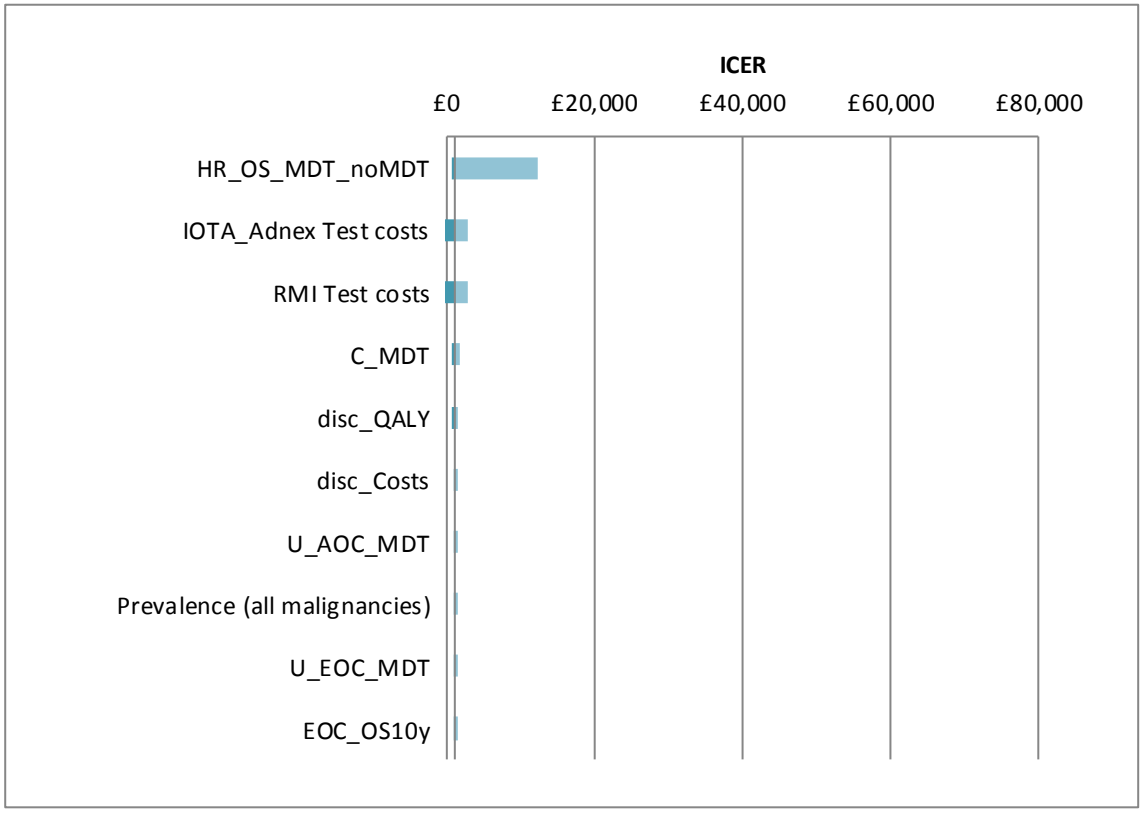
Top 10 influential parameters in the comparison of the RMI 1 (threshold 250) versus Overa (MIA2G) Vermillion (threshold 5 units)



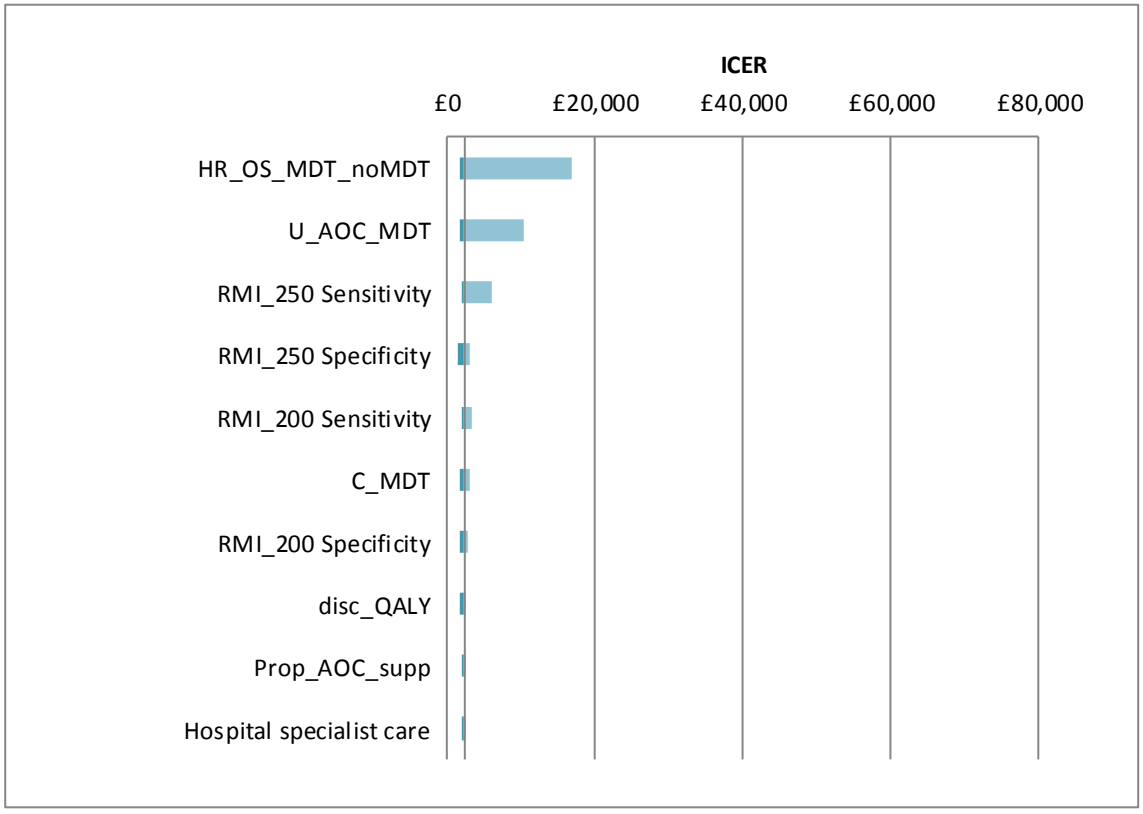
Top 10 influential parameters in the comparison of the RMI 1 (threshold 250) versus IOTA simple ultrasound rules (inconclusive assumed to be malignant)



Top 10 influential parameters in the comparison of the RMI 1 (threshold 250) versus IOTA ADNEX model (threshold 10%)



Top 10 influential parameters in the comparison of the RMI 1 (threshold 250) versus the RMI 1 (threshold 200)



APPENDIX 9: SCENARIO ANALYSES (DETERMINISTIC)

Assuming a prevalence of 20% for all malignancies

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,540	14.026	17.156	-£2	0.020	0.027	dominant	cheapest
RMI 1(threshold 250)	£5,542	14.006	17.129	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,546	14.007	17.131	£4	0.001	0.002	£2,511	dominated
IOTA ADNEX model (threshold 10%)	£5,572	14.028	17.158	£30	0.022	0.029	£1,333	£16,137
ROMA Abbott ARCHITECT	£5,580	14.010	17.136	£37	0.004	0.007	£9,008	dominated
ROMA Roche Elecsys	£5,586	14.012	17.138	£43	0.006	0.009	£7,169	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,648	14.022	17.151	£106	0.017	0.022	£6,385	dominated

Assuming a prevalence of 30% for all malignancies

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£6,432	12.709	15.626	-£1	0.032	0.041	dominant	cheapest
RMI 1 (threshold 250)	£6,433	12.677	15.585	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£6,438	12.679	15.589	£4	0.002	0.003	£2,059	dominated
IOTA ADNEX model (threshold 10%)	£6,464	12.712	15.630	£31	0.035	0.045	£888	£10,504
ROMA Abbott ARCHITECT	£6,473	12.683	15.595	£40	0.006	0.010	£6,408	dominated
ROMA Roche Elecsys	£6,478	12.687	15.600	£45	0.010	0.015	£4,466	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£6,539	12.703	15.619	£106	0.026	0.034	£4,105	dominated

Assuming 0% prevalence of non-ovarian malignancies

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,339	13.920	16.995	-£4	0.024	0.031	dominant	cheapest
RMI 1 (threshold 250)	£5,343	13.896	16.964	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,347	13.898	16.967	£4	0.002	0.002	£2,427	dominated
IOTA ADNEX model (threshold 10%)	£5,371	13.922	16.997	£28	0.026	0.033	£1,083	£15,094
ROMA Abbott ARCHITECT	£5,381	13.901	16.971	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,385	13.905	16.976	£42	0.009	0.012	£4,876	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,447	13.916	16.990	£104	0.020	0.026	£5,259	dominated

Assuming an equal proportion of early versus advanced stage ovarian cancer in the FN and TP groups (in the base case it was assumed that FN would predominantly/all be early stage)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 250)	£5,652	13.838	16.933	£0	0.000	0.000		cheapest
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,655	13.855	16.957	£2	0.017	0.024	£147	£147
RMI 1 (threshold 200)	£5,656	13.840	16.936	£3	0.002	0.003	£1,552	dominated
IOTA ADNEX model (threshold 10%)	£5,688	13.856	16.959	£35	0.018	0.026	£1,958	£27,656
ROMA Abbott ARCHITECT	£5,689	13.844	16.942	£36	0.006	0.009	£6,057	dominated
ROMA Roche Elecsys	£5,694	13.847	16.945	£42	0.008	0.012	£5,052	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,761	13.853	16.954	£109	0.015	0.021	£7,451	dominated

Assuming for IOTA simple ultrasound rules that subjective assessment would be used for inconclusive assessments (instead of assumed to be malignant)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,642	13.847	16.948	-£17	0.016	0.022	dominant	cheapest
RMI 1 (threshold 250)	£5,659	13.831	16.926	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,663	13.832	16.928	£4	0.002	0.002	£2,427	dominated
IOTA ADNEX model (threshold 10%)	£5,689	13.855	16.957	£30	0.024	0.031	£1,249	£5,922
ROMA Abbott ARCHITECT	£5,697	13.835	16.933	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,703	13.837	16.936	£44	0.007	0.010	£6,625	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,765	13.849	16.950	£106	0.018	0.024	£5,949	dominated

Assuming equal test costs for all risk scores

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 250)	£5,676	13.831	16.926	£0	0.000	0.000		cheapest
RMI 1 (threshold 200)	£5,680	13.832	16.928	£4	0.002	0.002	£2,427	ext dominated
ROMA Abbott ARCHITECT	£5,686	13.835	16.933	£10	0.004	0.007	£2,227	ext dominated
ROMA Roche Elecsys	£5,696	13.837	16.936	£20	0.007	0.010	£3,025	ext dominated
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,699	13.853	16.955	£23	0.022	0.029	£1,073	£1,073
IOTA ADNEX model (threshold 10%)	£5,706	13.855	16.957	£30	0.024	0.031	£1,249	£3,057

Overa (MIA2G) Vermillion (threshold 5 units)	£5,708	13.849	16.950	£33	0.018	0.024	£1,832	dominated

Assuming no ultrasound is performed in conjunction with ROMA and Overa (MIA2G) risk scores, thus reducing the costs of these risk scores

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
ROMA Abbott ARCHITECT	£5,621	13.835	16.933	-£39	0.004	0.007	-£8,759	cheapest
ROMA Roche Elecsys	£5,626	13.837	16.936	-£33	0.007	0.010	-£5,006	ext dominated
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,657	13.853	16.955	-£2	0.022	0.029	-£96	£2,109
RMI 1 (threshold 250)	£5,659	13.831	16.926	£0	0.000	0.000	£0	dominated
RMI 1 (threshold 200)	£5,663	13.832	16.928	£4	0.002	0.002	£2,427	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,689	13.849	16.950	£29	0.018	0.024	£1,645	dominated
IOTA ADNEX model (threshold 10%)	£5,689	13.855	16.957	£30	0.024	0.031	£1,249	£15,094

Assuming additional costs for FP (surgery costs with malignancy instead of without) and additional costs for FN (additional costs of benign surgery)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,759	13.853	16.955	-£174	0.022	0.029	-£7,986	cheapest
IOTA ADNEX model (threshold 10%)	£5,793	13.855	16.957	-£140	0.024	0.031	-£5,829	£16,372
ROMA Roche Elecsys	£5,904	13.837	16.936	-£29	0.007	0.010	-£4,384	dominated
ROMA Abbott ARCHITECT	£5,905	13.835	16.933	-£28	0.004	0.007	-£6,261	dominated
RMI 1 (threshold 200)	£5,915	13.832	16.928	-£18	0.002	0.002	-£11,809	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,921	13.849	16.950	-£12	0.018	0.024	-£675	dominated
RMI 1 (threshold 250)	£5,933	13.831	16.926	£0	0.000	0.000	£0	dominated

Assuming additional costs for FP (surgery costs with malignancy instead of without) and additional costs for FN (additional costs of benign surgery and SMDT costs)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,760	13.853	16.955	-£182	0.022	0.029	-£8,322	cheapest
IOTA ADNEX model (threshold 10%)	£5,794	13.855	16.957	-£147	0.024	0.031	-£6,158	£16,128
ROMA Roche Elecsys	£5,909	13.837	16.936	-£32	0.007	0.010	-£4,935	dominated
ROMA Abbott ARCHITECT	£5,911	13.835	16.933	-£30	0.004	0.007	-£6,851	dominated
RMI 1 (threshold 200)	£5,923	13.832	16.928	-£19	0.002	0.002	-£12,400	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,923	13.849	16.950	-£18	0.018	0.024	-£1,033	dominated
RMI 1 (threshold 250)	£5,942	13.831	16.926	£0	0.000	0.000	£0	dominated

Assuming a discount of 92% for carboplatin (CG122: discount England 91.8%, discount Wales 92.1%)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,363	13.853	16.955	-£1	0.022	0.029	dominant	cheapest
RMI 1 (threshold 250)	£5,364	13.831	16.926	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,368	13.832	16.928	£4	0.002	0.002	£2,427	dominated
IOTA ADNEX model (threshold 10%)	£5,395	13.855	16.957	£31	0.024	0.031	£1,280	£15,136
ROMA Abbott ARCHITECT	£5,402	13.835	16.933	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,408	13.837	16.936	£44	0.007	0.010	£6,629	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,470	13.849	16.950	£106	0.018	0.024	£5,977	dominated

Assuming a discount of 95% for paclitaxel (CG122: discount England 91.0%, discount Wales 95.4%)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,102	13.853	16.955	-£1	0.022	0.029	dominant	cheapest
RMI 1 (threshold 250)	£5,103	13.831	16.926	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,107	13.832	16.928	£4	0.002	0.002	£2,427	dominated
IOTA ADNEX model (threshold 10%)	£5,134	13.855	16.957	£32	0.024	0.031	£1,318	£15,186
ROMA Abbott ARCHITECT	£5,141	13.835	16.933	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,146	13.837	16.936	£44	0.007	0.010	£6,635	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,210	13.849	16.950	£107	0.018	0.024	£6,010	dominated

Assuming alternative HR for progression-free and overall survival for SMDT versus no SMDT (of 0.808)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,658	13.847	16.948	£0	0.043	0.057	dominant	cheapest
RMI 1 (threshold 250)	£5,659	13.804	16.891	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,664	13.807	16.896	£5	0.003	0.005	£1,664	dominated
IOTA ADNEX model (threshold 10%)	£5,690	13.851	16.953	£31	0.047	0.062	£660	£7,464
ROMA Abbott ARCHITECT	£5,700	13.812	16.905	£41	0.009	0.014	£4,812	dominated
ROMA Roche Elecsys	£5,707	13.816	16.911	£48	0.013	0.020	£3,760	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,767	13.839	16.938	£108	0.035	0.047	£3,078	dominated

Assuming alternative HR for progression-free and overall survival for SMDT versus no SMDT (of 0.990)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,656	13.858	16.961	-£4	0.002	0.003	dominant	cheapest
RMI 1 (threshold 250)	£5,660	13.856	16.958	£0	0.000	0.000		dominated
RMI 1 (threshold 200)	£5,663	13.856	16.959	£3	0.000	0.000	£16,921	dominated
IOTA ADNEX model (threshold 10%)	£5,689	13.858	16.961	£29	0.002	0.003	£12,374	£158,980
ROMA Abbott ARCHITECT	£5,694	13.856	16.959	£34	0.000	0.001	£78,602	dominated
ROMA Roche Elecsys	£5,699	13.856	16.959	£39	0.001	0.001	£60,637	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,763	13.858	16.961	£103	0.002	0.002	£60,005	dominated

Assuming that the proportion of patients receiving supportive care (for advanced stage) is 10% (instead of 5%)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 250)	£5,780	13.831	16.926	£0	0.000	0.000	£0	cheapest
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,781	13.853	16.955	£1	0.022	0.029	£37	£37
RMI 1 (threshold 200)	£5,784	13.832	16.928	£5	0.002	0.002	£3,005	dominated
IOTA ADNEX model (threshold 10%)	£5,813	13.855	16.957	£33	0.024	0.031	£1,368	£15,066
ROMA Abbott ARCHITECT	£5,820	13.835	16.933	£40	0.004	0.007	£9,106	dominated
ROMA Roche Elecsys	£5,827	13.837	16.936	£47	0.007	0.010	£7,134	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,889	13.849	16.950	£109	0.018	0.024	£6,119	dominated

Assuming alternative transvaginal ultrasound of £142.46 (MA36Z) (instead of £76.75 based on CG122)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,723	13.853	16.955	-£2	0.022	0.029	-£96	cheapest
RMI 1 (threshold 250)	£5,725	13.831	16.926	£0	0.000	0.000	£0	dominated
RMI 1 (threshold 200)	£5,729	13.832	16.928	£4	0.002	0.002	£2,427	dominated
IOTA ADNEX model (threshold 10%)	£5,755	13.855	16.957	£30	0.024	0.031	£1,249	£15,094
ROMA Abbott ARCHITECT	£5,763	13.835	16.933	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,768	13.837	16.936	£44	0.007	0.010	£6,625	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,831	13.849	16.950	£106	0.018	0.024	£5,949	dominated

Assuming alternative transvaginal ultrasound of £142.46 (MA36Z) (instead of £76.75 based on CG122) and increasing the transvaginal ultrasound for the IOTA risk scores by 20% (to reflect potential training costs)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 250)	£5,725	13.831	16.926	£0	0.000	0.000	£0	cheapest
RMI 1 (threshold 200)	£5,729	13.832	16.928	£4	0.002	0.002	£2,427	ext dominated
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,751	13.853	16.955	£26	0.022	0.029	£1,206	£1,206
ROMA Abbott ARCHITECT	£5,763	13.835	16.933	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,768	13.837	16.936	£44	0.007	0.010	£6,625	dominated
IOTA ADNEX model (threshold 10%)	£5,783	13.855	16.957	£58	0.024	0.031	£2,435	£15,094
Overa (MIA2G) Vermillion (threshold 5 units)	£5,831	13.849	16.950	£106	0.018	0.024	£5,949	dominated

Assuming additional costs SMDT costs of £2,500 to reflect higher surgery costs

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 250)	£6,162	13.831	16.926	£0	0.000	0.000	£0	cheapest
RMI 1 (threshold 200)	£6,219	13.832	16.928	£57	0.002	0.002	£36,724	ext dominated
ROMA Abbott ARCHITECT	£6,333	13.835	16.933	£171	0.004	0.007	£38,526	ext dominated
ROMA Roche Elecsys	£6,533	13.837	16.936	£371	0.007	0.010	£56,351	ext dominated
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£6,627	13.853	16.955	£464	0.022	0.029	£21,275	£21,275
IOTA ADNEX model (threshold 10%)	£6,807	13.855	16.957	£645	0.024	0.031	£26,929	£85,145
Overa (MIA2G) Vermillion (threshold 5 units)	£6,915	13.849	16.950	£753	0.018	0.024	£42,337	dominated

Assuming 90% of the non-malignancy surgery and complications costs for TN reflecting a scenario wherein 90% of the TN are operated (instead of all).

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 250)	£5,419	13.831	16.926	£0	0.000	0.000	£0	cheapest
RMI 1 (threshold 200)	£5,427	13.832	16.928	£8	0.002	0.002	£5,311	ext dominated
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,458	13.853	16.955	£39	0.022	0.029	£1,791	£1,791
ROMA Abbott ARCHITECT	£5,467	13.835	16.933	£48	0.004	0.007	£10,837	dominated
ROMA Roche Elecsys	£5,496	13.837	16.936	£77	0.007	0.010	£11,685	dominated
IOTA ADNEX model (threshold 10%)	£5,508	13.855	16.957	£89	0.024	0.031	£3,735	£23,755
Overa (MIA2G) Vermillion (threshold 5 units)	£5,593	13.849	16.950	£174	0.018	0.024	£9,783	dominated

Assuming avastin for advanced stage.

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£8,412	13.887	17.010	-£9	0.022	0.029	-£406	cheapest
RMI 1 (threshold 250)	£8,421	13.865	16.980	£0	0.000	0.000	£0	dominated
RMI 1 (threshold 200)	£8,425	13.867	16.983	£4	0.002	0.003	£2,284	dominated
IOTA ADNEX model (threshold 10%)	£8,443	13.889	17.012	£22	0.024	0.032	£904	£14,728
ROMA Abbott ARCHITECT	£8,459	13.870	16.988	£38	0.005	0.008	£7,894	dominated
ROMA Roche Elecsys	£8,464	13.872	16.991	£44	0.007	0.011	£6,151	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£8,522	13.883	17.005	£101	0.018	0.024	£5,526	dominated

Assuming a disutility for FP during the first year in state transition model of 0.100.

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,657	13.835	16.955	-£2	0.010	0.029	-£207	cheapest
RMI 1 (threshold 250)	£5,659	13.825	16.926	£0	0.000	0.000	£0	dominated
RMI 1 (threshold 200)	£5,663	13.825	16.928	£4	0.000	0.002	£13,465	dominated
IOTA ADNEX model (threshold 10%)	£5,689	13.832	16.957	£30	0.007	0.031	£4,257	dominated
ROMA Abbott ARCHITECT	£5,697	13.826	16.933	£38	0.002	0.007	£24,820	dominated
ROMA Roche Elecsys	£5,703	13.822	16.936	£44	-0.003	0.010	-£15,114	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,765	13.823	16.950	£106	-0.002	0.024	-£66,380	dominated

Assuming a disutility for FP during the first year in state transition model of 0.010.

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,657	13.851	16.955	-£2	0.021	0.029	-£102	cheapest
RMI 1 (threshold 250)	£5,659	13.830	16.926	£0	0.000	0.000	£0	dominated
RMI 1 (threshold 200)	£5,663	13.832	16.928	£4	0.001	0.002	£2,644	dominated
IOTA ADNEX model (threshold 10%)	£5,689	13.852	16.957	£30	0.022	0.031	£1,344	£20,023
ROMA Abbott ARCHITECT	£5,697	13.834	16.933	£38	0.004	0.007	£9,126	dominated
ROMA Roche Elecsys	£5,703	13.836	16.936	£44	0.006	0.010	£7,738	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,765	13.846	16.950	£106	0.016	0.024	£6,676	dominated

Comparison of different RMI 1 thresholds

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
RMI 1 (threshold 300)	£5,647	13.826	16.919	-£13	-0.004	-0.007	£2,865	cheapest
RMI 1 (threshold 250)	£5,659	13.831	16.926	£0	0.000	0.000		ext dominated
RMI 1 (threshold 200)	£5,663	13.832	16.928	£4	0.002	0.002	£2,427	ext dominated
RMI 1 (threshold 150)	£5,664	13.834	16.932	£4	0.004	0.006	£1,172	ext dominated
RMI 1 (threshold 100)	£5,671	13.838	16.937	£11	0.007	0.011	£1,619	ext dominated
RMI 1 (threshold 50)	£5,690	13.848	16.949	£30	0.017	0.023	£1,783	£2,006
RMI 1 (threshold 25)	£5,706	13.853	16.956	£46	0.023	0.030	£2,051	£2,890

Using the most optimal RMI threshold (i.e. RMI threshold cost effective at £20,000 and/or £30,000 per QALY gained in former scenario)

	Discounted			Compared with standard RMI				**Full incremental**
	Costs	QALYs	LYs	ΔCosts	ΔQALYs	ΔLYs	ICER	
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,657	13.853	16.955	-£2	0.022	0.029	dominant	cheapest
RMI 1 (threshold 250)	£5,659	13.831	16.926	£0	0.000	0.000		dominated
IOTA ADNEX model (threshold 10%)	£5,689	13.855	16.957	£30	0.024	0.031	£1,249	£15,094
ROMA Abbott ARCHITECT	£5,697	13.835	16.933	£38	0.004	0.007	£8,527	dominated
ROMA Roche Elecsys	£5,703	13.837	16.936	£44	0.007	0.010	£6,625	dominated
RMI 1 (threshold 25)	£5,706	13.853	16.956	£46	0.023	0.030	£2,051	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,765	13.849	16.950	£106	0.018	0.024	£5,949	dominated

APPENDIX 10: ADDITIONAL SUBGROUP ANALYSES (PROBABILISTIC)

Table 48: Probabilistic results for base case analysis: costs, QALYs and incremental analysis (subgroup aged 50 year)

Risk scores	Costs (95% CI)	QALYs (95% CI)	Compared to RMI 250			Full incremental
			ΔCosts	ΔQALYs	ΔCosts / ΔQALYs	ΔCosts / ΔQALYs
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,652 (95% CI: £4,544 to £6,922)	11.640 (95% CI: 11.306 to 11.911)	-£3	0.020	dominant	cheapest
RMI 1 (threshold 250)	£5,654 (95% CI: £4,542 to £6,924)	11.621 (95% CI: 11.287 to 11.891)				dominated
RMI 1 (threshold 200)	£5,658 (95% CI: £4,545 to £6,929)	11.622 (95% CI: 11.287 to 11.893)	£4	0.001	£2,561	dominated
IOTA ADNEX model (threshold 10%)	£5,684 (95% CI: £4,574 to £6,958)	11.642 (95% CI: 11.308 -to11.912)	£29	0.021	£1,371	£17,212
ROMA Abbott ARCHITECT	£5,691 (95% CI: £4,584 to £6,962)	11.625 (95% CI: 11.291 to 11.897)	£37	0.005	£7,719	dominated
ROMA Roche Elecsys	£5,698 (95% CI: £4,588 to £6,979)	11.627 (95% CI: 11.291 to 11.899)	£43	0.007	£6,657	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,760 (95% CI: £4,638 to £7,035)	11.637 (95% CI: 11.302 to 11.907)	£106	0.016	£6,602	dominated

Table 49: Probabilistic results for base case analysis: costs, QALYs and incremental analysis (subgroup of early stage ovarian cancer)

Risk scores	Costs (95% CI)	QALYs (95% CI)	Compared to RMI 250			Full incremental
			ΔCosts	ΔQALYs	ΔCosts / ΔQALYs	ΔCosts / ΔQALYs
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,460 (95% CI: £4,364 to £6,710)	14.711 (95% CI: 14.363 to 15.018)	-£10	0.029	dominant	cheapest
RMI 1 (threshold 250)	£5,470 (95% CI: £4,372 to £6,724)	14.681 (95% CI: 14.333 to 14.987)				dominated
RMI 1 (threshold 200)	£5,471 (95% CI: £4,373 to £6,726)	14.685 (95% CI: 14.337 to 14.991)	£2	0.004	£480	dominated
IOTA ADNEX model (threshold 10%)	£5,492 (95% CI: £4,387 to £6,750)	14.713 (95% CI: 14.365 to 15.019)	£22	0.031	£715	£15,631
ROMA Abbott ARCHITECT	£5,501 (95% CI: £4,403 to £6,754)	14.692 (95% CI: 14.343 to 14.999)	£32	0.010	£3,052	dominated
ROMA Roche Elecsys	£5,506 (95% CI: £4,404 to £6,757)	14.696 (95% CI: 14.348 to 15.003)	£36	0.014	£2,501	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,568 (95% CI: £4,469 to £6,825)	14.707 (95% CI: 14.359 to 15.013)	£98	0.025	£3,897	dominated

Table 50: Probabilistic results for base case analysis: costs, QALYs and incremental analysis (subgroup of advanced stage ovarian cancer)

Risk scores	Costs (95% CI)	QALYs (95% CI)	Compared to RMI 250			Full incremental
			ΔCosts	ΔQALYs	ΔCosts / ΔQALYs	ΔCosts / ΔQALYs
RMI 1 (threshold 250)	£5,719 (95% CI: £4,582 to £6,995)	13.556 (95% CI: 13.138 to 13.906)	£0	0.000	£0	cheapest
RMI 1 (threshold 200)	£5,712 (95% CI: £4,579 to £6,988)	13.544 (95% CI: 13.133 to 13.890)	£4	0.002		ext dominated
IOTA simple ultrasound rules (inconclusive assumed to be malignant)	£5,716 (95% CI: £4,583 to £6,992)	13.545 (95% CI: 13.134 to 13.893)	£7	0.012	£571	£571
ROMA Abbott ARCHITECT	£5,752 (95% CI: £4,621to £7,024)	13.556 (95% CI: 13.138 to 13.906)	£38	0.004	£8,837	dominated
IOTA ADNEX model (threshold 10%)	£5,750 (95% CI: £4,623 to £7,024)	13.548 (95% CI: 13.134 to 13.897)	£40	0.013	£3,104	£39,171
ROMA Roche Elecsys	£5,757 (95% CI: £4,623 to £7,028)	13.549 (95% CI: 13.135 to 13.898)	£45	0.006	£7,495	dominated
Overa (MIA2G) Vermillion (threshold 5 units)	£5,824 (95% CI: £4,693 to £7,105)	13.554 (95% CI: 13.137 to 13.904)	£112	0.010	£10,748	dominated