National Institute for Health and Care Excellence

DIAGNOSTICS ASSESSMENT PROGRAMME

Evidence overview

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

This overview summarises the key issues for the diagnostics advisory committee's consideration. This document is intended to be read in conjunction with the final scope issued by NICE for the assessment and the diagnostics assessment report. A glossary of terms can be found in Appendix B.

1 Background

1.1 Introduction

The purpose of this assessment is to evaluate the clinical and cost effectiveness of tests used in secondary care to help determine if a person referred with suspected ovarian cancer is likely to have a malignancy. The tests inform decisions about whether someone should be referred to a specialist multidisciplinary team (MDT) for further assessment and treatment.

Ovarian cancer starts in cells in, or near, the ovaries. There were about 7,200 new cases of ovarian cancer in the UK in 2013, accounting for 2% of all new cancer cases. Primary ovarian tumours are classified based on the tissue that they develop from. There are 3 main types:

- epithelial ovarian tumours
- sex cord-stromal tumours of the ovary
- germ cell tumours of the ovary.

Tumours can be benign, malignant or intermediate (borderline malignant); about 90% of primary ovarian cancers are malignant epithelial tumours.

This assessment will potentially update part of the NICE guideline on ovarian cancer, which gives recommendations on establishing a diagnosis of ovarian cancer in secondary care. This guideline focuses on epithelial ovarian cancer, and does not cover germ cell tumours or sex cord-stromal tumours of the ovary. It recommends measuring serum CA125 in people in secondary care with suspected ovarian cancer and then calculating a risk of malignancy index 1 (RMI 1) score. This score is based on ultrasound characteristics seen, CA125 serum levels and menopausal status. People with an RMI 1 score of 250 or more should be referred to a specialist MDT for further assessment and treatment. The Scottish Intercollegiate Guidelines Network (SIGN) guideline on management of epithelial ovarian cancer also recommends using the RMI 1 score to predict the likelihood of ovarian cancer. However, referral to a gynaecology-oncology MDT is recommended if the score is more than 200. Recommendations from the NICE guideline on ovarian cancer related to tumour markers, imaging and malignancy indices in secondary care can be found in appendix C. A flow chart showing these recommendations (from the full NICE guideline on ovarian cancer) can be found in appendix D.

Serum biomarker CA125 is widely used in secondary care, as part of the RMI 1 score, to decide if a referral to a specialist MDT is needed. However, patients with early stage epithelial ovarian cancer often do not have elevated CA125 levels. Also, elevated levels of CA125 are not always indicative of ovarian cancer - they may be raised because of other causes, such as endometriosis, fibroids, pregnancy or pelvic inflammatory disease. Tests and risk scores included in this assessment (ADNEX, Overa [MIA2G], RMI 1 at thresholds other than 250, ROMA and Simple Rules) may be better able to distinguish between benign and malignant ovarian tumours, and improve the accuracy of referral from secondary care to a specialist MDT.

Increasing the proportion of people with ovarian cancer who have their initial treatment determined by a specialist MDT is likely to improve patient outcomes. Conversely, 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 has the potential to reduce inappropriate referrals to specialist care for further assessment and treatment, and the costs and anxiety that this can cause.

Provisional recommendations on the use of these technologies will be formulated by the Diagnostics Advisory Committee at the Committee meeting on 20 June 2017.

1.2 Scope of the evaluation

Table 1 Scope of the evaluation

Decision question	What is the clinical and cost effectiveness of tests in secondary care (ADNEX, Overa [MIA2G], RMI 1, ROMA and Simple Rules) to identify people who are at high risk of ovarian cancer and who should be referred to a specialist multidisciplinary team?
Populations	People who are referred to secondary care with suspected ovarian cancer. Potential subgroups include: • people who are pre-menopausal • people who are post-menopausal.
Interventions	• ADNEX • Overa (MIA2G) • RMI 1 testing (with a value other than 250 as a cut-off for referral or incorporating HE4 serum levels) • ROMA • Simple Rules ultrasound-based testing.
Comparator	RMI 1 testing (with a score of 250 as a cut-off for referral).
Healthcare setting	Secondary care.
Outcomes	Intermediate measures for consideration may include: • diagnostic accuracy of testing • time to test result • number of inconclusive results

	• stage of ovarian cancer at diagnosis • number of people referred to gynaecological oncology multidisciplinary teams • number of cross sectional imaging scans for people with suspected ovarian cancer • number of people who have ovarian cancer whose initial surgery to remove a pelvic mass is not done by a gynaecological oncologist • adverse events from biopsy or surgery.
	Clinical outcomes for consideration may include: • morbidity associated with ovarian cancer (or surgery for suspected ovarian cancer) • mortality associated with ovarian cancer (or surgery for suspected ovarian cancer) • survival.
	Patient-reported outcomes for consideration may include: • health-related quality of life.
	Costs will be considered from an NHS and Personal Social Services perspective. Costs for consideration may include: • cost of equipment, reagents and consumables • cost of staff and associated training • costs of procedures, including biopsy, histological examination and surgery (including secondary surgery) and including any time related costs associated with these procedures • costs associated with treatment and subsequent testing to confirm diagnosis • costs arising from adverse events.
	The cost effectiveness of interventions should be expressed in terms of incremental cost per quality-adjusted life year.
Time horizon	The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared.

Further details including descriptions of the interventions, comparator, care pathway and outcomes can be found in the final scope. Table 2 provides an overview of the components of the included tests and risk scores.

Table 2 Summary of the components of included tests and risk scores

	Components of the tests/risk scores				
	CA125	**HE4**	**Other serum markers**	**Ultrasound scan features**	**Menopausal status**
ADNEX	X			X	
Overa (MIA2G)	X	X	X		
RMI 1	X			X	X
ROMA	X	X			X
Simple Rules				X	

2 The evidence

This section summarises data from the diagnostics assessment report compiled by the external assessment group (EAG).

2.1 Clinical effectiveness

The EAG did a systematic review to identify evidence on the clinical effectiveness of using tests and risk scores (ADNEX, Overa [MIA2G], RMI 1, ROMA and Simple Rules) in secondary care to guide referral decisions for people with suspected ovarian cancer (who had not previously been treated for ovarian cancer and who were not having chemotherapy). This included identifying studies that reported the accuracy of the tests and risk scores at different thresholds and also their use in combination, or in sequence with 1 or more additional tests. Details of the systematic review can be found starting on page 43 of the diagnostics assessment report.

Overview of the included studies

Fifty one studies were identified (in 65 publications). Also, an unpublished interim report of phase 5 of the IOTA study was provided as academic in confidence. All the included studies were diagnostic cohort studies that reported data on 1 or more of the included tests or risk scores. An overview of the included studies is provided in table 5 of the diagnostics assessment report, starting on page 52. No randomised controlled trials or controlled

clinical trials were identified. Ten studies had inclusion criteria which included people under 18 years; but the number of participants in this age group was not reported.

The methodological quality of the included diagnostic cohort studies was assessed using the QUADAS-2 tool. One study which reported the development and validation of the ADNEX model (Van Calster et al. 2014) was also assessed using the PROBAST tool. Full detail on the critical appraisal of the studies can be found in the diagnostics assessment report starting on page 55.

Thirty seven of the studies (73%) were rated as having an 'unclear' risk of bias and 19 studies (37%) were rated as 'high' risk of bias on at least 1 QUADAS-2 domain. The EAG commented that the flow and timing domain was the main source of concern, with 15 studies rated as having 'high' risk of bias. This was because not all patients were included in the analyses (13 out of the15 studies) and that patients did not all have the same reference standard. [redacted]

Regarding applicability, 24 of the studies (47%) were rated as 'unclear' on at least 1 domain and 26 studies (51%) were rated as 'high' 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. For example, in several studies the index test was done by people whose level of experience was likely to be higher than would be routinely available in secondary care. A further issue for applicability of studies was how the target condition was defined. Eighteen studies (35%) had 'high' risk for applicability because their target condition was defined as 'any malignant tumour', which could include non-ovarian primary cancers and also metastases to the ovaries. However, studies that limited their target condition to ovarian cancer excluded participants from the analysis; generally in practice by retrospectively removing cases (with non-

ovarian primary cancer) from the analysis on the basis of histopathology results. This may lead to overestimation of test performance and therefore any such studies were rated as having a ‘high’ risk of bias in the timing and flow domain (discussed above).

Diagnostic accuracy

All the included studies measured the accuracy of tests and risk scores to assess people with an adnexal/pelvic mass. Where 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.

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 other did not (and either classified them as disease negative or excluded them from analyses). Furthermore, studies varied in whether they included people with metastases to the ovaries and germ cell tumours in analyses. A description of how the EAG defined the target condition of studies in the diagnostics assessment report is shown in table 3. Study level detail of the histopathological diagnosis of participants can be found in table 36 in appendix 2 of the diagnostics assessment report.

Table 3 Definitions of 'target condition' used in the diagnostics assessment report

Target condition	Description
All ovarian malignancies	Participants with a non-ovarian primary cancer were excluded from estimates of test performance, even if it had metastasised to the ovaries (that is, the primary tumour must be ovarian).
All malignant tumours	Participants with a non-ovarian primary cancer were not excluded from estimates of test performance, and were considered as disease positive. This could include people with tumours on the ovaries that had metastasised there from another (non-ovarian) primary cancer and people with an adnexal/pelvic mass that was caused by a non-ovarian cancer (which had not metastasised to the ovaries).
Epithelial ovarian cancer	Participants with malignancies other than epithelial ovarian cancer (identified by post-operative histological diagnosis) were excluded from estimates of test performance.

Full detail on the diagnostic performance of tests included in this assessment can be found in the diagnostics assessment report from page 61.

Risk of Malignancy Index 1 (RMI 1) at decision thresholds other than 250

Ten identified studies reported diagnostic accuracy of the RMI 1 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 (Davies et al. 1993; Jacobs et al. 1990), 2 elsewhere in Europe (Italy and Norway; Morgante et al. 1999; Tingulstad et al. 1996) and 6 in non-European countries (Akturk et al. 2011; Asif et al. 2004; Lou et al. 2010; Manjunath et al. 2001; Ulusoy et al. 2007; Yamamoto et al. 2009). CA125 assays from various manufacturers were used in the studies. Full details can be found starting on page 92 of the diagnostics assessment report.

The EAG focussed on comparative accuracy of RMI 1 at decision thresholds of 200 and 250, shown in table 4. No statistically significant difference between the sensitivity and specificity of RMI 1 at these thresholds was seen in any of the target condition categories. Detail on the individual studies

included in analysis can be found in table 19 on page 95 of the diagnostics assessment report.

Table 4 Comparative accuracy of RMI 1 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=1079)	All	RMI (200)	70.8 (65.6 to 75.6)	91.2 (88.9 to 93.1)
		RMI (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 (200)	80.0 (65.2 to 89.5)	86.4 (81.8 to 89.9)
		RMI (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 (200)	73.5 (64.3 to 81.3)	89.6 (83.2 to 94.2)
		RMI (250)	66.4 (56.9 to 75.0)	93.3 (87.7 to 96.9)
Abbreviations: CI, confidence interval. Asif et al. (2004) had a target condition of all malignant tumours but it was unclear if borderline tumours were included (see table 46 in appendix 4 of the diagnostics assessment report for study results).				

Data on the accuracy of RMI 1 at additional decision thresholds in identified studies can be found in table 40 in appendix 4 of the diagnostics assessment report. As the threshold of RMI 1 used decreased, the sensitivity estimate increased and specificity estimate decreased.

Risk of Ovarian Malignancy Algorithm (ROMA)

Fourteen identified 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)

All of the 9 ROMA studies which used Abbott ARCHITECT assays were done outside the UK: 3 in European countries (Karlsen et al. 2012; Novotny et al.

2012; Presl et al. 2012), 4 in Asia (Chan et la. 2013; Clemente et al. 2015; Li et al. 2016; Winarto et al. 2014), 1 in the USA (Moore et al. 2011) and 1 in Oman (Al Musalhi et al. 2016). No direct comparisons (that is, where both tests were assessed in the same patient cohort) between ROMA and RMI 1 (threshold of 250) were identified. Three studies made a direct comparison between ROMA using Abbott ARCHITECT assays and RMI 1 (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 analysis (Al Musalhi et al. 2016) there was no significant difference between the sensitivity and specificity estimates of ROMA and RMI 1 (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 significantly lower for ROMA compared to RMI 1 (threshold of 200). Full results can be found starting on page 62 of the diagnostics assessment report, with results summarised in table 8 on page 65.

Table 5 Comparative accuracy of ROMA (using Abbott ARCHITECT assays) and RMI 1 (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*	75.0 (60.4 to 86.4)	87.9 (81.9 to 92.4)
		RMI (200)	77.1 (62.7 to 88.0)	81.8 (75.1 to 87.4)
	Pre-menopausal (n=162)	ROMA*	52.4 (29.8 to 74.3)	90.1 (83.9 to 94.5)
		RMI (200)	57.1 (34.0 to 78.2)	85.1 (78.1 to 90.5)
	Post-menopausal (n=51)	ROMA*	92.6 (75.7 to 99.1)	79.2 (57.8 to 92.9)
		RMI (200)	91.7 (73.0 to 99.0)	66.7 (46.0 to 83.5)
Target condition: Epithelial ovarian malignancies including borderline				
Winarto et	All	ROMA	91.0 (81.5 to 96.6)	42.6 (30.0 to 55.9)

al. 2014	(n=128)	RMI (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=1172)	ROMA	96.4 (93.6 to 98.2)	53.3 (50.0 to 56.7)
		RMI (200)	93.4 (90.0 to 95.9)	80.3 (77.5 to 82.9)
* Not using the manufacturer's suggested thresholds. Abbreviations: CI, confidence interval.				

Further studies were identified that assessed performance of the ROMA score (using the Abbott ARCHITECT assays and at the company's suggested thresholds) without comparison with RMI 1 and are reported in table 9 on page 66 of the diagnostics assessment report. None of these studies included all participants, regardless of final histopathological diagnosis, in the analysis. One study (Chan et al. 2013) reported that the sensitivity of the ROMA score was highest when the target condition was stage III/IV epithelial ovarian cancers (stage I/II and borderline tumours excluded from analysis) and that a small, non-significant decrease in sensitivity happened when stage I/II epithelial ovarian cancer was the target condition (people with borderline and higher stage ovarian cancer excluded from analysis). Accuracy data at ROMA thresholds different from those suggested by the manufacturer (for Abbott ARCHITECT assays) can be found in table 37 in appendix 4 of the diagnostics assessment report. The EAG commented that no alternative threshold offered a clear performance advantage.

Elecsys HE4 immunoassay (Roche Diagnostics)

All of the 5 ROMA studies that used Roche Elecsys assays were done outside the UK: 1 in a European country (Janas et al. 2015), 3 in Asia (Chen et al. 2015; Xu et al. 2016; Yanaranop et al. 2016) and 1 in the USA (Shulman et al. 2016). No direct comparisons (that is, where both tests were assessed in the same cohort) between ROMA and RMI 1 (threshold of 250) were identified. One study (Yanaranop et al. 2016) made a direct comparison between ROMA using Roche Elecsys assays and RMI 1 (threshold of 200), shown in table 6. In this study, people with a final histological diagnosis of borderline ovarian tumour were classified as disease negative. Differences between sensitivity

and specificity values for ROMA and RMI (threshold of 200) 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 1 (threshold of 200) increased, although not significantly. This study also presented results stratified by stage of ovarian cancer, which can be found in table 10 of the diagnostics assessment report, on page 71. Full discussion of the ROMA using Roche assays can be found in the diagnostics assessment report from page 68.

Table 6 Comparative accuracy of ROMA (using Roche Elecsys assays) and RMI (threshold of 200)

Study	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours – borderline tumours classified as disease negative				
Yanaranop et al. 2016	All (n=260)	ROMA	83.8 (73.4 to 91.3)	68.8 (61.6 to 75.4)
		RMI (200)	78.4 (67.3 to 87.1)	79.6 (73.1 to 85.1)
	Pre-menopausal (n=148)	ROMA	85.7 (67.3 to 96.0)	70.8 (61.8 to 78.8)
		RMI (200)	75.0 (55.1 to 89.3)	80.8 (72.6 to 87.4)
	Post-menopausal (n=112)	ROMA	82.6 (68.6 to 92.2)	65.2 (52.4 to 76.5)
		RMI (200)	80.4 (66.1 to 90.6)	77.3 (65.3 to 86.7)
Target condition: Epithelial ovarian malignancies – borderline tumours classified as disease negative				
Yanaranop et al. 2016	All (n=252)	ROMA	87.9 (77.5 to 94.6)	68.8 (61.6 to 75.4)
		RMI (200)	80.3 (68.7 to 89.1)	79.6 (73.1 to 85.1)
Abbreviations: CI, confidence interval.				

Four further studies assessed the ROMA score (using Roche Elecsys assays) without comparison with RMI 1. 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 7. Summary estimates differed from those of the comparative accuracy study (Yanaranop et al. 2016, where borderline tumours were considered disease negative; table 6, above), although not significantly.

Table 7 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= 1252)	All	79.1 (74.2, 83.5)	79.1 (76.3, 81.6)
Janas et al. 2015	Pre-menopausal (n=132)	90 (55.5, 99.7)	82.0 (74.0, 88.3)
	Post-menopausal (n=127)	78.6 (65.6, 88.4)	76.1 (64.5, 88.4)

Further studies were identified that assessed performance of the ROMA score (using Roche Elecsys assays at the company's suggested thresholds) without comparison with RMI 1 and are reported in table 11 page 72 of the diagnostics assessment report. Also, accuracy data at ROMA thresholds different from those suggested by the manufacturer can be found in table 37 in appendix 4 of the diagnostics assessment report. The EAG commented that no alternative threshold offered a clear performance advantage.

Lumipulse G HE4 (Fujirebio Diagnostics)

None of the included studies assessed the ROMA score and used the Fujirebio Lumipulse G HE4 assay. The EAG identified 2 studies (Langhe et al. 2013; Van Gorp et al. 2012) that used a ROMA score calculated using a manual Fujirebio tumour marker EIA assay (results can be found in tables 41 and 42 in appendix 4 of the diagnostics assessment report); however this assay was outside the scope of this assessment.

Between assay comparisons of the ROMA score

No identified study directly compared the performance of the ROMA score using different manufacturer's assays. Between study comparisons (when all study participants were included in the analyses regardless of final histopathological diagnosis) showed no significant difference in the estimates of sensitivity for ROMA scores using the Abbott ARCHITECT assays (from 1

study) and Roche Elecsys assays (from 2 studies). However the specificity estimate for ROMA using Abbott assays (87.9%; 95% CI: 81.9 to 92.4%) was higher than for ROMA using Roche assays (79.1%; 95% CI: 76.3 to 81.6%).

Simple Rules (IOTA group)

Seventeen published studies had data on the diagnostic accuracy of Simple Rules. Eleven of these studies (Abdalla et al. 2013; Alcazar et al. 2013; Baker et al. 2013; Fathallah et al. 2011; Knafel et al. 2016; Meys et al. 2016; Murala et al. 2014; Piovano et al. 2016; Ruiz de Gauna et al. 2015; Sayasneh et al. 2013; Testa et al. 2014) were done in Europe; 3 in the UK (Baker et al. 2013; Murala et al. 2014; Sayasneh et al. 2013). Two studies were multinational and included UK participants (Di Legge et al. 2012; Timmerman et al. 2010), 2 studies were done in Thailand (Tantipalakorn et al. 2014; Tinnangwattana et al. 2015), 1 was done in Brazil (Silvestre et al. 2015) and 1 study did not provide detail on location (Weinberger and Minar, 2013). [redacted]

In studies included in the analysis, Simple Rules was done by a level 2 or 3 examiner as per the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) classification system; 1 study (Knafel et al. 2015) also reported data from level 1 examiners. Studies in which participants with inconclusive results from Simple Rules were excluded from analysis, or which did not included sufficient detail about how these participants were considered, were not included in analysis by the EAG. Results of these studies can be found in table 39 of appendix 4 in the diagnostics assessment report.

The EAG commented that 3 studies were done by the IOTA study core group (Di Legge et al. 2012; Timmerman et al. 2010; Testa et al. 2014) and used data from various stages of the IOTA study. Data from phase 5 of the IOTA study was provided as unpublished interim report (as academic in confidence; IOTA 2017). The EAG commented that the largest data sets for Simple Rules

(and also ADNEX) came from the various phases of the IOTA study, and that these tended to dominate analyses.

The unpublished interim study (IOTA 2017) provided a direct comparison of Simple Rules and RMI 1 at a threshold of 250, shown in table 8. Four published studies (Adballa et al. 2013; Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014) and the unpublished interim report (IOTA 2017) provided direct comparison of the accuracy of Simple Rules and RMI 1 at a threshold of 200; summary estimates are shown in table 8. The summary estimate of sensitivity was significantly higher for Simple Rules when compared with RMI1 (threshold of 200); however the summary specificity estimate was significantly lower. All these studies included all participants in analysis, regardless of their final histopathological diagnosis (target condition all malignant tumours including borderline).

Table 8 Comparative accuracy of Simple Rules and RMI 1 (at thresholds of 200 and 250)

Study	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline				
IOTA 2017 (interim unpublished study; [redacted])	All	Simple Rules (inconclusive = malignancy)	[redacted]	[redacted]
		RMI (250)	[redacted]	[redacted]
Summary estimate (4 published studies and interim unpublished study; [redacted])	All	Simple Rules (inconclusive = malignancy)	93.9 (92.8 to 94.9)	74.2 (72.6 to 75.8)
		RMI (200)	66.9 (64.8 to 68.9)	90.1 (88.9 to 91.2)
Abbreviations: CI, confidence interval.				

Also, a further 4 studies (Alcazar et al. 2013; Knafel et al. 2015; Silvestre et al. 2015; Timmerman et al. 2010) had data on the accuracy of Simple Rules for the same target condition but without a direct comparison with RMI 1. Including data from these studies in summary estimates of Simple Rules accuracy (a total of 8 published studies and the unpublished interim work) did

not significantly alter sensitivity (94.2%; 95% CI: 93.3 to 95.1%) or specificity (76.1% (95% CI: 74.9 to 77.3%).

Three studies (Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014) directly compared Simple Rules and RMI 1 (threshold of 200) stratified by menopausal status (see table 9; full detail in table 16 on page 87 of the diagnostics assessment report). There was no significant difference between the sensitivity and specificity estimates produced for the pre- and post-menopausal subgroups. However if data from a further study (Knafel et al. 2015) which did not report a direct comparison with RMI 1 was added, the summary estimate for specificity was significantly higher for people who are pre-menopausal, when compared with people who are post-menopausal. Full data can be found in table 13 on page 83 of the diagnostics assessment report

Table 9 Comparative accuracy of Simple Rules and RMI 1 (threshold of 200) stratified by menopausal status

Study	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline				
Summary estimate (3 studies; n=1,647)	Pre-menopausal	Simple Rules (inconclusive = malignancy)	94.3 (91.7 to 96.3)	78.2 (75.7 to 80.5)
		RMI (200)	52.2 (47.4 to 56.9)	94.2 (92.7 to 95.5)
Summary estimate (3 studies; n=1,337)	Post-menopausal	Simple Rules (inconclusive = malignancy)	95.5 (93.7 to 96.9)	72.3 (68.9 to 75.5)
		RMI (200)	78.8 (75.7 to 81.7)	78.7 (75.2 to 81.9)
Abbreviations: CI, confidence interval.				

Differential assessment of inconclusive Simple Rules results

In the above estimates of accuracy for Simple Rules (tables 8 and 9), inconclusive results were treated as malignancy positive (inconclusive = malignancy). Test accuracy data was also available when these inconclusive results were instead classified by expert subjective assessment of the

ultrasound images (inconclusive = subjective assessment; Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014), shown in table 10. Full data can be found in table 16 on page 87 of the diagnostics assessment report. Only studies in which the subjective assessment was done by experts (this term was used by studies without further details) or by level 2 or 3 examiners as per the EFSUMB classification system were included.

Table 10 Comparative accuracy of Simple Rules and RMI 1 (at threshold of 200) when inconclusive results were subjectively assessed

Study	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline				
Summary estimate (3 studies; n= 2984)	All	Simple Rules (inconclusive = subjective assessment)	91.2 (89.4 to 92.8)	89.6 (88.1 to 91.0)
		RMI (200)	67.8 (65.0 to 70.4)	98.5 (98.3 to 98.7)
Summary estimate (3 studies; n=1647)	Pre-menopausal	Simple Rules (inconclusive = subjective assessment)	92.3 (89.4 to 94.7)	92.0 (90.3 to 93.5)
		RMI (200)	52.2 (47.4 to 56.9)	94.2 (92.7 to 95.5)
Summary estimate (3 studies; n=1337)	Post-menopausal	Simple Rules (inconclusive = subjective assessment)	92.3 (90.2 to 94.2)	80.3 (76.9 to 83.4)
		RMI (200)	78.8 (75.7 to 81.7)	78.7 (75.2 to 81.9)
Abbreviations: CI, confidence interval.				

Also, 4 studies of Simple Rules without comparison with RMI 1, and in the same target population, were identified in which participants with inconclusive assessments were classified by expert subjective assessment (Alcazar et al. 2013; Knafel et al. 2015; Piovanono et al. 2016; Timmerman et al. 2010). Summary estimates (of the 3 studies with comparative accuracy in table 10 and these 4 additional studies) were sensitivity of 88.4% (95% CI: 86.9 to 89.8%) and specificity of 92.5% (95% CI: 91.6 to 93.4%). The EAG commented that assessment of inconclusive results from Simple Rules using

expert subjective assessment (rather than assuming them to be malignant) significantly increased the specificity of the test, but significantly lowered sensitivity.

One study (Knafel et al. 2015) assessed the impact of training on the performance of Simple Rules. The study reported no statistically significant difference in test performance when EFSUMB level 2 examiners or EFSUMB level 1 examiners did the Simple Rules assessment (compare results in tables 13 and 14 in the diagnostics assessment report); however all examiners in this study had half-day practical training in the use of the Simple Rules.

Full accuracy data for the Simple Rules can be found in tables 13, 14, 15 and 16 (pages 83, 85, 86 and 87, respectively) of the diagnostics assessment report.

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

Six published studies had data on the diagnostic accuracy of the ADNEX model. One was done entirely in the UK (Moffatt et al. 2016) and 2 were multi-centre studies that included UK participants (Sayasneh et al. 2016; Van Calster et al. 2014). The remaining 3 studies were done elsewhere in Europe (Joyeux et al. 2016; Meys et al. 2016; Szubert et al. 2016). A further unpublished interim report (provided as academic in confidence) also had data on the diagnostic accuracy of the ADNEX model (IOTA 2017). The EAG focussed on test accuracy at the 10% threshold (reported by all studies). Data on the accuracy of the ADNEX model at other thresholds can be found in table 38 in appendix 4 of the diagnostics assessment report.

Four of the studies did not report details about the people doing the ultrasound scans (Joyeux et al. 2016; Meys et al. 2016; Moffatt et al. 2016; Szubert et al. 2016). In 1 study (Sayasneh et al. 2016) ultrasound scans were done by EFSUMB level 2 ultrasound examiners (non-consultant gynaecology specialist, gynaecology trainees doctors and gynaecology sonographers) and in another study (Van Calster et al. 2014) they were done by EFSUMB level 2

or 3 practitioners with 8 to 20 years' experience in gynaecological sonography. [redacted]

One published study (Meys et al. 2016) and the unpublished interim report (IOTA 2017) made a direct comparison between the ADNEX model and RMI 1 (threshold of 200), shown in table 11. Also, a further 2 studies reported on the accuracy of the ADNEX test in the same target population (all malignant tumours including borderline; Sayasneh et al. 2016; Van Calster et al. 2014) but without direct comparison with RMI 1. Inclusion of data from these studies in summary estimates did not cause a 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 (IOTA 2017) also directly compared the ADNEX model and RMI 1 (threshold of 250), shown in table 11. Meys et al. (2016) also reported accuracy data for ADNEX by menopausal status, also shown in table 11.

Table 11 Comparative accuracy of the ADNEX model and RMI 1 (at thresholds of 200 and 250)

Study	Subgroup	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline				
Summary estimate (1 study and 1 AIC submission; n=[redacted])	All	ADNEX	96.0 (94.5 to 97.1)	67.0 (64.2 to 69.6)
		RMI (200)	66.0 (62.9 to 69.0)	89.0 (87.0 to 90.7)
IOTA 2017 (n=[redacted])	All	ADNEX	[redacted]	[redacted]
		RMI (250)	[redacted]	[redacted]
Meys et al. 2016	Pre-menopausal (n=128)	ADNEX	100 (86.0 to 100)	71.0 (61.0 to 80.0)
		RMI (200)	42.0 (25.0 to 61.0)	94.0 (86.0 to 97.0)
	Post-menopausal (n=198)	ADNEX	98.0 (91.0 to 100)	54.0 (44.0 to 63.0)
		RMI (200)	82.0 (72.0 to 89.0)	66.0 (56.0 to 74.0)

Abbreviations: CI, confidence interval.

Two further studies (reporting 3 data sets; Joyeux et al. 2016; Szubert et al. 2016) had data on the accuracy of the ADNEX model without comparison with RMI 1. 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 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 significantly higher. Data stratified by menopausal status was available from 1 study (Szubert et al. 2016). No significant effect on sensitivity was reported, but specificity was significantly higher for people who were pre-menopausal than for people who were post-menopausal. Full data can be found in table 12 page 82 of the diagnostics assessment report.

Direct comparison of ADNEX and Simple Rules

One published study (Meys et al. 2016) and the unpublished interim analysis (IOTA 2017) directly compared the ADNEX model and Simple Rules (inconclusive results assumed to be malignant). The summary estimate of sensitivity was 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. Full data are presented in table 15 page 86 of the diagnostics assessment report.

Overa (MIA2G)

Three studies (in 4 publications: Coleman et al. 2016; Wolf et al. 2015; Shulman et al. 2016; Zhang et al. 2015) had data on the diagnostic performance of Overa (MIA2G). Only one study was available as a full paper (Coleman et al. 2016), reports of the other studies were available as a conference abstract or meeting slides. All the studies were done in the USA and used a score of 5 units as the threshold for the Overa (MIA2G). Full details are given in the diagnostics assessment report from page 89.

No studies were identified that directly compared Overa (MIA2G) with RMI 1 (at any threshold). However, Shulman et al. (2016) 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 (table 12). Overa (MIA2G) had a significantly higher sensitivity and significantly lower specificity than the ROMA in this study.

Table 12 Comparative accuracy of Overa (MIA2G) and ROMA

Study	Index test	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline			
Shulman et al. 2016 (n=993)	Overa (MIA2G)	91.0 (86.8 to 94.0)	65.5 (62.0 to 68.8)
	ROMA (Roche)	79.2 (73.7 to 83.8)	78.9 (75.8 to 81.7)
Abbreviations: CI, confidence interval.			

Two further studies reported the diagnostic accuracy of Overa (MIA2G) without comparison with other risk scores (Coleman et al. 2016; Zhang et al. 2015; see table 13). One study (Coleman et al. 2016) assessed subgroups of pre- and post-menopausal people; there was no statistically significant difference between these groups.

Table 13 Diagnostic accuracy of Overa (MIA2G)

Study	Subgroup	Sensitivity % (95% CI)	Specificity % (95% CI)
Target condition: All malignant tumours including borderline			
Summary estimate (2 studies, n=798)	All	90.2 (84.6 to 94.3)	65.8 (61.9 to 69.5)
Coleman et al. 2016	Pre-menopausal (n=276)	90.3 (75.1, 96.7)	71.4 (65.5, 76.7)
	Post-menopausal (n=217)	91.8 (82.2, 96.4)	65.4 (57.6, 72.4)
Abbreviations: CI, confidence interval.			

2.2 Costs and cost effectiveness

Systematic review of cost effectiveness evidence

The EAG did a systematic review to identify existing studies that assessed the cost effectiveness of the included tests to help identify people with ovarian cancer. Details of the review are reported in the diagnostics assessment report from page 97 onwards. 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 and a strategy of referring all people for specialist care (without testing). Conversely, in Forde et al. (2016) the multivariate index assay (MIA) dominated the use of CA125 alone (that is, it was cost saving and produced more QALYs). No identified studies assessed the cost effectiveness of all the tests and risk scores included in this assessment.

Economic analysis

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 inform decisions about the referral of people with suspected ovarian cancer to a specialist MDT:

- RMI 1 (threshold of 250)
- ROMA (using Abbott ARCHITECT assays)
- ROMA (using Roche Elecsys assays)
- Overa (MIA2G) (threshold of five units)
- IOTA Simple Rules (inconclusive results assumed to be malignant)
- IOTA ADNEX model (threshold of 10%)
- RMI 1 (threshold of 200)

The model does 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.

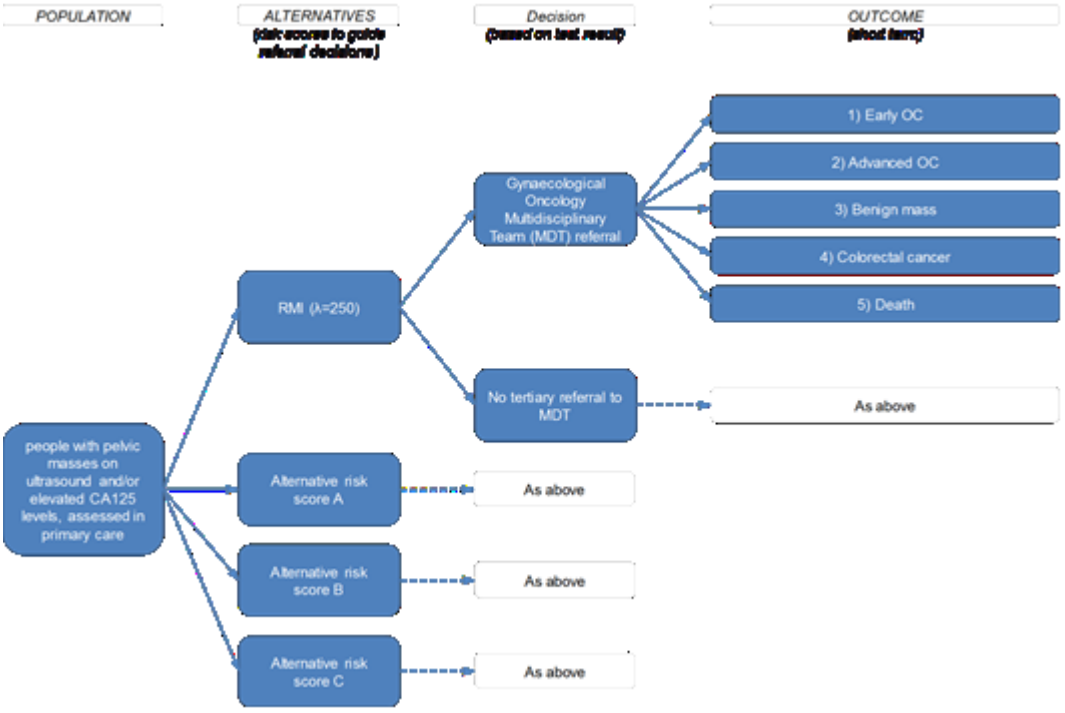
Model structure

Using modelling done for the NICE guideline on ovarian cancer as a starting point, the EAG developed a decision tree and Markov model for the current 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 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%. Full details on model structure can be found in the diagnostics assessment report from page 110.

In the decision tree (figure 1), the alternative tests and risk scores were assessed by their ability to inform a decision about referral to a specialist MDT. People who had a high risk score (true or false) were assumed to be referred to a specialist MDT for treatment (surgery done by a gynaecological oncology specialist), and those without a high risk score (true or false) were assumed to have their treatment in secondary care (surgery done by a secondary care gynaecologist). After the referral, people in the decision tree were allocated to 1 of the following states: early ovarian cancer (OC), advanced OC, benign mass, colorectal cancer or death (to account for 30 days post-surgery mortality).

A small proportion of people with a pelvic mass that test positive for ovarian cancer in secondary care will ultimately be diagnosed with a non-ovarian malignancy. A simplifying assumption made in the model (consistent with modelling for the NICE guideline on ovarian cancer) was that all such non-ovarian malignancies were colorectal cancer, with 2.9% of people with a positive test result (and who are referred to a specialist MDT) being assumed to have colorectal cancer and being allocated to this state in the decision tree.

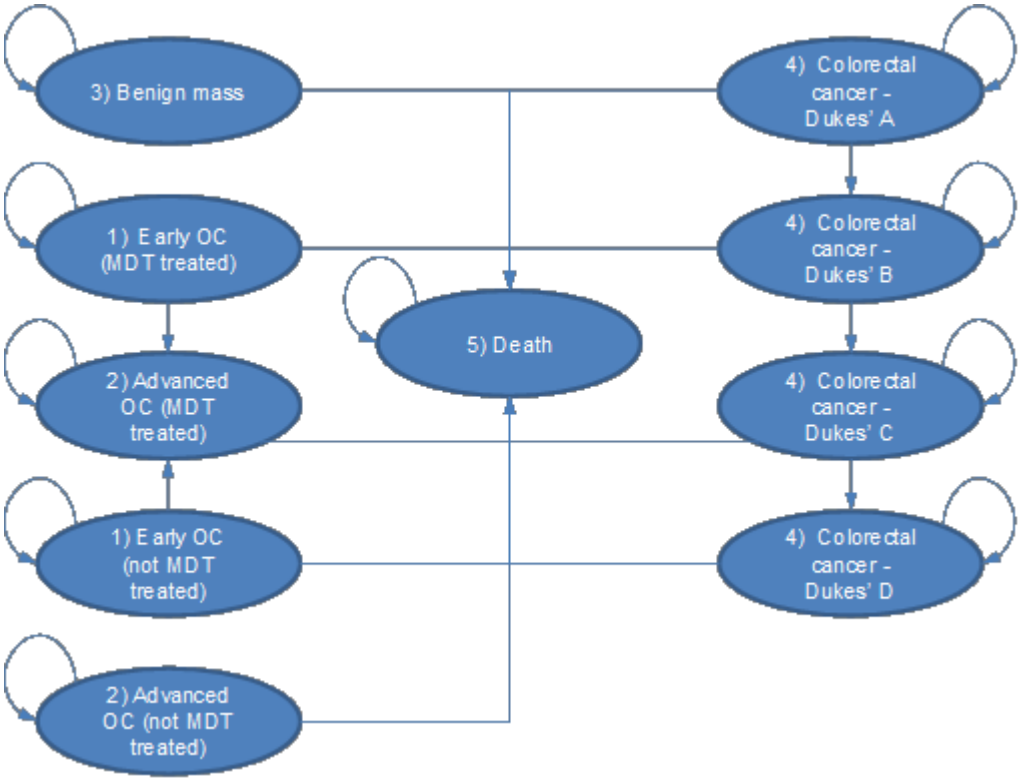
Figure 1 Decision tree structure



For people with a false negative diagnosis (that is, they have a malignancy but are not referred to a specialist MDT), there is an increased risk of progression to advanced ovarian cancer and death. People with a false positive diagnosis are referred to a specialist MDT and incur additional costs related to this referral. No disutility is incurred for false positive patients because it was assumed that they are recognised as having a benign mass by the specialist MDT.

Longer-term costs and QALYs (over a lifetime horizon) were estimated using a Markov cohort model, shown in figure 2. Separate states dependent on whether referral to a specialist MDT happened only exist for ‘early OC’ and ‘advanced OC’ because this referral was only assumed to have an impact on long term outcomes for people with ovarian cancer.

Figure 2 Markov cohort model structure



Model inputs

Parameter values used in the model were taken from several sources. These included the clinical-effectiveness review, focussed searches of literature and by consulting experts for unpublished data. The same sources used in economic modelling for the NICE guideline on ovarian cancer (with costs updated) were used when possible for consistency.

The effect of treatment for people with ovarian cancer in a specialist MDT (rather than secondary care) was estimated from a Cochrane review (Woo et al. 2012). This study reported a hazard ratio of 0.90 (95% CI 0.82 to 0.99) for overall survival of people with ovarian cancer treated in a teaching compared with a general hospital. The EAG assumed that this hazard ratio would also apply for progression-free survival, because Woo et al. (2012) commented that hazard ratios for overall and progression-free survival are very similar.

Full details of the parameter values and sources can be found in the diagnostics assessment report from page 113.

Diagnostic accuracy estimates used in cost effectiveness modelling

The accuracy of the assessed tests and risk scores used in the model were obtained from the clinical-effectiveness review and are shown in table 14. 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 risk score performance most representative of clinical practice. Further detail can be found in section 3.2.7 of the diagnostics assessment report, on page 96.

The EAG used all available studies with accuracy data for this target condition to calculate summary diagnostic accuracy estimates for modelling. That is, not just studies that reported a direct comparison with RMI 1. Summary estimates of test and risk score accuracy calculated from all studies and from only studies with a direct comparison with RMI 1 did not differ significantly, the EAG therefore used the larger data set to make maximum use of the available data.

Table 14 Diagnostic accuracy estimates used in the model

	Sensitivity (standard error)	**Specificity (standard error)**	**Source**
RMI 1 (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%)	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	94.2%	76.1%	Summary estimate from 1

Rules (inconclusive assumed to be malignant)	(0.5%)	(0.6%)	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 1 (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)

The prevalence of malignancies used in the model (21.3%; including ovarian malignancies, including borderline, and non-ovarian malignancies) was a summary estimate obtained from diagnostic cohort studies identified in the clinical-effectiveness review.

Costs

The costs associated with the use of the different risk scores used in the model are shown in table 15. Manufactures indicated that lower costs may be available for higher volume orders of components. A full description of costs used can be found in the diagnostics assessment report from page 118. Costs were obtained from companies, published literature and routine sources of NHS costs.

Table 15 Risk score costs used in modelling

Test	Ultrasound cost [a] (£)	Test cost per kit (£)	Total HE4 test related costs [b]	CA125 cost [c] (£)	Total cost (£)
ADNEX	76.75	-	-	25.58	102.34
Overa (MIA2G)	76.75	99.00	-	-	175.80
RMI 1	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

[a] Calculated from the cost of transvaginal ultrasound scans used in economic modelling for NICE guideline CG122 and inflated to 2015/16 values.

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

[c] Cost of carrying out a CA125 assay calculated from NICE guideline CG122.

In the model, all people with a high risk score were referred to a specialist MDT. The cost of this MDT meeting to discuss a case was assumed to be £116, based on the cost of a specialist MDT meeting from NHS reference costs (2015 to 2016). Further costs related to treatment and follow-up for ovarian cancer were obtained from modelling done for the NICE guideline on ovarian cancer, relevant NHS reference costs, Personal Social Services Research Unit (PSSRU) publications and further identified literature. More detail on the costs used in modelling can be found in table 27 on page 122 of the diagnostics assessment report.

Health-related quality of life and QALY decrements

Utility estimates for people with early ovarian cancer were taken from Havrilesky et al. (2009) and for advanced ovarian cancer from Grann et al. (1998), as used in economic modelling for the NICE guideline on ovarian cancer. The utility scores were not adjusted depending on whether treatment occurred in a specialist MDT. For people with a benign mass, age-dependent general population utility estimates were used, and utilities for people with colorectal cancer were derived from Ness et al. (1999). See table 16 for

values used. Full details can be found in the diagnostics assessment report on page 118.

Table 16 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	SMDT treated	0.83	Havrilesky et al. (2009)
	Not SMDT treated	Equal to SMDT treated	Assumption
Advanced ovarian cancer	SMDT treated	0.63	Grann et al. (1998)
	Not SMDT treated	Equal to SMDT treated	Assumption
Colorectal cancer	Dukes' A	0.74	Ness et al. (1999)
	Dukes' B	0.67	
	Dukes' C	0.50	
	Dukes' D	0.25	
Abbreviations: SMDT, specialist multidisciplinary team			

Base-case results

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

- All non-ovarian malignancies were assumed to be colorectal cancer.
- People classified as false negative were more likely to be early, rather than advanced, stage ovarian cancer.
- Inconclusive results from the 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).

In the base-case model analysis, the EAG did a pairwise analysis comparing the costs and QALYs resulting from use of the included tests and risk scores with RMI 1 (threshold of 250), and also a fully incremental analysis (table 17). Use of Simple Rules (inconclusive assumed to be malignant) was the cheapest and second most effective, and dominated RMI 1 (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 ICER of £15,304 per QALY gained. Use of the ROMA and Overa (MIA2G) were dominated.

Table 17 Base-case analysis results

	Compared to RMI 1 (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 1 (threshold of 250)	£0	0	N/A	Dominated
RMI 1 (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

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 1 (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%.

Full analysis, including cost effectiveness acceptability curves, can be found in the diagnostics assessment report starting on page 126.

Sensitivity analyses

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.

Full details of sensitivity analysis can be found in the diagnostics assessment report starting at page 129 and in appendix 8.

Analysis of alternative scenarios

The EAG did several scenario analyses to test assumptions made about parameter values used in the base-case model analysis. A full list of the scenario analyses done can be found in the diagnostics assessment report starting on page 124. Use of the ADNEX model remained cost effective in most scenario analysis. However, in the following scenarios Simple Rules (inconclusive results assumed to be malignant) was cost effective at a maximum acceptable ICER of £20,000 per QALY:

- For false negative results, an equal proportion of early and advanced stage ovarian cancer was assumed, rather than predominantly early stage as in the base-case analysis.
- Only 90% of true negatives were operated on, rather than 100% as in the base-case analysis (with an associated decrease in costs of non-malignancy surgery and complications).
- A disutility of 0.01 was assumed for false positive cases in the first year.

Also, in the following scenarios Simple Rules was cost effective at maximum acceptable ICERs of both £20,000 and £30,000 per QALY gained:

- A decrease in the benefit of surgery for ovarian malignancies done by a specialist MDT, compared with secondary care; with the hazard ratio for progression-free and overall survival at the upper bound of the confidence interval used in base-case analysis (0.99).
- A disutility of 0.1 was assumed for false positive cases in the first year.

In a scenario analysis in which a higher cost of surgery done by a specialist MDT was used, RMI 1 (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 (the exact cost of this was unknown).

In a scenario analysis using sensitivities and specificities of different RMI 1 thresholds, RMI 1 (threshold of 25) was cost effective at all maximum acceptable ICERs above £2,890 per QALY gained. However, RMI 1 at this threshold was still dominated when included in base-case analysis.

Full scenario analysis results can be found in appendix 9 of the diagnostics assessment report.

Subgroup analysis

The EAG also did several analyses based on subgroup populations. Full analysis can be found in the diagnostics assessment report from page 131. Results were similar to the base-case analysis when the starting age of the cohort was 50 years and also when only early stage cancer was considered. However, when 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. Full results can be found in appendix 10 of the diagnostics assessment report.

People who are pre-menopausal

The starting age for the cohort in this analysis was 38 years; based on age-dependent prevalence data from Cancer Research UK. Accuracy data for the tests and risk scores used in this subgroup were obtained from the clinical-effectiveness review and are shown in table 18. Prevalence of malignancy in this subgroup was also adjusted to 16.2%; a pooled estimate of the prevalence of malignancy in pre-menopausal study populations identified in the clinical-effectiveness review.

Table 18 Diagnostic accuracy values used in subgroup analysis for people who are pre-menopausal

	Sensitivity	Specificity	Source [a]
RMI 1 (threshold of 250)	64.4%	91.8%	No data available specifically for people who are pre-menopausal. Data for all population used instead.
ROMA (Abbott ARCHITECT)	52.4%	90.1%	Al Musalhi et al. (2016)
ROMA (Roche Elecsys)	90.0%	82.0%	Janas et al. (2015)
Overa (MIA2G)	90.3%	71.4%	Coleman et al. 2016
Simple Rules (inconclusive results treated as malignant)	94.5%	79.3%	Summary estimate from 4 studies (Knafel et al. 2016; Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014)
ADNEX	97.0%	71.0%	Meys et al. 2016
RMI 1 (threshold of 200)	53.3%	93.5%	Summary estimate from 5 studies (Al Musalhi et al. 2016; Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014; Van Gorp et al. 2012)
[a] Estimates taken from studies that reported subgroup data for the target condition 'all malignant tumours'.			

Outputs from the model in this subgroup are shown in table 19. At a maximum acceptable ICER of £20,000 per QALY gained, the probability of the ADNEX model being cost effective was 46%, for Simple Rules this was 37% and for

the ROMA (Roche Elecsys) this was 16%. At a maximum acceptable ICER of £30,000 per QALY gained, probabilities were 52% for ADNEX, 27% for Simple Rules and 19% for ROMA (Roche Elecsys). The probability of RMI 1 (threshold of 250) being cost effective at these ICERs was less than 1%, and for all other risk scores was less than 2%. Full analysis can be found in the diagnostics assessment report from page 131.

Table 19 Subgroup analysis for people who are pre-menopausal

	Compared to RMI 1 (threshold of 250)			Full incremental analysis
	Difference in costs	Difference in QALYs	Difference in costs / difference in QALYS	
RMI 1 (threshold of 200)	-£7	-0.003	£1,954	Cheapest
Simple Rules (inconclusive assumed to be malignant)	-£6	0.016	Dominant	£15
RMI 1 (threshold of 250)	£0	0.000	N/A	Dominated
ROMA (Abbott ARCHITECT)	£24	-0.004	Dominated	Dominated
ADENX (10% of threshold)	£28	0.018	£1,564	£18,466
ROMA (Roche Elecsys)	£40	0.013	£2,993	Dominated
Overa (MIA2G)	£100	0.013	£7,748	Dominated

People who are post-menopausal

The starting age for the cohort in this analysis was 68 years; based on age-dependent prevalence data from Cancer Research UK. Accuracy data for the tests and risk scores used in this subgroup were obtained from the clinical-effectiveness review and are shown in table 20. Prevalence of malignancy in this group was adjusted to 45.9%; a pooled estimate of the prevalence of malignancy in post-menopausal study populations identified in the clinical-effectiveness review.

Table 20 Diagnostic accuracy values used in subgroup analysis for people who are post-menopausal

	Sensitivity	Specificity	Source [a]
RMI 1 (threshold of 250)	64.4%	91.8%	No data available specifically for people who are post-menopausal. Data for all population used instead.
ROMA (Abbott ARCHITECT)	92.6%	79.2%	Al Musalhi et al. (2016)[a]
ROMA (Roche Elecsys)	78.6%	76.1%	Janas et al. (2015)
Overa (MIA2G)	91.8%	65.4%	Coleman et al. 2016
Simple Rules (inconclusive results treated as malignant)	95.4%	67.3%	Summary estimate from Knafel et al. 2016; Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014
ADNEX	98.0%	54.0%	Meys et al. 2016
RMI 1 (threshold of 200)	79.4%	79.2%	Summary estimate from Al Musalhi et al. 2016; Meys et al. 2016; Sayasneh et al. 2013; Testa et al. 2014; Van Gorp et al. 2012

[a] Estimates taken from studies that reported subgroup data for the target condition 'all malignant tumours'.

Outputs from the model in the post-menopausal subgroup are shown in table 21. At a maximum acceptable ICER of £20,000 per QALY gained the probabilities of ADNEX and Simple Rules being cost effective were 59% and 40%, respectively. At £30,000 per QALY gained, this was 74% (ADNEX) and 24% (Simple Rules). The probability of RMI 1 (threshold of 250) being cost effective was less than 2%, and less than 1% for all other tests and risk scores. Full analysis can be found in the diagnostics assessment report from page 134.

Table 21 Subgroup analysis for people who are post-menopausal

	Compared to RMI 1 (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)	-£1	0.028	Dominance	Cheapest
RMI 1 (threshold of 250)	£0	0.000	N/A	Dominated
RMI 1 (threshold of 200)	£22	0.013	£1,746	Dominated
ADNEX (threshold of 10%)	£31	0.031	£1,013	£12,876
ROMA (Abbott ARCHITECT)	£45	0.026	£1,759	Dominated
ROMA (Roche Elecsys)	£46	0.012	£3,738	Dominated
Overa (MIA2G)	£99	0.025	£3,992	Dominated

3 Summary

Clinical effectiveness

The comparator for this assessment was RMI 1 at a threshold of 250; however, few studies directly compared the included tests and risk scores with RMI 1 at this threshold. Instead, studies generally used RMI 1 at a threshold of 200. Summary estimates produced from studies that directly compared RMI 1 at thresholds of 200 and 250 showed higher sensitivity (and lower specificity) for RMI 1 at threshold of 200; but differences did not differ significantly between the thresholds.

Identified studies differed in their target condition (that is, what was considered a reference standard positive result) and whether final histological diagnoses were excluded from estimates of test and risk score accuracy. Accuracy estimates for some test or risk scores varied depending on which target population was considered.

Summary estimates of the sensitivity of ROMA were highest in studies that excluded borderline ovarian tumours and malignancies other than epithelial ovarian cancer from analyses (but estimates of specificity were lowest from such studies). Similar effects were also seen for RMI 1 (threshold of 200). In direct comparisons of RMI 1 and ROMA, results differed depending on target condition. In 2 studies there was no statistically significant difference in the sensitivity and specificity values of ROMA and RMI 1 (threshold of 200). In a further direct comparison of these tests (using summary estimates produced from 2 studies with a target condition of epithelial ovarian cancer), the summary estimate of specificity was significantly lower for ROMA when compared to RMI 1 (threshold of 200).

No studies reported a direct comparison between Overa (MIA2G) and RMI 1; but 1 study did provide a comparison of Overa (MIA2G) with ROMA. The sensitivity estimate for Overa (MIA2G) was significantly higher than for ROMA, and specificity estimate significantly lower.

Several studies (4 plus an unpublished interim report) reported a direct comparison between RMI 1 (threshold of 200) and Simple Rules. Summary estimates of sensitivity for Simple Rules were significantly higher, and summary estimates of specificity were significantly lower, than summary estimates for RMI 1 (threshold of 200). [redacted]. When inconclusive results from the Simple Rules were assessed by expert subjective assessment, summary estimates of specificity were significantly higher, and estimates of sensitivity significantly lower, than when inconclusive results were assumed to be malignant. One study reported no significant effect on Simple Rules' accuracy when scans were done by less experienced practitioners.

Fewer studies (1 study and an unpublished interim report) reported direct comparisons between ADNEX and RMI 1 (threshold of 200). Sensitivity was significantly higher for ADNEX and specificity was significantly lower than the

summary estimates for RMI 1 (threshold of 200). [redacted]. In a direct comparison between Simple Rules and ADNEX, summary estimates of specificity were similar and the summary estimate of sensitivity was higher for ADNEX.

Menopausal status

Not all studies provided accuracy estimates stratified by menopausal status. Several studies reported the accuracy of ROMA for pre- and post-menopausal sub-populations. The studies were variable in terms of the size and direction of effect that menopausal status had on sensitivity and specificity estimates. Summary estimates for Simple Rules and ADNEX suggest that the specificity of these tests may be higher for people who are pre-menopausal when compared to people who are post-menopausal. Only 1 study reported Overa (MIA2G) accuracy stratified by menopausal status; without significant difference in sensitivity and specificity reported.

Cost effectiveness

The studies used to provide estimates of diagnostic accuracy for the economic model were those with a target condition of all malignant tumours including borderline. This was because the EAG considered that the populations included in analysis in these studies most closely represented the population that the tests would be used on in clinical practice. The number of identified studies with such data available varied between tests; with estimated accuracy for some tests being based on 1 study (ROMA using Abbott ARCHITECT) or 2 studies (Overa [MIA2G] and ROMA using Roche Elecsys).

ADNEX was generally cost effective at thresholds of £20,000 and £30,000 per QALY gained, although Simple Rules was cost effective in some scenario analyses. This included a scenario where the beneficial effect of treatment for people with an ovarian malignancy in a specialist MDT (rather than secondary care) was at its lowest level. Also, RMI 1 (threshold of 250) was cost effective at a maximum acceptable ICER of £20,000 per QALY gained if additional

surgical costs in the specialist MDT were assumed (based on an assumption that some patients will have extensive surgery for ovarian cancer). In analyses done for pre-menopausal and post-menopausal subgroups, the ADNEX model was cost effective at thresholds of £20,000 and £30,000 per QALY gained.

4 Issues for consideration

Clinical effectiveness

No study directly compared all test and risk scores (that is, assessed their diagnostic accuracy in the same patient cohort). Studies were identified that had direct comparisons between the tests or risk scores and RMI 1 (except Overa [MIA2G]); however, this was mostly for RMI 1 at a threshold of 200, rather than 250 (the comparator for this assessment).

Studies identified for a particular test or risk score differed in the target condition that they assessed diagnostic accuracy for (that is, studies varied in which conditions they considered to be reference standard positive results). None of the target conditions in included studies exactly matched the scope of the NICE guideline on ovarian cancer (which predominantly considered epithelial ovarian cancer). Studies which used epithelial ovarian cancer as a target condition did so by retrospectively excluding participants from analyses based on their final histological diagnosis, therefore study populations used to produce accuracy estimates would differ from populations that the tests would be applied to in clinical practice.

The diagnostic accuracy estimates of tests and risks scores differed depending on which target condition was being considered. For example, the accuracy of the ROMA and RMI 1 differed depending on whether non-epithelial ovarian cancers and borderline ovarian tumours were included in analysis. However, when participants with borderline ovarian tumours and non-epithelial ovarian cancer were excluded from analyses, the sensitivity estimate for ROMA was not significantly different from RMI 1 (threshold of

200), and the specificity estimate was significantly lower. Although based on small patient numbers, analyses of included studies suggested that non-epithelial ovarian cancer accounted for a large proportion of false negative results for the ROMA. The NICE guideline on ovarian cancer recommends the use of alternative serum markers alongside CA125 and the RMI 1 (alpha fetoprotein [AFP] and beta human chorionic gonadotrophin [beta-hCG]) for women under 40 years with suspected ovarian cancer to help identify non-epithelial ovarian cancer; therefore potentially cases of non-ovarian cancer that were missed by ROMA would be detected by these tests. Further analysis of this was outside the scope of this assessment.

A potential difference between the patient population in the included studies and clinical practice is that all participants in studies had surgery (allowing their disease status to be confirmed by histology). In clinical practice, tests and risk scores may be used not only to decide where surgery should be done, but also to decide between surgery and surveillance or conservative management. Therefore not all people in the NHS secondary care on whom the tests or risk scores are used will have surgery. The EAG commented that this difference in populations may account for the relatively high prevalence of malignancy derived from studies that were used to produce summary estimates of test and risk score accuracy (21.3%), which was used in modelling. The EAG suggested that a lower prevalence of malignancy may affect test and risk score performance in practice.

The performance of the ADNEX model and Simples Rules may not be as effective in NHS secondary care if the level of skill and expertise of ultrasonographers in routine NHS practice differs from those in the identified studies. The EAG assessed the impact of practitioner expertise on ADNEX and Simple Rules performance; however studies did not always report levels of experience or expertise. The largest datasets for both ADNEX and Simple Rules came from the IOTA study cohort which used experienced ultrasound practitioners or used pre-study training in use of the tools. One study did report no difference in Simple Rules performance between EFSUMB level 2 or

3 and level 1 examiners (with the provision of a half-day training in use of the Simple Rules tool). Further Simple Rules studies (Alcazar et al. 2013; Sayasneh et al. 2013) also used less experienced ultrasound operators and reported estimates of Simple Rules accuracy that were similar to overall summary estimates.

Data from an interim report of phase 5 of the IOTA study were included in summary estimates of the accuracy of Simple Rules and the ADENX model. This data has not been published and therefore has not been peer reviewed.

Not all studies provided estimates of risk score accuracy according to menopausal status (or an indication of the proportion of participants who were pre- or post-menopausal). Potential differences in the performance of the tests and risk scores between pre- and post-menopausal populations may therefore not be apparent from the available data.

The NIHR funded ROCkeTS (Refining Ovarian Cancer Test Accuracy Scores) study is currently underway and is due to report in 2019/2020. This study will evaluate the diagnostic accuracy and clinical utility of existing and novel risk prediction models in secondary care in the NHS. This will potentially include the RMI 1, ROMA and Simple Rules, as well as novel models not included in the scope of this assessment. The study is not restricted to people who are scheduled for surgery (that is, people who will have tissue available for histology to be used as a reference standard); participants who do not have surgery will be followed up at 12 months to assess their status. It is likely that the results of this study will be extremely relevant to future potential updates of this guidance.

It is not clear if any of the identified studies assessed the performance of tests and risk scores in people under 18 years. Ten studies could have included people from this age group (based on inclusion criteria); but it was not reported if any did so. Therefore the applicability of accuracy scores from the identified studies to this age group is uncertain.

Cost effectiveness

Varying numbers of studies were available to inform the estimates of test and risk score accuracy used in the base-case analysis. For the ROMA and Overa (MIA2G), only 1 or 2 studies informed estimates, none of which were done in the UK.

The EAG used data from studies with a target condition of 'all malignant tumours including borderline' for accuracy estimates used in modelling. The EAG considered that these studies most closely represented clinical practice, in that they did not exclude participants from analysis based on knowledge of their diagnosis. However it will include people with non-ovarian primary cancers and people with cancers that were excluded from the scope of the NICE guideline on ovarian cancer (such as germ cell tumours and sex cord-stromal tumours of the ovary).

Cost effectiveness analysis showed that tests or risk scores with higher sensitivities (ADNEX and Simple Rules) tended to be cost effective; with analysis robust to changes in most parameter values. However, the ADNEX model and Simple Rules both have lower specificity than the RMI 1. Use of these models would therefore increase the number of people with a benign mass who are referred to a specialist MDT. This can be illustrated with an example of a cohort of 1,000 people (assuming a prevalence of malignancy of 21.3%). The expected number of 'false' results (positive and negative) in this cohort are shown in table 22.

Table 22 Numbers of incorrectly diagnosed people in a hypothetical cohort of 1,000 patients, using sensitivity and specificity values from the economic model [a].

	Expected number of people with a malignancy not referred to a specialist MDT (false negatives)	Expected number of people referred to specialist MDT with a benign condition (false positive)
ADNEX [b]	8	243
Simple Rules [b]	12	188
RMI 1 (threshold of	68	78

200) [b]		
RMI 1 (threshold of 25) [c]	11	385
[a] Values differ from those reported in the diagnostics assessment report which were calculated using different sensitivity and specificity values. [b] Accuracy of tests used from summary estimates used in cost effectiveness modelling, see tables 23 and 24in the diagnostics assessment report. [c] Summary estimates used to provide diagnostic accuracy values, see appendix 7 on page 297 of the diagnostics assessment report for details.		

Any effect of an increased workload for specialist MDTs that may occur if ADNEX or Simple Rules is adopted has not been considered in the model. The EAG noted clinical opinion that a major impact of false positive results is the time and resources taken away from true positive cases (that is, people who have a malignancy). Larger workloads in specialist MDTs may increase waiting times and adversely affect outcomes for patients. Also, limits to the service capacity of specialist MDTs was also not taken into account in the model.

5 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

All people with cancer are covered under the disability provision of the Equality Act (2010) from the point of diagnosis.

The Simple Rules classification system and the ADNEX model have not been validated for use with people who are pregnant. The use of transvaginal ultrasound probes for scans (which is needed by these models) may also be inappropriate for people under 18 years.

The ROMA has not been validated for use with people under 18 years. Also, the Overa (MIA2G) test is only indicated for use with people 18 years and older.

6 Implementation

Currently CA125 assays are widely used by laboratories, and a range of different analysers are available to run the assays. HE4 assays from different manufacturers may need particular analyser platforms to run the assay, which may differ from analysers currently used by laboratories to run CA125 assays. Also, manufacturers will often recommend that particular CA125 assays are used with their HE4 assays to calculate ROMA scores, which may differ from CA125 assays currently used by laboratories.

Adoption of tests or risk scores that result in an increased number of referrals to specialist MDTs (for example, more benign cases referred) will potentially have consequences for the operation of these services.

Transvaginal ultrasound (which is needed for both the Simple Rules and ADNEX model) may not be widely available in secondary care. Also, it may not be an acceptable procedure for all people (in comparison with transabdominal ultrasound).

Expertise needed to do and interpret ultrasound scans according to Simple Rules and ADNEX requirements may not be widely available, or may need training before they can be routinely used. Also, while protocols for obtaining samples for HE4 assays, and running the assay, may be similar to existing CA125 protocols, expertise in interpreting results (such as ROMA scores) may not be widely available.

7 Authors

Thomas Walker
Topic Lead

Frances Nixon
Technical Adviser

June 2017

Appendix A: Sources of evidence considered in the preparation of the overview

A. The diagnostics assessment report for this assessment was prepared by Kleijnen Systematic Reviews Ltd:

Westwood M, Ramaekers B, Lang S et al. Tests in secondary care to identify people at high risk of ovarian cancer: A systematic review and cost effectiveness analysis. May 2017.

B. The following organisations accepted the invitation to participate in this assessment as stakeholders. They were invited to attend the scoping workshop and to comment on the diagnostics assessment report.

Manufacturer(s) of technologies included in the final scope:

- Abbott Laboratories
- Fujirebio Diagnostics AB
- International Ovarian Tumor Analysis (IOTA) group
- Roche Diagnostics Ltd
- Vermillion, Inc

Other commercial organisations:

- None

Professional groups and patient/carer groups:

- British Association of Gynaecological Pathologists
- Institute of Biomedical Science
- Medicines & Healthcare Products Regulatory Agency
- Ovacome
- Royal College of Pathologists
- Royal College of Physicians
- Target Ovarian Cancer

Research groups:

- Ovarian Cancer Action

Associated guideline groups:

- None

Others:

- British In Vitro Diagnostics Association
- Department of Health
- Healthcare Improvement Scotland
- NHS England
- Welsh Government

Appendix B: Glossary of terms

Adnexal mass
A mass in the pelvis close to one or other side of the womb.

Endometriosis
A condition where tissue that behaves like the lining of the womb (the endometrium) is found outside the womb.

False negative
A result that appears negative but should have been positive, i.e. a test failure.

False positive
A result that appears positive but should have been negative, i.e. a test failure.

Gynaecological oncologist
A surgeon who is an expert in the treatment of cancer affecting the female reproductive system.

Menopause
The permanent cessation of ovarian function.

Metastases/Metastatic
Spread of cancer away from the original site to somewhere else in the body, usually via the bloodstream or the lymphatic system.

Overall survival
The time someone lives after a diagnosis of cancer. Often quoted as a percentage chance of living a number of years (e.g. 5 or 10).

Sensitivity
The proportion of individuals who have disease correctly identified by the study test.

Specificity
The proportion of individuals who do not have a disease and who are correctly identified by the study test.

Appendix C: Recommendations from NICE CG122

1.2 Establishing the diagnosis in secondary care

1.2.1 Tumour markers: which to use?

1.2.1.1 Measure serum CA125 in secondary care in all women with suspected ovarian cancer, if this has not already been done in primary care.

1.2.1.2 In women under 40 with suspected ovarian cancer, measure levels of alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (beta-hCG) as well as serum CA125, to identify women who may not have epithelial ovarian cancer.

1.2.2 Malignancy indices

1.2.2.1 Calculate a risk of malignancy index I (RMI I) score (after performing an ultrasound; see recommendation 1.2.3.1) and refer all women with an RMI I score of 250 or greater to a specialist multidisciplinary team.

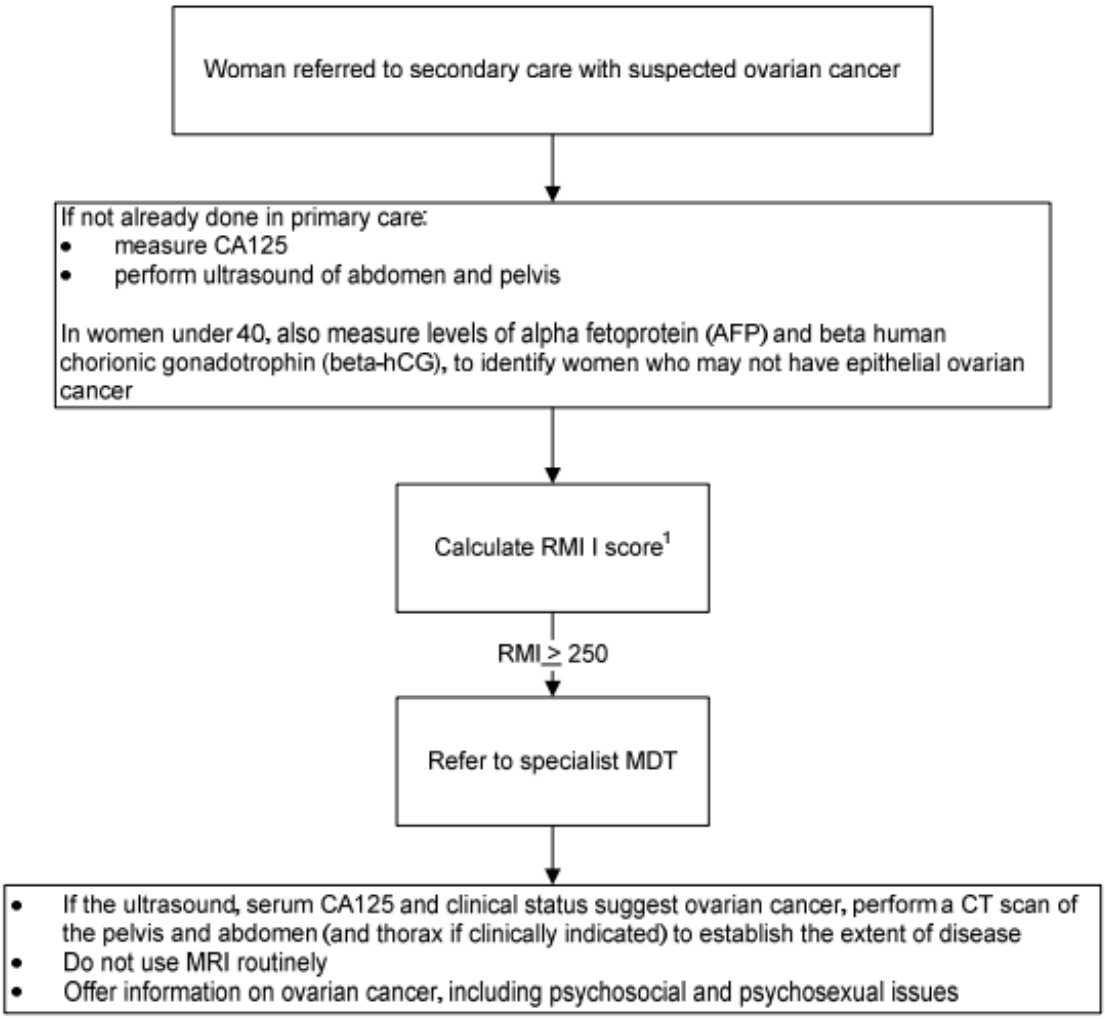
1.2.3 Imaging in the diagnostic pathway: which procedures?

1.2.3.1 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.

1.2.3.2 If the ultrasound, serum CA125 and clinical status suggest ovarian cancer, perform a CT scan of the pelvis and abdomen to establish the extent of disease. Include the thorax if clinically indicated.

1.2.3.3 Do not use MRI routinely for assessing women with suspected ovarian cancer.

Appendix D: Algorithm describing testing in secondary care for people referred with suspected ovarian cancer (from NICE CG122 full guideline).



[1] Risk of malignancy index (RMI I) calculated as described in Appendix D of NICE CG122.