NICE National Institute for Health and Care Excellence

NICE guidance

Tests in secondary care to identify people at high risk of ovarian cancer

Diagnostics guidance
Published: 15 November 2017

www.nice.org.uk/guidance/dg31

Your responsibility

This guidance represents the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, healthcare professionals are expected to take this guidance fully into account, and specifically any special arrangements relating to the introduction of new interventional procedures. The guidance does not override the individual responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or guardian or carer.

All problems (adverse events) related to a medicine or medical device used for treatment or in a procedure should be reported to the Medicines and Healthcare products Regulatory Agency using the Yellow Card Scheme.

Commissioners and/or providers have a responsibility to implement the guidance, in their local context, in light of their duties to have due regard to the need to eliminate unlawful discrimination, advance equality of opportunity, and foster good relations. Nothing in this guidance should be interpreted in a way that would be inconsistent with compliance with those duties. Providers should ensure that governance structures are in place to review, authorise and monitor the introduction of new devices and procedures.

Commissioners and providers have a responsibility to promote an environmentally sustainable health and care system and should assess and reduce the environmental impact of implementing NICE recommendations wherever possible.

Contents

1 Recommendations

1.1 There is currently not enough evidence to recommend the routine adoption of the IOTA ADNEX model, Overa (MIA2G), RMI I (at thresholds other than 200 or 250), ROMA or IOTA Simple Rules in secondary care in the NHS to help decide whether to refer people with suspected ovarian cancer to a specialist multidisciplinary team (MDT).

1.2 The NICE guideline on ovarian cancer recommends that people with an RMI I of 250 or more are referred to a specialist MDT. Evidence suggests that there is no substantial change in accuracy if the threshold for RMI I is lowered to 200.

1.3 The IOTA ADNEX model, Overa (MIA2G), RMI I (at thresholds other than 250), ROMA and IOTA Simple Rules show promise. Further research is recommended on test accuracy and the impact of the test results on clinical decision-making (see section 6 for detailed research recommendations).

2 Clinical need and practice

The problem addressed

2.1 Tests and risk scores are used in secondary care to help determine if a person referred with suspected ovarian cancer is likely to have an ovarian malignancy. Results inform decisions about whether they should be referred to a specialist multidisciplinary team (MDT) for further assessment and treatment. Currently, serum biomarker CA125 and pelvic ultrasound scans are widely used in secondary care, as part of the risk of malignancy index 1 (RMI I) score, in deciding whether a referral to a specialist MDT is needed. However, not all ovarian malignancies show elevated CA125 levels (particularly early stage ovarian cancer). Also elevated levels of CA125 are not always indicative of ovarian cancer, because they may be raised from other causes, such as endometriosis, fibroids, pregnancy, pelvic inflammatory disease, liver disease or heart failure. Tests and risk scores included in this assessment (ADNEX, Overa [MIA2G], RMI I at thresholds other than 250, ROMA and Simple Rules) may be better able to distinguish between benign and malignant ovarian tumours, and increase the proportion of people with a correct referral from secondary care to a specialist MDT.

2.2 Increasing the proportion of people with ovarian cancer who get a correct referral to a specialist MDT is likely to improve patient outcomes. Also, improved testing could lead to more accurate recognition of people referred to secondary care with suspected ovarian cancer who do not have the condition. This could reduce inappropriate referrals to specialist care for further assessment and treatment, as well as the costs and anxiety that this can cause.

The condition

2.3 Ovarian cancer starts in cells in, or near, the ovaries. Primary ovarian tumours are classified based on the tissue that they develop from, with 3 main types: epithelial ovarian tumours, sex cord-stromal tumours of the

ovary and germ cell tumours of the ovary. Each subtype of tumour can be benign, malignant or intermediate (borderline malignant). About 90% of primary ovarian cancers are malignant epithelial tumours. Non-epithelial ovarian cancers make up a higher proportion of ovarian cancer in people who are premenopausal.

2.4 Data from Cancer Research UK (ovarian cancer statistics) suggests:

- There were about 7,400 new cases of ovarian cancer in the UK in 2014, accounting for 2% of all new cancer cases.
- The incidence of ovarian cancer increases with age, with more than half of cases between 2012 and 2014 happening in people aged 65 years and over.
- There were about 50 new cases in people under 19 years in this time period, about 600 new cases in people under 40 years and about 1,400 new cases in people under 50 years.

The diagnostics and care pathways

Diagnosis

2.5 The NICE guideline on ovarian cancer includes recommendations on criteria and tests to use in primary care when deciding whether to refer someone to secondary care with suspected ovarian cancer. Recommendations from this guideline have also been incorporated in the NICE guideline on suspected cancer.

2.6 The NICE guideline on ovarian cancer also provides recommendations on diagnosing suspected ovarian cancer in secondary care. An ultrasound of the abdomen and pelvis is recommended as the first imaging test in secondary care for people with suspected ovarian cancer (if this has not already been done in primary care), as well as measuring serum CA125 (if not already done in primary care). The guideline recommends calculating an RMI I score, based on characteristics seen on ultrasound, CA125 serum levels and menopausal status (described in more detail in section 3). It states that people with an RMI I score of 250 or more should be referred to a specialist MDT.

2.7 For people under 40 years with suspected ovarian cancer, the NICE guideline on ovarian cancer recommends measuring the levels of alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (beta-hCG), as well as CA125, to identify non-epithelial ovarian cancer.

2.8 The NICE guideline on ovarian cancer also provides recommendations on further imaging to characterise the extent and spread of ovarian cancer, and also on getting a tissue sample to confirm a diagnosis of ovarian cancer. Histopathology is generally used as the reference standard for assessing the accuracy of tests to identify people who are likely to have ovarian cancer. As well as distinguishing between malignant and benign tumours, this testing can also determine the type of ovarian cancer present. If tissue samples are not taken, clinical follow-up may be needed to determine the presence, or absence, of ovarian cancer.

Care pathway

2.9 The NICE guideline on ovarian cancer contains recommendations for the management of early (stage I) and advanced (stages II to IV) ovarian cancer.

3 The diagnostic tests

The assessment compared 5 interventions with 1 comparator.

The interventions

The assessment of different neoplasias in the adnexa (ADNEX) model

3.1 The ADNEX model was developed by the International Ovarian Tumor Analysis (IOTA) group to assess people with an adnexal mass who are considered to need surgery. The model uses 3 clinical predictors and 6 ultrasound-derived predictors to estimate the probability that a pelvic tumour is benign or malignant (see table 1). Also, the model estimates probabilities that a tumour is borderline, stage I cancer, stage II to IV cancer or secondary metastatic cancer. The ADNEX model formulas are available in published literature (Van Calster et al. 2014) and the model is further described on the IOTA website. The terminology used in the model is as defined in a publication by the IOTA group (Timmerman et al. 2000), and the group run courses that teach the terms, definitions and measurement techniques needed to assess pelvic masses for the ADNEX model. An online training tool for NHS practitioners is also currently in development.

Table 1 Criteria included in the ADNEX model

Clinical predictors	Ultrasound derived predictors

Age (years) Serum CA125 level (units per millilitre [U/ml]) Type of centre (oncology centre or other hospital)[1]	Maximum diameter of lesion (mm) Proportion of solid tissue (ratio of the maximum diameter of the largest solid component and the maximum diameter of the lesion) More than 10 cyst locules (yes or no) Number of papillary projections (0, 1, 2, 3 or more than 3) Acoustic shadows (yes or no) Ascites (yes or no)
[1] Oncology centre defined as a tertiary referral centre with a specific gynaecology oncology unit (Van Calster et al. 2014).	

3.2 The ultrasound variables for the ADNEX model need B mode imaging and the IOTA group states that any modern ultrasound machine with a high-frequency (more than 6 Hz) transvaginal probe can be used. The ADNEX model has not been validated for use in people who are pregnant.

Overa (MIA2G) serum test (Vermillion)

3.3 The Overa (MIA2G) is a CE-marked qualitative serum test that combines the results of 5 immunoassays into a single numeric result (the Overa Risk Score). The 5 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 (CA125). The serum levels of these biomarkers are determined using immunoassays run on the Roche cobas 6000 system. The Overa 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 indicates a low probability of malignancy and a score of 5.0 or more indicates a high probability of malignancy. The assay is for use in people over 18 years with a pelvic mass for whom surgery may be considered. It is intended to be part of preoperative assessment to help decide if a person presenting with a pelvic mass has a high or low risk of ovarian malignancy.

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

(MIA2G) is available to the NHS through a private laboratory which tests samples and provides Overa Risk scores.

Risk of malignancy index 1 (RMI I) with thresholds other than 250

3.5 The RMI I tool combines 3 pre-surgical features (measured serum CA125 levels [CA125], ultrasound imaging [U] and menopausal status [M]) to create an index score: RMI I score = U×M×CA125. Definitions of these terms from the NICE guideline on ovarian cancer are in table 2.

Table 2 Definitions of RMI I terms

Terms in RMI I equation	Description
U	Ultrasound score based on 1 point scored for the presence of each of the following features: multilocular cysts, solid areas, metastases, ascites, bilateral lesions. U=0 (0 points), U=1 (1 point) or U=3 (2 to 5 points).
M	Menopausal status: M=1 (premenopausal) or M=3 (postmenopausal). The classification of 'postmenopausal' is a woman who has had no period for more than 1 year or a woman over 50 who has had a hysterectomy.
CA125	Serum CA125 concentration measured in units per millilitre (U/ml).

3.6 The NICE guideline on ovarian cancer recommends that people with an RMI I score of 250 or more should be referred to a specialist MDT (the RMI I at this threshold is the comparator for this assessment, see section 3.15). However, this guideline also includes a research recommendation stating that further research should be done to determine the optimum RMI I threshold that should be applied in secondary care to guide the management of suspected ovarian cancer. The subsequently published Scottish Intercollegiate Guidelines Network (SIGN) guideline on the management of epithelial ovarian cancer (SIGN 135) recommends referring people with an RMI I score of more than 200 to a gynaecological oncology multidisciplinary team.

Risk of ovarian malignancy algorithm (ROMA)

3.7 The ROMA combines serum CA125 and HE4 levels with a person's menopausal status to estimate the probability that they have epithelial ovarian cancer. Different equations are used depending on whether the person is pre- or postmenopausal (Moore et al. 2009). Cut-off values for the ROMA score stratify individuals as being at a high or low risk of having epithelial ovarian cancer. Cut-off values vary depending on which manufacturers' HE4 and CA125 assays are being used. The ROMA has not been validated in people under 18 years old, people being treated with chemotherapy and people who have previously been treated for a malignancy.

3.8 Three assays that measure HE4 serum levels using automated immunoassay analysers, and that are available to the NHS, are described in the following sections.

ARCHITECT HE4 (Abbott Diagnostics)

3.9 A CE-marked chemiluminescent microparticle immunoassay designed for use on the Abbott ARCHITECT i2000SR or ARCHITECT i1000SR immunoassay analysers. It is intended for use with the ARCHITECT CA125 II assay, with results of both assays used in the ROMA to help estimate the risk that someone presenting with an adnexal mass and who will have surgery has epithelial ovarian cancer. The following cut-off values are suggested for ROMA to determine if there is a high or low risk of epithelial ovarian cancer: 7.4% for people who are premenopausal; 25.3% for people who are postmenopausal.

Lumipulse G HE4 (Fujirebio Diagnostics)

3.10 A CE-marked chemiluminescent enzyme immunoassay designed for use on the LUMIPULSE G System (either the LUMIPULSE G1200 or LUMIPULSE G600 immunoassay analysers). It is intended for use with the Lumipulse G CA125 II assay, with results of both assays used in the ROMA to help estimate the risk that someone presenting with an adnexal mass and who will have surgery has epithelial ovarian cancer. The following cut-off values are suggested for ROMA to determine if there is

a high or low risk of epithelial ovarian cancer: 13.1% for people who are premenopausal; 27.7% for people who are postmenopausal.

Elecsys HE4 immunoassay (Roche Diagnostics)

3.11 A CE-marked immunoassay test that uses Roche's ElectroChemiLuminescence detection technology designed for use on the following immunoassay analysers: Modular analytics E170, cobas e 411, cobas e 601/e 602 and cobas e 801. It is intended for use with the Elecsys CA 125 II assay, with results of both assays used in the ROMA to help estimate the risk that someone presenting with a pelvic mass has epithelial ovarian cancer. The following cut-off values are suggested for ROMA to determine if there is a high or low risk of epithelial ovarian cancer: 11.4% for people who are premenopausal; 29.9% for people who are postmenopausal.

Simple Rules ultrasound classification system

3.12 Simple Rules was developed by the IOTA group to assess people with a pelvic mass who are considered to need surgery. It is a scoring system based on the presence of ultrasound features, to characterise an ovarian tumour before surgery as benign or malignant. No specific make or model of ultrasound device is needed to use the Simples Rules system. A transvaginal probe is needed and image quality must be of sufficient quality to allow the ultrasound features specified by the Simple Rules system to be seen.

3.13 Terms and definitions used in the classification system are as defined by the IOTA group. The group run courses that teach the terms, definitions and measurement techniques needed to assess pelvic masses for the Simple Rules. An online training tool for NHS practitioners is also currently under development. Simple Rules has not been validated for use in people who are pregnant.

3.14 There are 5 rules that predict a malignant tumour (M-rules) and 5 rules that predict a benign tumour (B-rules), as described in table 3. If any M-rules apply (and no B-rules) then the mass is classified as malignant. If any B-rules apply (and no M-rules) then the mass is classified as benign.

However, if both M- and B-rules apply, or neither, then the result is inconclusive, and is either classed as malignant or further criteria are needed to assess whether the mass is likely to be malignant; for example, further expert subjective assessment of the ultrasound images.

Table 3 Simple Rules ultrasound classification system

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 100 mm or more Very strong blood flow (colour score 4)	Unilocular Solid components present, with largest solid component having a largest diameter of less than 7 mm Acoustic shadows present Smooth multilocular tumour with largest diameter less than 100 mm No blood flow (colour score 1)

The comparator

3.15 The comparator for this assessment is the RMI I used at a threshold of 250, as currently recommended in the NICE guideline on ovarian cancer.

4 Evidence

The diagnostics advisory committee (section 8) considered evidence on tests used in secondary care to help identify people at high risk of ovarian cancer from several sources. Full details of all the evidence are in the committee papers.

Clinical effectiveness

4.1 Fifty-one diagnostic cohort studies were identified (in 65 publications) that reported data on 1 or more of the included tests or risk scores. Also, an unpublished interim report of phase 5 of the International Ovarian Tumor Analysis (IOTA) study was available to the external assessment group (EAG) and committee as academic in confidence. No randomised controlled trials or controlled clinical trials were identified; neither were studies that reported how test results affect clinical management decisions. Ten studies had inclusion criteria which allowed people under 18 years to take part; but the number of participants in this age group was not reported.

4.2 All the included studies reported the accuracy of tests and risk scores to assess people with an adnexal or pelvic mass. When summary estimates of sensitivity and specificity from multiple studies were calculated, these were separate pooled estimates produced using random-effects logistic regression. The bivariate/hierarchical summary receiver operating characteristic model was not used because data sets were either too small or too heterogeneous.

4.3 Histopathology was the reference standard used to assess test accuracy in all of the identified studies. The target condition (that is, what was considered a positive reference standard test result) varied between the included studies. Some studies classified borderline ovarian tumours as positive, but others did not (and either classified them as disease negative or excluded them from analyses). Furthermore, studies varied as to whether they included people with metastases to the ovaries and germ cell tumours in analyses.

4.4 The methodological quality of the diagnostic cohort studies was assessed using the QUADAS-2 tool. Fifteen studies had a high risk of bias in the 'flow and timing' domain, most commonly because not all patients were included in the analyses and patients did not all have the same reference standard. Regarding applicability, 26 studies were rated as 'high' concern on at least 1 domain. The EAG commented that areas of concern for applicability included how the index test was applied and whether this could be considered to be representative of routine practice. A further issue for applicability of studies was how the target condition was defined. One study, which reported the development and validation of the ADNEX model (Van Calster et al. 2014), was also assessed using the PROBAST tool; a tool developed to assess the methodological quality of prediction modelling studies.

Assessment of test accuracy

Risk of malignancy index 1 (RMI I) at decision thresholds other than 250

4.5 Ten studies reported diagnostic accuracy of the RMI I using a decision threshold of 250 (the comparator for this assessment) and at least 1 further threshold value. Two studies were done in the UK, 2 elsewhere in Europe and 6 in non-European countries. CA125 assays from various manufacturers were used in the studies.

4.6 In studies that directly compared RMI I at a threshold of 250 and 200, no statistically significant difference between the sensitivity and specificity of RMI I at these thresholds was seen in any of the target condition categories (see table 4).

Table 4 Comparative accuracy of RMI I at thresholds of 200 and 250

Source	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline				

Summary estimates (6 studies; n=1,079)	All	RMI I (200)	70.8 (65.6 to 75.6)	91.2 (88.9 to 93.1)
		RMI I (250)	69.0 (63.7 to 73.9)	91.6 (89.3 to 93.5)
Target condition: Ovarian malignancies including borderline				
Yamamoto et al. 2009 (n=253)	All	RMI I (200)	80.0 (65.2 to 89.5)	86.4 (81.8 to 89.9)
		RMI I (250)	72.5 (57.2 to 83.9)	88.7 (84.4 to 92.0)
Target condition: All malignant tumours excluding borderline				
Summary estimates (2 studies; n=248)	All	RMI I (200)	73.5 (64.3 to 81.3)	89.6 (83.2 to 94.2)
		RMI I (250)	66.4 (56.9 to 75.0)	93.3 (87.7 to 96.9)
Abbreviations: CI, confidence interval; RMI I, risk of malignancy index 1.				

Risk of ovarian malignancy algorithm (ROMA)

4.7 Fourteen studies (in 22 publications) reported diagnostic accuracy data for the ROMA using either Abbott ARCHITECT assays (9 studies) or Roche Elecsys assays (5 studies). No studies were identified that used the Fujirebio Lumipulse G automated CLEIA system.

ARCHITECT HE4 (Abbott Diagnostics)

4.8 All of the 9 ROMA studies which used Abbott ARCHITECT assays were done outside the UK: 3 in European countries, 4 in Asia, 1 in the US and 1 in Oman. No direct comparisons (that is, when both tests were assessed in the same patient cohort) between ROMA and RMI I (threshold of 250) were identified.

4.9 Three studies made a direct comparison between ROMA using Abbott ARCHITECT assays and RMI I (threshold of 200), shown in table 5. One study (Al Musalhi et al. 2016) did not exclude participants from analysis

based on their final histopathological diagnosis; but the other 2 studies did. Sensitivity was highest when people with borderline tumours and non-epithelial ovarian cancers were excluded from analysis, and lowest when all participants (regardless of final histopathological diagnosis) were included. The reverse was true for specificity. When all participants were included in the analysis (Al Musalhi et al. 2016) there was no statistically significant difference between the sensitivity and specificity estimates of ROMA and RMI I (threshold of 200). This was also true for the summary sensitivity estimate when the target condition was 'epithelial ovarian malignancies excluding borderline'; however specificity was statistically significantly lower for ROMA compared with RMI I (threshold of 200).

Table 5 Comparative accuracy of ROMA (using Abbott ARCHITECT assays) and RMI I (threshold of 200)

Source	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline				
Al Musalhi et al. 2016	All (n=213)	ROMA[1]	75.0 (60.4 to 86.4)	87.9 (81.9 to 92.4)
		RMI I (200)	77.1 (62.7 to 88.0)	81.8 (75.1 to 87.4)
	Premenopausal (n=162)	ROMA[1]	52.4 (29.8 to 74.3)	90.1 (83.9 to 94.5)
		RMI I (200)	57.1 (34.0 to 78.2)	85.1 (78.1 to 90.5)
	Postmenopausal (n=51)	ROMA[1]	92.6 (75.7 to 99.1)	79.2 (57.8 to 92.9)
		RMI I (200)	91.7 (73.0 to 99.0)	66.7 (46.0 to 83.5)
Target condition: Epithelial ovarian malignancies including borderline				

Winarto et al. 2014	All (n=128)	ROMA	91.0 (81.5 to 96.6)	42.6 (30.0 to 55.9)
		RMI I (200)	80.6 (69.1 to 89.2)	65.6 (52.3 to 77.3)
Target condition: Epithelial ovarian malignancies excluding borderline				
Summary estimate (2 studies)	All (n=1,172)	ROMA	96.4 (93.6 to 98.2)	53.3 (50.0 to 56.7)
		RMI I (200)	93.4 (90.0 to 95.9)	80.3 (77.5 to 82.9)

[1] Manufacturer's suggested thresholds not used.

Abbreviations: CI, confidence interval; RMI I, risk of malignancy index 1; ROMA, risk of ovarian malignancy algorithm.

4.10 Further identified studies assessed the performance of the ROMA score (using the Abbott ARCHITECT assays and at the company's suggested thresholds) without comparison with RMI I, across a range of target conditions. These included epithelial ovarian malignancies (both including and excluding borderline tumours). One study reported that the sensitivity of the ROMA was higher when the target condition was stage III or IV epithelial ovarian cancer, rather than stage I or II. Also, accuracy data at ROMA thresholds different from those suggested by the manufacturer were identified, but the EAG commented that no alternative threshold offered a clear performance advantage.

Elecsys HE4 immunoassay (Roche Diagnostics)

4.11 All of the 5 ROMA studies that used Roche Elecsys assays were done outside the UK: 1 in a European country, 3 in Asia and 1 in the US. No direct comparisons (that is, when both tests were assessed in the same cohort) between ROMA and RMI I (threshold of 250) were identified. One study (Yanaranop et al. 2016) made a direct comparison between ROMA using Roche Elecsys assays and RMI I (threshold of 200). In this study, people with a final histological diagnosis of borderline ovarian tumour were classified as disease negative. Differences between the ROMA and RMI I (threshold of 200) sensitivity (83.8% compared with 78.4%) and

specificity (68.6% compared with 79.6%) values were not statistically significant. The data were similar when stratified by menopausal status. When people with non-epithelial ovarian cancer were excluded from analysis in this study (target condition epithelial ovarian malignancies), sensitivity for both ROMA and RMI I (threshold of 200) increased, but not statistically significantly. Sensitivity was higher for ROMA when the target condition was stage II to IV epithelial ovarian malignancies (97.2%; 95% confidence interval [CI] 85.5 to 99.9%) when compared with stage I epithelial ovarian malignancies (76.7%; 95% CI 57.7 to 90.1%). This was also the case for RMI I (threshold of 200).

4.12 Four further studies assessed the ROMA score (using Roche Elecsys assays) without comparison with RMI I. Two of these studies included all participants in analyses (Janas et al. 2015; Shulman et al. 2016; target condition all malignant tumours including borderline), shown in table 6.

Table 6 Diagnostic accuracy of ROMA (using Roche Elecsys assays and manufacturer's suggested thresholds)

Source	Subgroup	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline			
Summary estimate (2 studies; n=1,252)	All	79.1 (74.2 to 83.5)	79.1 (76.3 to 81.6)
Janas et al. 2015	Premenopausal (n=132)	90.0 (55.5 to 99.7)	82.0 (74.0 to 88.3)
	Postmenopausal (n=127)	78.6 (65.6 to 88.4)	76.1 (64.5 to 88.4)
Abbreviation: CI, confidence interval.			

4.13 Two studies assessed the performance of the ROMA score (using the Roche Elecsys assays and at the company's suggested thresholds) without comparison with RMI I and with a target condition of ovarian malignancies excluding borderline tumours. The sensitivity estimates from these studies were very different (95.5% and 53.8%) and no

summary estimate was calculated. Also, accuracy data at ROMA thresholds different from those suggested by the manufacturer were identified, but the EAG commented that no alternative threshold offered a clear performance advantage.

Lumipulse G HE4 (Fujirebio Diagnostics)

4.14 None of the included studies assessed the ROMA score and used the Fujirebio Lumipulse G HE4 assay. The EAG identified 2 studies that used a ROMA score calculated using a manual Fujirebio tumour marker enzyme immunometric assay (EIA) assay; however this assay was outside the scope of this assessment.

Simple Rules

4.15 Seventeen published studies had data on the diagnostic accuracy of Simple Rules. Eleven of these studies were done in Europe, including 3 in the UK. Two studies were multinational and included UK participants, 2 studies were done in Thailand, 1 was done in Brazil and 1 study did not provide detail on location. Also, the provided interim report (academic in confidence) had diagnostic accuracy results for Simple Rules. In studies included in summary estimates of sensitivity and specificity, Simple Rules was done by a level 2 or 3 examiner as defined by the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) classification system; 1 study also reported data from level 1 examiners.

4.16 Four published studies and the unpublished interim report provided direct comparison of the accuracy of Simple Rules and RMI I at a threshold of 200. The summary estimate of sensitivity was statistically significantly higher for Simple Rules (93.9%; 95% CI 92.8 to 94.9%) when compared with RMI I (threshold of 200; 66.9%; 95% CI 64.8 to 68.9%); however the summary specificity estimate was statistically significantly lower (74.2% [95% CI 72.6 to 75.8%] compared with 90.1% [95% CI 88.9 to 91.2%]). All these studies included all participants in analysis, regardless of their final histopathological diagnosis (target condition all malignant tumours including borderline). The unpublished interim report also directly compared Simple Rules and RMI I (threshold of 250; academic in confidence).

4.17 A further 4 studies had data on the accuracy of Simple Rules for the same target condition but without a direct comparison with RMI I. There was no statistically significant change in sensitivity (94.2%; 95% CI 93.3 to 95.1%) or specificity (76.1%; 95% CI 74.9 to 77.3%) when data from these studies were included in the summary estimates of Simple Rules accuracy (a total of 8 published studies and the unpublished interim work).

4.18 Three studies directly compared Simple Rules and RMI I (threshold of 200) stratified by menopausal status. There was no statistically significant difference between the sensitivity and specificity estimates for Simple Rules produced for the pre- and postmenopausal subgroups. However if data from a further study (which did not report a direct comparison with RMI I) were added, the summary estimate for specificity was statistically significantly higher for people who are premenopausal (79.3%; 95% CI 77.0 to 81.5%), when compared with people who are postmenopausal (67.3%; 95% CI 63.5 to 70.9%).

4.19 In the above estimates of accuracy for Simple Rules, inconclusive results were treated as malignancy positive. Test accuracy data were also available from some studies in which inconclusive results were instead classified by expert subjective assessment of the ultrasound images. Assessment of inconclusive results from Simple Rules using expert subjective assessment (rather than assuming them to be malignant) statistically significantly increased the specificity of the test, but statistically significantly lowered sensitivity.

The ADNEX model

4.20 Six published studies had data on the diagnostic accuracy of the ADNEX model. One was done entirely in the UK and 2 were multicentre studies that included UK participants. The remaining 3 studies were done elsewhere in Europe. A further unpublished interim report (provided as academic in confidence) also had data on the diagnostic accuracy of the ADNEX model. Four of the studies did not report details about the people doing the ultrasound scans. In 1 study, ultrasound scans were done by EFSUMB level 2 ultrasound examiners (non-consultant gynaecology specialists, gynaecology trainee doctors and gynaecology sonographers)

and in another study they were done by EFSUMB level 2 or 3 practitioners with 8 to 20 years' experience in gynaecological sonography.

4.21 The EAG focused on test accuracy at the 10% threshold. One published study and the unpublished interim report made a direct comparison between the ADNEX model and RMI I (threshold of 200). Sensitivity was statistically significantly higher for ADNEX (96.0%; 95% CI 94.5 to 97.1%) than RMI I (threshold 200; 66.0%; 95% CI 62.9 to 69.0%), but specificity was statistically significantly lower (67.0% [95% CI 64.2 to 69.6%] compared with 89.0% [95% CI 87.0 to 90.7%]). Also, a further 2 studies reported on the accuracy of the ADNEX model in the same target population (all malignant tumours including borderline) but without direct comparison with RMI I. Inclusion of data from these studies in summary estimates did not cause a statistically significant change to sensitivity (96.3%; 95% CI 95.3 to 97.1%) or specificity (69.1%; 95% CI 67.4 to 70.8%) of the ADNEX model. The unpublished interim report also directly compared the ADNEX model and RMI I (threshold of 250; academic in confidence).

4.22 Two further studies had data on the accuracy of the ADNEX model without comparison with RMI I. These studies excluded people with histopathological diagnoses other than primary ovarian cancer from analysis (target condition ovarian malignancies including borderline). The summary estimate of sensitivity from these studies did not differ significantly from that of studies that included all participants in analysis; however the summary estimate of specificity (77.6%; 95% CI 73.6 to 81.2%) was statistically significantly higher.

4.23 Data stratified by menopausal status was available from 1 study. No statistically significant effect on sensitivity was reported, but specificity was statistically significantly higher for people who were premenopausal than for people who were postmenopausal.

4.24 One published study and the unpublished interim analysis directly compared the ADNEX model and Simple Rules (inconclusive results assumed to be malignant). The summary estimate of sensitivity was statistically significantly higher for ADNEX (96.0%; 95% CI 94.5 to 97.1%)

than Simple Rules (92.8%; 95% CI 90.9 to 94.3%). Summary estimates of specificity were similar.

Overa (MIA2G)

4.25 Three studies (in 4 publications) had data on the diagnostic performance of Overa (MIA2G). All the studies were done in the USA and used a score of 5 units as a threshold. No studies were identified that directly compared Overa (MIA2G) with RMI I (at any threshold). However, 1 study assessed the accuracy of the Overa (MIA2G) and ROMA (using Roche Elecsys assays and manufacturer suggested thresholds for ROMA) in the same population with a target condition of all malignancies including borderline. Overa (MIA2G) had a statistically significantly higher sensitivity (91.0% [95% CI 86.8 to 94.0%] compared with 79.2% [73.7 to 83.8%]) and statistically significantly lower specificity (65.5% [95% CI 62.0 to 68.8%] compared with 78.9% [75.8 to 81.7%]) than the ROMA in this study.

4.26 Two further studies reported the diagnostic accuracy of Overa (MIA2G) without comparison with other risk scores. The summary estimate of sensitivity was 90.2% (95% CI 84.6 to 94.3%), and specificity was 65.8% (95% CI 61.9 to 69.5%). One of these studies assessed subgroups of people who were pre- and postmenopausal; there was no statistically significant difference between these groups.

Cost effectiveness

Systematic review of cost-effectiveness evidence

4.27 The EAG did a systematic review to identify existing studies that assessed the cost effectiveness of the included tests and risk scores to help identify people with ovarian cancer. Five studies were identified, however 2 of these related to the use of tests in screening so were not applicable to the scope of this assessment. One of the studies (Havrilesky et al. 2015) included the ROMA and the Multivariate Index Assay algorithm (MIA; from Vermillion who also produce the Overa [MIA2G; multivariate index assay 2nd generation]). Both were dominated

(that is, they cost more and produced less life years) by the use of CA125 alone or by a strategy of referring all people for specialist care (without testing). Conversely, in Forde et al. (2016) MIA dominated the use of CA125 alone (that is, it was cost saving and produced more quality-adjusted life years [QALYs]). No identified studies assessed the cost effectiveness of all the tests and risk scores included in this assessment.

Modelling approach

4.28 The EAG developed a de novo economic model designed to assess the cost effectiveness of the following tests and risk scores when used in secondary care to help decide whether to refer people with suspected ovarian cancer to a specialist multidisciplinary team (MDT):

- RMI I – threshold of 250
- ROMA – using Abbott ARCHITECT assays
- ROMA – using Roche Elecsys assays
- Overa (MIA2G) – threshold of 5 units
- IOTA Simple Rules – inconclusive results assumed to be malignant
- IOTA ADNEX model – threshold of 10%
- RMI I – threshold of 200.

4.29 The model did not include assessment of the ROMA using Fujirebio Diagnostics' Lumipulse G HE4 assay because no studies were identified that provided data on the accuracy of the ROMA using this assay. In the base-case analysis the starting cohort was assumed to be 40 years old, consistent with the modelling produced for the NICE guideline on ovarian cancer. All costs and effects included in the model were discounted by 3.5%.

Model structure

4.30 The EAG developed a decision tree and Markov model for the assessment. The decision tree was used to model short-term outcomes

(up to 30 days after surgery) and the Markov model for longer-term outcomes over a lifetime horizon. In the decision tree, the alternative tests and risk scores were assessed by their ability to help decision-making about referral to a specialist MDT. After the referral, people in the decision tree were classified as being in 1 of the following states: early ovarian cancer, advanced ovarian cancer, benign mass, colorectal cancer or death (to account for mortality 30 days after surgery).

4.31 Longer-term costs and QALYs (over a lifetime horizon) were estimated using a Markov cohort model. This model included separate states for people with ovarian cancer who were treated in a specialist MDT and those who were not, to allow a beneficial effect for treatment in a specialist MDT to be applied. This treatment effect was a hazard ratio of 0.90 (95% CI 0.82 to 0.99) applied to overall survival of people with ovarian cancer (for people with both early and advanced stage ovarian cancer) and to progression-free survival for people with early stage ovarian cancer. This effect size was taken from a Cochrane review (Woo et al. 2012) which reported this hazard ratio for overall survival of people with ovarian cancer who had treatment in institutions with gynaecologic oncologists on site compared with community or general hospitals. The EAG assumed that this hazard ratio would also apply for progression-free survival, based on data in Woo et al. (2012).

Model inputs

4.32 The accuracy of the assessed tests and risk scores used in the model were taken from the clinical-effectiveness review and are shown in table 7. The EAG used diagnostic accuracy estimates derived from studies in which the target condition was 'all malignant tumours including borderline'; that is, studies that did not exclude participants from analysis on the basis of their final histological diagnosis. This was because the EAG considered that this population would produce estimates of test performance most representative of clinical practice. The prevalence of malignancies used in the model (21.3%; comprising ovarian malignancies, including borderline, and non-ovarian malignancies) was a summary estimate calculated from diagnostic cohort studies identified in the clinical-effectiveness review.

Table 7 Diagnostic accuracy estimates used in the model

	Sensitivity (standard error)	Specificity (standard error)	Source
RMI I – threshold of 250	64.4% (1.4%)	91.8% (0.7%)	Summary estimate from 1 unpublished study (IOTA 2017) and 6 studies (Davies et al. 1993; Jacobs et al. 1990; Lou et al. 2010; Morgante et al. 1999; Tingulstad et al. 1996; Ulusoy et al. 2007).
ROMA Abbott ARCHITECT	75.0% (6.6%)	87.9% (2.7%)	Summary estimate from Al Musalhi et al. (2016).
ROMA Roche Elecsys	79.1% (2.4%)	79.1% (1.4%)	Summary estimate from 2 studies (Janas et al. 2015; Shulman et al. 2016).
Overa (MIA2G) – threshold of 5 units	90.2% (2.5%)	65.8% (1.9%)	Summary estimate from 2 studies (Coleman et al. 2016; Zhang et al. 2015).
IOTA Simple Rules – inconclusive assumed to be malignant	94.2% (0.5%)	76.1% (0.6%)	Summary estimate from 1 unpublished study (IOTA 2017) and 8 studies (Adballa et al. 2013; Alcazar et al. 2013; Knafel et al. 2015; Meys et al. 2016; Sayasneh et al. 2013; Silvestre et al. 2015; Testa et al. 2014; Timmerman et al. 2010).
IOTA ADNEX model – threshold of 10%	96.3% (0.5%)	69.1% (0.9%)	Summary estimate from 1 unpublished study (IOTA 2017) and 3 studies (Meys et al. 2016; Sayasneh et al. 2016; Van Calster et al. 2014).

RMI I – threshold of 200	68.1% (0.9%)	90.1% (0.5%)	Summary estimate from 1 unpublished study (IOTA 2017) and 12 studies (Abdalla et al. 2013; Al Musalhi et al. 2016; Davies et al. 1993; Jacobs et al. 1990; Lou et al. 2010; Meys et al. 2016; Morgante et al. 1999; Sayasneh et al. 2013; Testa et al. 2014; Tingulstad et al. 1996; Ulusoy et al. 2007; Van Gorp et al. 2012).
Abbreviations: RMI I, risk of malignancy index 1; ROMA, risk of ovarian malignancy algorithm.			

Costs

4.33 The costs associated with the use of the different risk scores used in the model are shown in table 8. Costs were taken from companies, published literature and routine sources of NHS costs. Further costs used in modelling were taken from modelling done for the NICE guideline on ovarian cancer, relevant NHS reference costs, Personal Social Services Research Unit publications and further identified literature. No costs related to the training needed for the use of Simple Rules and ADNEX model were included in base-case analysis. However, the effect of additional costs (to reflect potential training costs) for these tests was investigated in scenario analysis.

Table 8 Risk score costs used in modelling

Test	Ultrasound cost1 (£)	Test cost per kit (£)	Total HE4 test-related costs2 (£)	CA125 cost3 (£)	Total cost (£)
ADNEX	76.75	–	–	25.58	102.34
Overa (MIA2G)	76.75	99.00	–	–	175.80
RMI I	76.75	–	–	25.58	102.34
ROMA (Abbott ARCHITECT)	76.75	21.33	6.64	25.58	130.31
ROMA (Roche Elecsys)	76.75	15.95	7.81	25.58	126.09

Simple Rules	76.75	–	–	–	76.75

[1] Calculated from the cost of transvaginal ultrasound scans used in economic modelling for the NICE guideline on ovarian cancer and inflated to 2015/16 values.

[2] Includes capital, quality control, maintenance, shipping, calibration and personnel costs, as set out in appendix 6 of the diagnostics assessment report.

[3] Cost of doing a CA125 assay calculated from the NICE guideline on ovarian cancer (adjusted for inflation).

Abbreviations: RMI I, risk of malignancy index 1; ROMA, risk of ovarian malignancy algorithm.

Health-related quality of life and quality-adjusted life year decrements

4.34 Utility estimates used in modelling are shown in table 9.

Table 9 Utility scores used in modelling

		Utility value estimate	Source
Benign mass (assumed equal to general population)		Age dependent	Ara et al. (2010)
Early ovarian cancer	Treated by specialist MDT	0.83	Havrilesky et al. (2009)
	Not treated by specialist MDT treated	Equal to treated by specialist MDT	Assumption
Advanced ovarian cancer	Treated by specialist MDT	0.63	Grann et al. (1998)
	Not treated by specialist MDT treated	Equal to treated by specialist MDT	Assumption
Colorectal cancer	Dukes' A	0.74	Ness et al. (1999)
	Dukes' B	0.67	
	Dukes' C	0.50	
	Dukes' D	0.25	

Abbreviations: MDT, multidisciplinary team.

Base-case results

4.35 The following assumptions were applied in the base-case analysis:

- All non-ovarian malignancies were assumed to be colorectal cancer.
- People with a false negative diagnosis were more likely to have early-, rather than advanced-, stage ovarian cancer.
- Inconclusive results from Simple Rules were assumed to be malignant.
- All people with a false positive and false negative diagnosis were operated on for a benign mass.
- No disutility was applied for people who were incorrectly told that they have ovarian cancer (false positives).

4.36 In the base-case model analysis, the EAG did a pairwise analysis comparing the costs and QALYs resulting from using the included tests and risk scores with RMI I (threshold of 250), and also a fully incremental analysis (table 10). Use of Simple Rules (inconclusive assumed to be malignant) was the cheapest and second most effective, and dominated RMI I (at a threshold of 200 and 250). Use of the ADNEX model was most effective (that is, produced the most QALYs) and when compared with Simple Rules produced an incremental cost-effectiveness ratio (ICER) of £15,304 per QALY gained. Use of the ROMA and Overa (MIA2G) were dominated.

Table 10 Base-case analysis results

	Compared with RMI I (threshold of 250)			Full incremental analysis
	Difference in costs	Difference in QALYs	Difference in costs / difference in QALYS	
Simple Rules – inconclusive assumed to be malignant	−£2	0.021	Dominant	Cheapest

RMI I – threshold of 250	£0	0	N/A	Dominated
RMI I – threshold of 200	£4	0.002	£2,483	Dominated
ADNEX – threshold of 10%	£30	0.023	£1,274	£15,304
ROMA – Abbott ARCHITECT	£38	0.005	£7,506	Dominated
ROMA – Roche Elecsys	£44	0.007	£6,409	Dominated
Overa (MIA2G) – threshold of 5 units	£105	0.017	£6,038	Dominated
Abbreviations: QALY, quality-adjusted life year; RMI I, risk of malignancy index 1; ROMA, risk of ovarian malignancy algorithm.				

4.37 At a maximum acceptable ICER of £20,000 per QALY gained, the ADNEX model and Simple Rules had a probability of being cost effective of 60% and 39% respectively. At a maximum acceptable ICER of £30,000 per QALY gained, these probabilities were 75% (ADNEX) and 23% (Simple Rules). The probability of RMI I (threshold of 250) being cost effective at both thresholds was about 1%, and the probabilities of the other tests and risk scores was less than 1%.

Sensitivity analysis

4.38 Use of the ADNEX model remained cost effective at £20,000 and £30,000 per QALY gained in one-way deterministic sensitivity analysis when most parameters were altered. Simple Rules became cost effective in some analyses, typically when the costs of using the ADNEX model were increased (or Simple Rules costs were decreased) or the diagnostic accuracy of the Simple Rules was improved relative to ADNEX. Also, when the upper bound value for the overall-survival hazard ratio for people with an ovarian malignancy treated in a specialist MDT (rather than secondary care) was used, (that is, the beneficial effect of surgery done by a specialist MDT was at its lowest level in the model), Simple Rules became cost effective at both £20,000 and £30,000 per QALY gained.

Alternative scenario analyses

4.39 The EAG did several scenario analyses to test assumptions made about parameter values used in the base-case model analysis. Use of the ADNEX model remained cost effective in most scenario analysis. However, in some scenarios Simple Rules (inconclusive results assumed to be malignant) was cost effective. These included when a disutility (the value of which was arbitrary) was applied for people with a false positive diagnosis for 1 year and when the benefit of treatment in specialist care was reduced.

4.40 In a scenario analysis in which a higher cost of surgery done by a specialist MDT was used, RMI I (threshold of 250) was cost effective at a maximum acceptable ICER of £20,000 per QALY gained and Simple Rules was cost effective at a maximum acceptable ICER of £30,000 per QALY gained. In this scenario, an additional cost of £2,500 was added to the average cost of surgery done by a specialist MDT, to reflect expert opinion that some patients referred to a specialist MDT will have extensive surgery for ovarian cancer.

Subgroup analyses

4.41 Results from subgroup analyses were similar to the base-case analyses when the starting age of the cohort was 50 years and also when only early stage cancer was considered. However, when the analysis was run for advanced stage cancer, Simple Rules (rather than ADNEX) was cost effective at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained. No changes to sensitivity or specificity values for tests were made in these subgroup analyses (because of a lack of data on test performance in these populations).

4.42 The EAG also did subgroup analyses for populations who were pre- and postmenopausal. Sensitivity and specificity estimates for tests or risk scores in these subgroups were taken from the clinical-effectiveness review; but relatively few studies were available to inform these estimates. A different starting age of the cohort and prevalence of malignancy (compared with the base-case analysis) was also used for these subgroups. The ADNEX model was cost effective at thresholds of

£20,000 and £30,000 per QALY gained for both these subgroup analyses.

5 Committee discussion

5.1 The committee discussed the potential benefits of correctly identifying people referred to secondary care with suspected ovarian cancer who have a benign or malignant mass. It heard from patient and clinical experts that a correct diagnosis of a malignant mass at an early stage will increase the likelihood of survival. Patient experts also suggested that even for people with stage III ovarian cancer, earlier identification of the condition could mean that there is a lower volume of tumour on which to operate if surgery is indicated.

5.2 The committee discussed the potential disadvantages of incorrectly referring people with a benign mass to a specialist multidisciplinary team (MDT). The committee heard that false positive results (that is, people with a benign mass who are incorrectly told that it is likely to be malignant) lead to unnecessary anxiety for patients and their families, and may result in an increased number of people having surgery when in fact no surgery, or less extensive surgery, could be considered. The committee noted that this is particularly an issue for people who are premenopausal and who may wish to consider fertility-conserving surgery, as well as for people who are older or frail and wanting to avoid surgery if possible.

Clinical effectiveness

5.3 The committee discussed the diagnostic accuracy data available for the included tests. It noted that the studies in the clinical-effectiveness review varied in which target condition they used (that is, what was considered a positive reference standard test result). The committee heard from clinical experts that although epithelial cancers are the most common form of ovarian malignancy, people referred to secondary care with suspected ovarian cancer will include those with non-ovarian tumours and non-epithelial ovarian tumours. The committee noted that studies which used a target condition of ovarian cancer or epithelial ovarian cancer retrospectively excluded patients from analysis based on their reference standard diagnosis, and that these studies had been

assessed as having a high risk of bias by the external assessment group (EAG). The committee concluded that studies which used a target condition of all malignant (including borderline) tumours were most representative of clinical practice.

5.4 The committee considered the generalisability of the evidence to clinical practice in the NHS. It heard from clinical experts that the prevalence of malignancy in study populations was considerably higher than would be expected for people referred to secondary care with suspected ovarian cancer in the NHS. Clinical experts commented that a prevalence of less than 10% would be expected, and suggested 5% as a realistic prevalence of malignancy in this population. The committee heard from the EAG that, in addition, most studies did not report the distribution of disease stages among patients with ovarian cancer. Therefore, it noted that the spectrum of disease in the studies may not reflect that seen in secondary care in the NHS. The committee concluded that the study populations, for all tests, may not be representative of the clinical population for this assessment. In particular, the differing levels of disease severity in the study populations and in secondary care in the NHS could mean that the sensitivity and specificity estimates obtained from these studies are not accurate estimates of how the tests would perform in secondary care in the NHS.

5.5 The committee considered the test accuracy of the Simple Rules and ADNEX. It noted that studies showed that they were statistically significantly more sensitive than RMI I. The committee then considered the expertise of practitioners doing and interpreting ultrasound scans in studies assessing the Simple Rules system and the ADNEX model. The committee heard from clinical experts that practitioners in these studies had a higher level of skill and experience than is generally available in secondary care in the NHS, and noted that there were limited data available on the accuracy of tests done by less experienced operators. The committee heard that although image acquisition would generally be straightforward for NHS practitioners, additional training would be needed in interpreting the images to use the Simple Rules or ADNEX model. The committee heard from clinical experts that the local organisation of care would be likely to affect test performance in practice; and that there is uncertainty about the effect of NHS service

models on how well the Simple Rules and ADNEX model would perform in secondary care. For example, if a model of delivery in which patients are seen, scanned and their condition managed in 1 setting was used, this may lead to improved performance of the tests through a concentration of services and skills. It also heard from clinical experts that no data have been published on inter-observer variation using the IOTA tests in the NHS. The committee therefore concluded that there is considerable uncertainty about the likely performance of the Simple Rules and ADNEX tests in secondary care in the NHS, and that the accuracy reported in the studies may not be achieved in clinical practice in the NHS.

5.6 The committee considered the test accuracy data for the ROMA and Overa (MIA2G). It noted that relatively few studies were identified for these tests, particularly studies with a direct comparison with RMI I. The committee concluded that there is considerable uncertainty about the diagnostic accuracy of these tests; however it is possible that the tests may offer improved accuracy relative to RMI I.

5.7 The committee considered data on the accuracy of RMI I at thresholds other than 250. It noted that most studies with a direct comparison of RMI I at 250 and another threshold used a threshold of 200, and that there were relatively few studies with other alternative thresholds. It also noted that there was very little difference in the summary estimates of sensitivity and specificity for RMI I at thresholds of 200 and 250 obtained from studies with a direct comparison. The committee therefore concluded that RMI I used with a threshold of 200 was unlikely to offer accuracy benefits over using this test with a threshold of 250, and noted that the use of RMI I at a threshold of 250 is recommended in the NICE guideline on ovarian cancer.

5.8 The committee discussed how the stage of ovarian cancer (early or advanced) could affect test accuracy. It heard from clinical experts that about 70% of people identified with ovarian cancer have advanced stage cancer. However, the main benefit of tests such as RMI I in secondary care is in identifying early stage ovarian cancer, with advanced stage ovarian cancer being more apparent from imaging. The committee heard from the EAG that very few data were available to inform estimates of test accuracy by stage of ovarian cancer. Two studies assessing the

ROMA reported that sensitivity was lower for detecting early stage ovarian cancers. However most studies did not provide details on the stage of cancer of study participants included in analysis. The committee heard from clinical experts that populations in studies were likely to include more cases of advanced than early stage ovarian cancers, and that the performance of tests to detect early and advanced stage ovarian cancer could differ substantially. The committee concluded that there is uncertainty about how accurate the tests are at correctly detecting early stage ovarian cancer, potentially the most relevant group for this assessment. The committee also concluded that data on the accuracy of tests to detect early stage ovarian cancer would be important for any future assessment (see section 6.1).

Cost effectiveness

5.9 The committee discussed the sensitivity and specificity estimates used in the cost-effectiveness modelling. It noted that these estimates were taken from the clinical-effectiveness review and that the concerns about the applicability of data from these studies to NHS secondary care (see sections 5.4 and 5.5) also apply to the model. The committee also noted that relatively few studies were available to inform test accuracy estimates for the ROMA and Overa (MIA2G) tests (see section 5.6).

5.10 The committee noted that tests with the highest sensitivity (ADNEX and Simple Rules) that resulted in more people with ovarian cancer being referred to a specialist MDT tended to be cost effective. It considered the parameter used in the model for the beneficial effect of a referral to a specialist MDT for people with ovarian cancer; a hazard ratio for overall and progression-free survival obtained from a Cochrane review (see section 4.31). The committee heard that estimates from this review were based on studies with a mixed cohort of early and advanced stage ovarian cancer, and that because advanced stage was likely to be predominant in this cohort, the summary estimate may not be an accurate reflection of the beneficial effect of specialist MDT treatment on people with early stage ovarian cancer. It also heard that the benefits for people with early stage ovarian cancer of having their surgery done by a gynaecological oncology specialist are potentially more difficult to assess than those for people with advanced stage cancer. An improved

quality of surgery is likely to lead to more accurate staging of the cancer, which will help with subsequent treatment decisions. For example, accurate staging may show that chemotherapy is not needed (for low-risk stage I disease); but inadequate staging in a non-specialist centre could lead to inappropriate use of chemotherapy or the need for further surgery to accurately stage the cancer. The committee also heard from clinical experts that people with a false negative test result will have their condition managed in secondary care, and if their ovarian malignancy is recognised at a later date they will then be referred to a specialist MDT. The effect of this delayed referral, rather than lack of referral, to a specialist MDT on patient outcomes is unclear. The committee concluded that, because of a lack of data, there is considerable uncertainty about the effect of false negative test results on people with ovarian cancer, particularly if they have early stage ovarian cancer.

5.11 The committee discussed the costs included in the model for people referred to a specialist MDT. It noted that the cost of an MDT meeting had been included, but heard from clinical experts that additional costs may be incurred when a patient is referred to a specialist MDT for discussion; such as costs for the time taken by radiologists to review images in advance of the meeting. The committee heard from the EAG that these additional costs were not captured in the model. The committee also noted that in the base case, the model used NHS reference costs for surgery, and that this may not adequately capture the cost of extensive surgery potentially needed for people with advanced stage cancer, who make up 75% of the population with an ovarian malignancy in the model. It heard from clinical experts who suggested that the costs associated with more extensive surgery should be included in the model. The committee noted that in a scenario analysis in which additional surgery costs were assumed, RMI I (threshold 250) was cost effective at a maximum acceptable incremental cost-effectiveness ratio of £20,000 per quality-adjusted life year gained. The committee concluded that the costs of a referral to, and treatment by, a specialist MDT may have been underestimated in the model, and that this could affect the model results, such as which tests seemed to be cost effective.

5.12 The committee discussed the costs of CA125 testing included in the

model. It heard from the EAG that the economic model assumed that all patients have a CA125 test in secondary care, even if they previously had one in primary care. The committee heard from clinical experts that the reasons for patients having another CA125 test in secondary care are: CA125 levels may have changed since the first test was carried out; some risk scores are only compatible with a specific brand of CA125 test; and some tests include CA125 as part of an array. The committee concluded that the assumption in the economic model in relation to CA125 testing costs was valid.

5.13 The committee considered its discussions on the clinical- and cost-effectiveness evidence. It concluded that:

- There is considerable uncertainty about the estimates of test accuracy used in modelling because: relatively few studies were found to inform estimates (for the ROMA and Overa [MIA2G]; see section 5.6); the high prevalence of malignancy in studies suggested that they were not representative of clinical populations in secondary care in the NHS (see section 5.4); and that the level of expertise of people interpreting scans for the Simple Rules and the ADNEX model in studies was higher than would be routinely available in the NHS (see section 5.5).
- There is uncertainty about the accuracy of tests to detect early stage ovarian cancer (see section 5.8), and about the likely effect on outcomes for people with early stage ovarian cancer who have a delayed referral to a specialist MDT, as a result of an initial false negative test result (see section 5.10).
- There is uncertainty about the costs of assessment and treatment at a specialist MDT, and that higher costs would impact model results (see section 5.11).

Other considerations

5.14 The committee discussed the accuracy of tests, noting that tests with higher sensitivity had lower specificity. The committee heard from clinical experts that tests with high sensitivity reduce the number of missed cases of ovarian cancer, but that lower test specificity will result in more false positive referrals to specialist MDTs. The committee heard from

clinical experts that there is very limited specialist MDT capacity for personnel in relation to the current case demand, and that increasing the number of false positive referrals to specialist MDTs would reduce the quality of assessment by limiting time available for discussion for each patient. Clinical experts commented that this could adversely affect, or delay, decisions made in specialist MDT meetings about patient treatment. The committee concluded that using tests with lower specificity would have a large impact on specialist MDT services, but the model has not captured the impact because of a lack of data on the effect this would have on patient care and clinical outcomes and because of its structure.

5.15 The committee considered the accuracy of the tests for people who are pre- and postmenopausal. It noted that relatively few studies provided accuracy estimates stratified by menopausal status. The committee further noted that the EAG did cost-effectiveness analyses for pre- and postmenopausal subgroups; however there were relatively few data to inform the tests' performance in pre- and postmenopausal populations in this analysis. The committee heard from clinical experts that menopausal status could considerably affect the performance of the tests, and that further data are needed to assess the performance of the tests in these subgroups (see section 6.1).

5.16 The committee discussed the likely effect of test results on decisions made about patient care. It noted that no data were available on the effect of test results on decisions about patient care or referral. The committee heard from clinical experts that in practice results from tests such as the RMI I are used alongside further information, such as imaging, when making decisions about patient care and referral. It noted therefore that increased accuracy of testing may not correspond to changes in decision-making. The committee concluded that there is uncertainty about how the results of the tests included in the assessment would be used in clinical practice in the NHS, and that further research on this would be useful (see section 6.2).

5.17 The committee heard from patient experts that access to tests in primary care for people with ovarian cancer symptoms varied, and that getting a referral to secondary care could take a long time. It heard further from

clinical experts that about 60% of people referred to secondary care have a late stage of cancer, and they are often referred from colorectal and urological cancer services and emergency care. The committee agreed that differences in initial assessment may lead to variation in patient outcomes. The committee noted that tests for suspected ovarian cancer in primary care were outside of the scope of this assessment and that evidence from this setting had not been reviewed.

5.18 The committee discussed the use of tests in sequence. It heard from clinical experts that the use of a highly sensitive test followed by a highly specific test may improve accuracy of referral. The committee noted that the EAG had looked for studies that assessed the use of included tests in combination or sequence in the clinical-effectiveness review; however very few data were found (1 study with a lack of detail about how test results were combined to produce a positive result). The committee considered that the use of tests in sequence could be cost effective and further research is needed to inform accuracy estimates (see section 6.3).

Research considerations

5.19 The committee noted that there is an ongoing National Institute for Health Research funded diagnostic test accuracy study; the ROCkeTS (Refining Ovarian Cancer Test Accuracy Scores) study, which will report in 2019 or 2020. This study will evaluate existing and new risk prediction models for people with symptoms of suspected ovarian cancer against a comparator of RMI I (threshold of 250). It will assess patients using the IOTA Simple Rules and ADNEX tests, as well as biomarker assays. Practitioners doing ultrasound scans as part of the study will have an IOTA training course; therefore the study will provide test accuracy data for NHS practitioners with a defined amount of training. The study will also report costs and resource use associated with the diagnostic tests. The recruited study population will be people referred to secondary care in the NHS; the committee noted that the results are likely to be very relevant to future updates of this guidance.

6 Recommendations for further research

6.1 Further diagnostic accuracy studies, or analyses of existing data sets, are recommended to assess the accuracy of the tests included in this assessment in the following subgroups:

- people who are premenopausal
- people who are postmenopausal
- people with suspected early stage ovarian cancer, that is, disease apparently confined to the pelvis.

  Future studies should be done in populations that are representative of people with suspected ovarian cancer who are assessed in NHS secondary care.

6.2 Further research is recommended to assess:

- inter-observer reproducibility of tests involving ultrasound scans (the ADNEX model and Simple Rules)
- changes in clinical management based on test results from the ADNEX model, Overa (MIA2G), ROMA and Simple Rules.

6.3 Further research is recommended to assess the diagnostic accuracy of the tests included in this assessment when used in combination; for example sequentially.

7 Implementation

NICE will support this guidance through a range of activities to promote the recommendations for further research. The research proposed will be considered by the NICE Medical Technologies Evaluation Programme research facilitation team for the development of specific research study protocols as appropriate. NICE will also incorporate the research recommendations in section 6 into its guidance research recommendations database (available on the NICE website) and highlight these recommendations to public research bodies.

8 Diagnostics advisory committee members and NICE project team

Diagnostics advisory committee

The diagnostics advisory committee is an independent committee consisting of 22 standing members and additional specialist members. A list of the committee members who participated in this assessment appears below.

Standing committee members

Professor Adrian Newland
Chair, Diagnostics Advisory Committee and Professor of Haematology, Barts Health NHS Trust

Dr Mark Kroese
Vice Chair, Diagnostics Advisory Committee and Consultant in Public Health Medicine, PHG Foundation, Cambridge and UK Genetic Testing Network

Mr John Bagshaw
In-vitro Diagnostics Consultant

Professor Enitan Carrol
Chair in Paediatric Infection, University of Liverpool

Dr Sue Crawford
GP Principal, Chillington Health Centre

Dr Owen Driskell
Lead for Laboratory Medicine, National Institute for Health Research Clinical Research Network West Midlands

Dr Steve Edwards
Head of Health Technology Assessment, BMJ Evidence Centre

Dr Simon Fleming
Consultant in Clinical Biochemistry and Metabolic Medicine, Royal Cornwall Hospital

Dr James Gray
Consultant Microbiologist, Birmingham Children's Hospital

Professor Steve Halligan
Professor of Radiology, University College London

Mr John Hitchman
Lay Member

Professor Chris Hyde
Professor of Public Health and Clinical Epidemiology, Peninsula Technology Assessment Group (PenTAG)

Mr Patrick McGinley
Head of Costing and Service Line Reporting, Maidstone and Tunbridge Wells NHS Trust

Dr Michael Messenger
Deputy Director and Scientific Manager, National Institute for Health Research Diagnostic Evidence Co-operative, Leeds

Mrs Alexandria Moseley
Lay Member

Dr Peter Naylor
GP, Wirral

Dr Dermot Neely
Consultant in Clinical Biochemistry and Metabolic Medicine, Newcastle upon Tyne NHS Trust

Professor Matt Stevenson
Professor of Health Technology Assessment, School of Health and Related Research, University of Sheffield

Professor Anthony Wierzbicki

Consultant in Metabolic Medicine/Chemical Pathology, St Thomas' Hospital

Specialist committee members

Professor Richard Edmondson
Professor of Gynaecological Oncology, The University of Manchester

Dr Jurjees Hasan
Consultant Medical Oncologist, The Christie NHS Foundation Trust, Manchester

Mr Jeremy Hawe
Consultant Gynaecologist, The Countess of Chester NHS Foundation Trust

Dr Tracie Miles
Clinical Nurse Specialist, Royal United Hospital, Bath. Information Nurse Specialist to The Eve Appeal

Mr Stuart Morgan
Lay Specialist Committee Member

Dr Hilary Morrison
Lay Specialist Committee Member

Dr Cathie Sturgeon
Consultant Clinical Scientist, Royal Infirmary of Edinburgh

Dr Sudha Sundar
Senior Lecturer/Consultant in Gynaecological Oncology, City Hospital, West Midlands

Dr Michael Weston
Consultant Radiologist, St James's University Hospital

NICE project team

Each diagnostics assessment is assigned to a team consisting of a technical analyst (who acts as the topic lead), a technical adviser and a project manager.

Thomas Walker

Topic Lead (until July 2017)

Jessica Maloney
Topic Lead (from August 2017)

Frances Nixon
Technical Adviser

Robert Fernley
Project Manager (until July 2017)

Donna Barnes
Project Manager (from August 2017)

ISBN: 978-1-4731-2726-5