

Diagnostics Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence – Protocol

Title of project

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

Name of External Assessment Group (EAG) and project lead

Kleijnen Systematic Reviews Ltd. Assessment Group

Project lead: Marie Westwood

Second Contact: Shona Lang

Kleijnen Systematic Reviews Ltd

Unit 6, Escrick Business Park

Riccall Road

Escrick

York YO19 6FD

Tel: 01904 727983

Email: marie@systematic-reviews.com; shona@systematic-reviews.com

Health economics lead: Bram Ramaekers

Second contact: Manuela Joore

Department of Clinical Epidemiology and Medical Technology Assessment

Maastricht University Medical Centre & CAPHRI School for Public Health and Primary Care

Department of Health Services Research

Maastricht University

P.O. Box 5800

6200 AZ Maastricht

The Netherlands

Tel: +31-43-3877399 Email: bram.ramaekers@mumc.nl; m.joore@mumc.nl

Plain English Summary

Ovarian cancer is the sixth most common cancer in women in the UK and it is more common in older women, with approximately half of new cases occurring in women over 65. There is a much better chance that ovarian cancer can be treated successfully if it is found early, however, early stage ovarian cancer can be difficult to diagnose. Symptoms, such as feeling bloated, feeling full early or having poor appetite, abdominal or pelvic pain, and needing to urinate more often or more urgently can be early warning signs of ovarian cancer, but can also be caused by other conditions (e.g. infections or benign growths).

A diagnosis of ovarian cancer is confirmed using tissue samples taken using needle biopsy (a procedure done under local anaesthetic, involving insertion of needles into the abdomen in order to take tissue samples, usually from a number of sites) or during exploratory laparoscopy (a procedure which uses a small camera, inside a flexible tube, to see inside the body, and which requires a general anaesthetic). Because these procedures can be unpleasant, result in bruising and discomfort and carry a small risk of complications, it is important to find tests which can help to select those women who really need to have them, i.e., those who are more likely to have ovarian cancer.

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

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

1 Decision problem

1.1 Population

The primary indication for this assessment is optimisation of the routine secondary care assessment of people with suspected ovarian cancer to decide whether a patient should be referred to a specialist gynaecological oncology multi-disciplinary team (MDT). The assessment is being conducted in the context of an up-date to current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management.¹ Recommendations on primary care assessment are being up-dated separately by the guidelines group. The relevant population is people of any age, including pre- and post-menopausal women, who have been referred to secondary care for the investigation of suspected ovarian cancer. This assessment will include data for people of any age, with a view to informing possible research recommendations for the population under 18 years, but no cost-effectiveness modelling will be undertaken for this population. People with a previous history of ovarian cancer, who are being monitored for possible recurrence, and those referred directly from primary care to a specialist MDT, are outside the scope of this assessment.

1.2 Target condition

The target condition, for this assessment, is ovarian cancer. Ovarian cancer is a term describing a group of cancers arising from cells in or near the ovaries. Ovarian cancers can be classified based on tissue type (epithelial ovarian tumours, sex cord-stromal tumours and germ cell tumours), with epithelial carcinomas being the most common type (90%) of primary ovarian cancer; non-epithelial ovarian cancers are more common in pre-menopausal women.² The target conditions included in this assessment are those covered by NICE clinical guideline (CG122), i.e. epithelial ovarian cancer, fallopian tube carcinoma, primary peritoneal carcinoma, and borderline ovarian cancer;¹ excluded target conditions are pseudomyxoma peritonei, relapsed ovarian, fallopian tube or peritoneal cancer, germ cell tumour of the ovary, sex cord stromal tumours of the ovary.

Ovarian cancers are staged using the four-stage FIGO system:³

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

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

Grade 1	Well differentiated
Grade 2	Moderately differentiated
Grade 3	Poorly differentiated/undifferentiated

Ovarian cancer is the sixth most common cancer in females in the UK (2013), accounting for 4% of all new cancer cases in females.^{4, 5} In 2013, there were 7,284 new cases of ovarian cancer in the UK, giving an age-standardised incidence rate of 23.3 per 100,000.^{4, 5} Ovarian cancer accounts for around 5% of cancer deaths in women in the UK; 2014 statistics recorded 4,100 ovarian cancer deaths.⁶ Incidence of ovarian cancer is strongly related to age, with 2011-2013 data indicating that approximately half (53%) of new cases were diagnosed in women over 65 years of age.^{4, 5} Ovarian cancer mortality is also strongly related to age at diagnosis.⁶

Data from the Office of National Statistics, published by Cancer Research UK, indicate that, although ovarian cancer incidence rates have increased, overall, since the 1970's, UK age standardised incidence rates decreased by 6% in the decade between 2002-2004 and 2011-2013.⁷ However, it remains the case that a high proportion of women (58%) are diagnosed at an advanced stage (stage III or IV) and 21% have metastases at diagnosis.⁸ Ovarian cancer survival is strongly related to stage at diagnosis; 2012 data showed that the one year and five year survival rates for women diagnosed at stage I were 97% and 90% versus 53% and 4% for women diagnosed at stage IV.⁶ Improving early diagnosis is therefore a priority and variation in the performance of testing strategies for the detection of different stages of ovarian cancer should be considered. The majority of studies about ovarian cancer diagnosis concern epithelial carcinomas, however, there is some evidence to indicate that the diagnostic performance of tumour markers and risk scores may vary between tumours of different tissue types;⁹ possible effects of tumour tissue type on estimates of test performance should also be considered.

The overall prevalence of malignancy in women with a symptomatic ovarian cyst is lower in pre-menopausal women (approximately 1 in 1000) than in women over the age of 50 (approximately 3 in 1000).¹⁰ It has been suggested that CA125 results should be interpreted cautiously in pre-menopausal women because of the high rate of false positives resulting from various non-malignant conditions (fibroids, endometriosis, adenomyosis, pelvic infection).¹⁰ It is therefore important to consider the effects of menopausal status on the performance of testing strategies, either by stratification of data from test accuracy studies or by including menopausal status in risk models (as in RMI).

1.3 Intervention technologies

Serum tumour markers are used in the secondary care investigation of people with suspected ovarian cancer; they are not considered to be 'stand-alone' diagnostic tests, but are used in conjunction with other test, signs and symptoms to assess risk of malignancy. An estimate of an individual's risk of malignancy can inform decisions about specialist referral,

further testing and treatment. It is anticipated that these risk assessment tools will be used in secondary care, for people in whom ultrasound imaging suggests confined disease or low volume disease outside the pelvis (stage I to IIIb).

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

The current standard assessment (RMI) has been reported as having poor sensitivity, approximately 63% at an operating threshold of 200.¹² If referral decisions are based on RMI score at this threshold, there remains the potential for significant numbers of people with ovarian cancer to remain un-referred and to experience consequential delays in diagnosis and detrimental effects on prognosis. A systematic review of studies comparing HE4, CA125 and the ROMA score reported similar overall sensitivity estimates for HE4 and CA125 (76% and 79%, respectively) and a higher sensitivity (85%) for the ROMA score.⁹ Sensitivity estimates were lower for early stage cancer (55% for both HE4 and CA125, and 74% for the ROMA score).⁹ Risk scores with higher sensitivity are needed to facilitate prompt referral of the appropriate patient group.

1.3.1 ROMA score

The ROMA score uses serum HE4 and serum CA125 levels, along with menopausal status, to generate an individualised estimate of the risk that a person has ovarian cancer. Initially a predictive index (PI) value is calculated using a formula which differs depending on whether the person is pre- or postmenopausal (equations (1) and (2) in Table 1). This PI value can then be used to calculate the ROMA score (equation (3) in Table 1).¹³ The ROMA score is intended for use in women who present with adnexal mass (i.e. following ultrasound examination). Manufacturers of HE4 assays have obtained CE marking for the use of these assays, in the context of a ROMA score, in combination with a specific CA126 assay or assays; if a CA125 level has been obtained in primary care, using a different assay, this will need to be repeated in secondary care before a ROMA score can be calculated.

Table 1: ROMA equations

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

Cut-off values for ROMA are used to classify individuals as having low or high risk of having epithelial ovarian cancer. Recommended thresholds can differ depending on the tumour marker assays used (Section 1.3.2 below).

The manufacturers of tumour marker assays state that the ROMA has not been validated for the following patient groups: people previously treated for malignancy, people currently being treated with chemotherapy, and people less than 18 years of age; this assessment will include data for people of any age, with a view to informing possible research recommendations for the population under 18 years.

1.3.2 Serum HE4 assays

There are currently four commercial HE4 assays that are available for use in the UK NHS; a summary of the key technical characteristics of these assays is provided below (Table2).

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

Name of assay	Company	Detection limit	Detection range	Assay time
ARCHITECT HE4	Abbott Diagnostics	15 pmol/L	20 – 1500 pmol/L	28 minutes*
HE4 EIA	Fujirebio Diagnostics	15 pmol/L	15-900 pmol/L	Approximately 3-4 hours**
Elecsys® HE4 - Human epididymal protein 4	Roche Diagnostics	15 pmol/L	15-1,500 pmol/L	18 min***
* Time is for analyser to complete sample analysis once initiated ** Time is for one EIA plate (42 samples, duplicate measurement) *** Report time is dependent on whether other tests are carried out on the same sample, but typically take less than 30 minutes				

ARCHITECT HE4 (Abbott Diagnostics)

The ARCHITECT HE4 assay is a chemiluminescent micro particle immunoassay (CMIA) for the quantitative determination of HE4 in human serum. The assay is designed for use on an immunoassay analyser, specifically the Abbott ARCHITECT i2000SR or ARCHITECT i1000SR analysers. Additional materials required to run the assay are: ARCHITECT HE4 Assay software file, ARCHITECT HE4 Calibrators, ARCHITECT HE4 Controls, ARCHITECT Multi-Assay Manual Diluent, ARCHITECT Pre-Trigger Solution, ARCHITECT Trigger Solution, ARCHITECT Wash Buffer, ARCHITECT Reaction Vessels, ARCHITECT Sample Cups, ARCHITECT Septum and ARCHITECT Replacement Caps.

The results of the assay are intended to be used in conjunction with the ARCHITECT CA 125 II assay as an aid in estimating the risk of epithelial ovarian cancer in people presenting with a pelvic mass who will undergo surgical intervention. The company recommends that the HE4 assay results are used in the ROMA, using the following cut-off values for ROMA scores, based on obtaining a specificity of 75%:

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

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

Lumipulse G HE4 (Fujirebio Diagnostics)

The Lumipulse G HE4 is a Chemiluminescent Enzyme Immunoassay (CLEIA) for the quantitative measurement of HE4 in human serum. The assay is designed for use on the LUMIPULSE G System (either the LUMIPULSE G1200 or LUMIPULSE G600 immunoassay analysers). Samples are run using immunoreaction cartridges, which contain reagents and into which samples are added. Further materials required for the assay are: Lumipulse G HE4 calibrators, Lumipulse G substrate solution, Lumipulse G wash solution, Lumipulse G specimen diluent I, sampling tips for Lumipulse system, soda lime for Lumipulse system and Lumipulse G dilution cartridges.

The assay is intended for use in conjunction with CA125 levels (measured using the Lumipulse G CA125II assay) as an aid in estimating the risk of epithelial ovarian cancer in

premenopausal and postmenopausal women presenting with a pelvic mass who will undergo surgical intervention.

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

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

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

A further HE4 assay is also available from Fujirebio: the HE4 EIA assay, a manual, enzyme immunometric assay for the quantitative determination of HE4 in human serum. Clinical experts commented that manual kits would be too labour intensive for general use in the NHS; therefore this assay has not been included in the scope of this assessment.

Table 3: Inclusion criteria Recommended cut-off points for ROMA values produced using the ARCHITECT HE4 EIA assay, in conjunction with various CA125 assays, to provide a specificity of 75%

Assays used to determine HE4 and CA125 serum concentrations	Cut-off ROMA values
ARCHITECT HE4 EIA + ARCHITECT CA125II	<p><i>Premenopausal patients:</i> ROMA value $\geq 13.1\%$ indicates a high risk of finding epithelial ovarian cancer. ROMA value $< 13.1\%$ indicates a low risk of finding epithelial ovarian cancer.</p> <p><i>Postmenopausal patients:</i> ROMA value $\geq 27.7\%$ indicates a high risk of finding epithelial ovarian cancer. ROMA value $< 27.7\%$ indicates a low risk of finding epithelial ovarian cancer.</p>
ARCHITECT HE4 EIA + CanAg CA125 II	<p><i>Premenopausal patients:</i> ROMA value $\geq 12.5\%$ indicates a high risk of finding epithelial ovarian cancer ROMA value $< 12.5\%$ indicates a low risk of finding</p>

	epithelial ovarian cancer. <i>Postmenopausal patients:</i> ROMA value $\geq 14.4\%$ indicates a high risk of finding epithelial ovarian cancer ROMA value $< 14.4\%$ indicates a low risk of finding epithelial ovarian cancer.
ARCHITECT HE4 EIA + Lumipulse G CA125II	<i>Premenopausal patients:</i> ROMA value $\geq 13.1\%$ indicates a high risk of finding epithelial ovarian cancer. ROMA value $< 13.1\%$ indicates a low risk of finding epithelial ovarian cancer. <i>Postmenopausal patients:</i> ROMA value $\geq 27.7\%$ indicates a high risk of finding epithelial ovarian cancer. ROMA value $< 27.7\%$ indicates a low risk of finding epithelial ovarian cancer.

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

Elecsys HE4 immunoassay (Roche Diagnostics)

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

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

The assay is intended to be used in conjunction with the Elecsys CA 125 II assay as an aid in estimating the risk of epithelial ovarian cancer in premenopausal and postmenopausal people with a pelvic mass. The company recommend that the HE4 and CA125 assay results are used in the ROMA.

A ROMA calculator is available on the Roche website and is also available as an app. Roche specifies that this calculator is for use only with values derived from Elecsys CA 125II and HE4 immunoassays from Roche Diagnostics. The company suggest the following cut-off values for ROMA scores, based on obtaining a specificity of 75%:

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

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

1.3.3 Simple rules ultrasound model (International Ovarian Tumour Analysis [IOTA] group)

Several models designed to estimate the risk of malignancy based on features observed by ultrasound scans have been developed by the International Ovarian Tumour Analysis (IOTA) group, including the Simple Rules model. The group are currently in the process of applying for a CE mark for this model. The model uses a morphological scoring system based on the presence of ultrasound features (described as rules) to characterise an ovarian mass as benign or malignant and requires the use of transvaginal ultrasound. There are five 'rules' describing features of malignant tumours (M-rules) and five rules that describe benign tumours (Table4).^{14, 15} Because use of the simple rules model requires specialist training in interpreting ultrasound images in relation to these rules, it is assumed that using the simple rules model in the specified population will require a secondary care ultrasound examination (i.e. repeat examination where ultrasound has been conducted in primary care).

Table 4: Simple Rules ultrasound model (IOTA group)

M-rules (rules for predicting a malignant tumour)	B-rules (rules for predicting a benign tumour)
<ul style="list-style-type: none"> • Irregular solid tumour • Ascites present • Four or more papillary structures • Irregular multilocular solid tumour with largest diameter 100mm or more • Very strong blood flow (colour score 4). 	<ul style="list-style-type: none"> • Unilocular • Solid components present, with largest solid component having a largest diameter of less than 7mm • Acoustic shadows present • Smooth multilocular tumour with largest diameter less than 100mm • No blood flow (colour score 1).

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

- if any M-rules (and no B-rules) apply the mass is classified as malignant

- if any B-rules (and no M-rules) apply the mass is classified as benign
- if both M and B rule (or neither) apply the mass is unclassifiable and the IOTA group state that there are then a number options:
 - classify the mass as malignant
 - refer the patient to an expert ultrasound operator for a second opinion
 - use alternative imaging techniques
 - use the Simple Rules risk model¹⁶ to calculate risk of malignancy using the ultrasound features described in the Simple Rules model

No specific make or model of ultrasound device is required for model inputs. A transvaginal probe is required and image quality must be of sufficient quality to allow the ultrasound features specified by the model to be seen. The group state that the approach to evaluating masses required by the model is not more time consuming than a standard ultrasound scan. The IOTA group organise 1 day courses that teach the techniques for classifying masses required by the model, with participants assessed with a multiple choice test. An on-line training tool which will be freely accessible to NHS practitioners is also being developed. In addition to this training, the group also recommend that practitioners should have completed 300 gynaecological scans. Software is not required to run the Simple Rules model, however a program is available from the IOTA website <http://www.iotagroup.org/simplerules/>.

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

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

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

1.3.5 Ova2/Overa (Vermillion)

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

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

The levels of these biomarkers present in serum are determined using immunoassays run on the Roche cobas 6000 system. The Overa 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 indicated for use in people over the age of 18 years with a pelvic mass for whom surgery may be considered. It is intended for use as part of preoperative assessment to help decide if a person presenting with a pelvic mass has a high or low risk of ovarian malignancy.

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

1.3.6 The Risk of Malignancy Index (RMI)

The RMI, used at thresholds other than that currently recommended in NICE clinical guidelines (see section 1.4 below) will be considered as an alternative intervention technology.

1.4 Comparator

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

$$\text{RMI I score} = U \times M \times \text{CA125}$$

U: ultrasound score – 1 point scored for the presence of each of the following features: multilocular cysts, solid areas, metastases, ascites, bilateral lesions. U=0 (0 points), U=1 (1 point) or U=3 (2-5 points).

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

CA125: serum CA125 concentration – measured in IU/ml

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

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

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

1.5 Reference standard

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

1.6 Care pathway

Primary care assessment and criteria for referral to secondary care

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

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

The recommendations about serum CA125 testing were contentious at the time of publication²⁰ and were based on limited evidence; the evidence review supporting CG122²¹ cited two systematic reviews that included data about the accuracy of CA125, neither of which were specific to primary care settings. One systematic review included studies of women with suspected adnexal mass who were being treated surgically (secondary care, higher prevalence population)²² and the other included studies in any population (screening to women with known and adnexal mass).² Whilst the use of CA125 testing in primary care may result in the detection of some additional cancers in women with non-specific symptoms, it should be noted that the majority of the studies that informed the evidence

summary were conducted in women with abdominal or pelvic masses; CA125 is only raised in around 50% of early stage disease.^{10, 23} The summary sensitivity estimates derived from the two reviews included in the evidence summary for CG122 were 80% (range 28 to 100%) and 78% (range 50 to 100%), respectively. Myers 2006 also stratified accuracy data by menopausal status and noted that sensitivity estimates for CA125, in pre-menopausal women, ranged from 50% to 74%.² These data would appear to suggest that a negative CA125 test result alone is unlikely to be considered adequate to rule-out ovarian cancer in symptomatic women without a pelvic or abdominal mass, particularly if they are pre-menopausal.

More recent guidance about cancer diagnosis, NICE guidance NG12 (Suspected cancer recognition and referral), published in 2015,²⁴ reproduces the recommendations from NG122 with no up-date; therefore, the criteria for referring a woman with suspected ovarian cancer to secondary care are not specifically defined.

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

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

Establishing a diagnosis in secondary care

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

- Measure serum CA125 in secondary care in all women with suspected ovarian cancer, if this has not already been done in primary care
- In women under 40 with suspected ovarian cancer, measure levels of alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (beta-hCG) as well as CA125, to identify women who may not have epithelial ovarian cancer
- Calculate the risk of malignancy index I (RMI I) score (after performing ultrasound) and refer all women with an RMI I score of 250 or greater to a specialist multidisciplinary team

- Perform an ultrasound of the abdomen and pelvis as the first imaging test in secondary care for women with suspected ovarian cancer, if this has not already been done in primary care
- If the ultrasound, serum CA125 and clinical status suggest ovarian cancer, perform a CT scan of the pelvis and abdomen to establish the extent of disease. Include the thorax if clinically indicated
- Do not use MRI routinely for assessing women with suspected ovarian cancer
- If offering cytotoxic chemotherapy to women with suspected ovarian cancer, first obtain a confirmed tissue diagnosis by histology (or by cytology if histology is not appropriate) in all but exceptional cases
- Offer cytotoxic chemotherapy for suspected ovarian cancer without a tissue diagnosis (histology or cytology) only:
 - in exceptional cases, after discussion at the multidisciplinary team **and**
 - after discussing with the woman the possible benefits and risks of starting chemotherapy without a tissue diagnosis
- If surgery has not been performed, use histology rather than cytology to obtain a tissue diagnosis. To obtain tissue for histology:
 - use percutaneous image-guided biopsy if this is feasible
 - consider laparoscopic biopsy if percutaneous image-guided biopsy is not feasible or has not produced an adequate sample.
 Use cytology if histology is not appropriate

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

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

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

The Royal College of Obstetricians & Gynaecologists (RCOG) and the British Society for Gynaecological Endoscopy (BSGE) have produced a joint guideline about the management of suspected ovarian masses in pre-menopausal women. This guideline aimed to clarify when ovarian masses can be managed in a 'benign' gynaecological service and when referral

to a gynaecological oncological service is needed.¹⁰ the guideline notes the importance of thorough history taking, including risk factors, and careful physical examination, including abdominal and vaginal examination and determination of the presence or absence of local lymphadenopathy. Specific statements about pre-operative assessment are:

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

The Royal College of Radiologists iRefer radiological investigation guidelines tool recommends that CT of the abdomen and pelvis has a role in identifying people who may benefit from chemotherapy or cytoreductive surgery. MRI of the abdomen and pelvis is recommended for specialised investigation when enhanced CT is contraindicated or for problem-solving. PET-CT is indicated as a specialised investigation for difficult management situations.²⁵

Management of early (stage I) ovarian cancer

NICE guideline NG122 includes the following recommendations about the overall management of women with suspected early (stage I) ovarian cancer¹ and NICE Technology Appraisal Guidance (TA55) provides recommendations about first-line chemotherapy regimens²⁶:

- perform retroperitoneal lymph node assessment as part of optimal surgical staging

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

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

Management of advanced (stage II to IV) ovarian cancer

NICE guideline NG122 includes the following recommendations about the management of women with advanced (stage II to IV) ovarian cancer¹ and NICE Technology Appraisal Guidance (TA55 and TA284) provides recommendations about first-line chemotherapy regimens^{26, 27}:

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

Further recommendations about chemotherapy regimens for women with recurrent ovarian cancer can be found in NICE Technology Appraisal Guidance TA389, TA381 and TA285.²⁸⁻³⁰

1.7 Summary of the decision problem

Current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management¹ recommends that serum cancer antigen 125 (CA125) should be measured in secondary care, in all women with suspected ovarian cancer for whom serum CA125 has not already been measured in primary care. CA125 levels can inform clinical decision making in secondary care and are not used in isolation; NICE guideline CG122 specifically recommends the calculation of a risk malignancy index I (RMI I) score, which includes CA125. CG122 does not currently include any recommendations on HE4 or risk scores or testing algorithms (other than RMI). An up-date to the section of CG122 that deals with establishing a diagnosis in secondary care is planned in order to assess the potential role of alternative risk scores in assessing people with suspected ovarian cancer for possible referral to a specialist gynaecological oncology MDT and to consider the best way to incorporate tumour markers and other tests in the decision making process.

This assessment will systematically review the evidence about the comparative performance of alternative risk scores that include CA125, HE4 or ultrasound (detailed in section 1.3 above) to guide referral decisions for people with suspected ovarian cancer in secondary care. The assessment will focus on direct comparisons between the interventions described and RMI I, using the referral threshold of ≥ 250 , (current practice as indicated in CG122),¹ in order to best inform comparative cost-effectiveness modelling. However, we will also include assessments of the accuracy of individual risk scores. Data will be collected on the accuracy and comparative accuracy of different risk scores, alternative cut-offs and risk scores used in combination or sequence with one or more additional tests, in order to determine the best way to incorporate tumour markers and ultrasound findings in the diagnostic process. We will also include prediction modelling studies, which report the development and validation of multivariable prediction models intended to be used to guide individual patient care. Studies of this type can provide information on the independent predictive value of tumour markers and ultrasound findings in clinical context, i.e. does adding a tumour marker or markers or a specific ultrasound finding to the diagnostic work-up significantly change our ability to reach the correct diagnosis or determine prognosis?

2 Objectives

The overall objective of this assessment is to summarise the evidence on the clinical and cost-effectiveness of using alternative risk scores that include CA125, HE4 or ultrasound (detailed in section 1.3 above) to guide referral decisions for people with suspected ovarian cancer in secondary care. Current guidance, NICE clinical guideline (CG122) Ovarian cancer: recognition and initial management¹ recommends that serum cancer antigen 125 (CA125) should be measured in secondary care, in all people with suspected ovarian cancer for whom serum CA125 has not already been measures in primary care. CA125 levels are a component of secondary care investigation and are not used in isolation; NICE guideline CG122 specifically recommends the calculation of a risk malignancy index I (RMI I) score, which includes CA125, and referral to a specialist MDT for people with an RMI score ≥ 250 . CG122 Does not currently include any recommendations on HE4 testing or alternative methods of risk scoring. An evaluation of current evidence is needed to assess the clinical utility and cost-effectiveness of alternative methods of risk scoring. This assessment will inform a planned up-date to the section of CG122 that deals with establishing a diagnosis in secondary care, by addressing the following research questions:

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

3 Methods for assessing clinical effectiveness

Systematic review methods will follow the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care³¹ and the NICE diagnostics assessment programme manual.³²

3.1 Inclusion and exclusion criteria

Separate inclusion criteria were developed for each of the clinical-effectiveness questions. These are summarised in Table 5.

Table 5: Inclusion criteria

Question	What is the accuracy of alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥ 250 (current practice), where the target condition is histologically confirmed ovarian?	What are the effects of using alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥ 250 (current practice), on clinical management decisions and clinical outcomes?
Participants:	People of any age with suspected ovarian cancer, who have not previously been treated for ovarian cancer and are not currently receiving chemotherapy Secondary care*	
Setting:		
Interventions (index test):	Alternative methods of risk scoring or RMI used at thresholds other than 250, as described in section 1.3 above**	
Comparators:	Risk malignancy Index (RMI) [§]	
Reference standard:	Histological examination of surgically resected tissue sample ^{§§}	Not applicable
Outcomes:	Diagnostic accuracy (the numbers of true positive, false negative, false positive and true negative test results), where the target condition is histologically confirmed ovarian cancer	Diagnosis of ovarian cancer confirmed by pathological examination of a biopsy, or prognostic outcomes for ovarian cancer (e.g. stage at diagnosis, differentiation status, suitability for surgical intervention/curative treatment, overall survival, progression-free survival)
Study design[§]:	Diagnostic cohort studies, directly comparing one or more interventions (index tests) and the comparator~	Prediction modelling studies, randomised and non-randomised controlled trials

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

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

[§] Not applicable for prediction modelling studies

^{§§} Studies which use histological examination of a biopsy sample or follow-up (ideally for a minimum of 12 months) of people with a risk score below the referral threshold, who do not have a pelvic mass requiring surgery as the reference standard, will also be included

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

3.2 Search strategy

Search strategies will be based on the specified risk scores (ROMA, IOTA simple ultrasound rules, ADNEX, OVA2/Overa and RMI) and target condition (ovarian cancer), as recommended in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.^{31, 33, 34} Additional supplementary searches will be carried out as necessary. Searches for studies for costs and quality of life will be developed separately where required.

Candidate search terms will be identified from target references, browsing database thesauri (e.g. Medline MeSH and Embase Emtree), existing reviews identified during initial scoping searches. These scoping searches will be used to generate test sets of target references, which will inform text mining analysis of high-frequency subject indexing terms using Endnote reference management software. Strategy development will involve an iterative approach testing candidate text and indexing terms across a sample of bibliographic databases, aiming to reach a satisfactory balance of sensitivity and specificity. Search strategies will be developed specifically for each database and the keywords ovarian cancer and RMI, ROMA, IOTA simple ultrasound rules, OVA2/Overa and ADNEX will be adapted according to the configuration of each database.

The following databases will be searched for relevant studies:

- MEDLINE (Ovid)
- MEDLINE In-Process Citations and Daily Update (Ovid)
- EMBASE (Ovid)
- Cochrane Database of Systematic Reviews (CDSR) (Wiley)
- Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley)
- Database of Abstracts of Reviews of Effects (DARE) (Wiley)
- Health Technology Assessment Database (HTA) (Wiley)
- International Network of Agencies for Health Technology Assessment (INAHTA) Publications (Internet) <http://www.inahta.org/publications/>
- NIHR Health Technology Assessment Programme (Internet) <http://www.nets.nihr.ac.uk/programmes/hta>
- Aggressive Research Intelligence Facility (ARIF) database (Internet) <http://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/index.aspx>
- PROSPERO (International Prospective Register of Systematic Reviews) (Internet) <http://www.crd.york.ac.uk/prospero/>

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

- NIH ClinicalTrials.gov (Internet) (<http://www.clinicaltrials.gov/>)

- EU Clinical Trials Register (Internet) <https://www.clinicaltrialsregister.eu/ctr-search/search>
- WHO International Clinical Trials Registry Platform (ICTRP) (Internet) <http://www.who.int/ictcp/en/>

Key conference proceedings, to be identified in consultation with clinical experts, will be screened for the last five years. References in retrieved articles and relevant systematic reviews will be checked.

There will be no restrictions on date, language or publication status and searches will take into account generic and other product names for the intervention. An example search strategy is presented in Appendix 1; this will be adapted as necessary following consultation with clinical experts. The main Embase strategy for each search will be independently peer reviewed by a second Information Specialist, using the CADTH Peer Review checklist.³⁵ Identified references will be downloaded in Endnote X6 software for further assessment and handling. References in retrieved articles will be checked for additional studies. The final list of included papers will also be checked on PubMed for retractions, errata and related citations.³⁶⁻³⁹

3.3 Review strategy

Two reviewers will independently screen titles and abstracts of all reports identified by the searches and discrepancies will be discussed. Full copies of all studies deemed potentially relevant, after discussion, will be obtained and two reviewers will independently assess these for inclusion; any disagreements will be resolved by consensus or discussion with a third reviewer.

Where available, data will be extracted on the following: study design/details, participant characteristics (e.g. age, pre- or post-menopause, presenting symptoms, other risk factors, etc.), details of the risk score and its component tests (and any other test used), manufacturer, antibody, limit of quantitation, definition of cut-off, detection method [including analyser used], ultrasound method, definition of a positive ultrasound, etc.), details of reference standard (including imaging method used to guide biopsy, number of samples, laparoscopy or open surgery, etc.) details of the covariables (e.g. clinical risk factors, family history, ultrasound or other test results) and the dependent variable(s) (e.g. confirmed diagnosis of ovarian cancer, stage at diagnosis or survival outcomes) used in any prediction model studies, and test combination or prediction model performance outcome measures. Data will be extracted by one reviewer, using a piloted, standard data extraction form. A second reviewer will check data extraction and any disagreements will be resolved by consensus or discussion with a third reviewer.

3.4 Quality assessment strategy

The methodological quality of included diagnostic accuracy studies will be assessed using QUADAS-2⁴⁰ and the methodological quality of any prediction model studies will be assessed using the PROBAST tool.⁴¹ If any randomised controlled trials (RCTs) are identified, these will be assessed using the Cochrane risk of bias tool.⁴² The results of the quality assessment will be used for descriptive purposes to provide an evaluation of the overall quality of the included studies and to provide a transparent method of recommendation for design of any future studies. Where sufficient data are available the results of quality assessment may be used to inform stratified meta-analyses in order to explore the impact of individual components of study quality upon the findings of the review. Quality assessment will be undertaken by one reviewer and checked by a second reviewer, any disagreements will be resolved by consensus or discussion with a third reviewer.

3.5 Methods of analysis/synthesis

If available data allow, summary estimates of the comparative sensitivity and specificity, together with 95% confidence intervals (CIs) and prediction regions, of different risk scores and referral thresholds (alone or in series with other tests), versus RMI using a referral threshold of 250, will be calculated. We will use the bivariate/hierarchical summary receiver operating characteristic (HSROC) random effects model to generate summary estimates and an SROC curve.^{28, 43-45} The results of any prediction model studies will be summarised in relation to the comparative accuracy results and the potential of individual tumour markers and ultrasound findings to improve model performance will be considered. If more than one RCT evaluates the same clinical outcome and the same intervention, then data will be pooled on treatment effect (e.g. hazard ratio, odds ratio, relative risk, weighted mean difference). The DerSimonian and Laird random effects model will be used to generate summary estimates together with 95% CIs. All estimates of the relative accuracy/effectiveness of different interventions will be derived from direct, within study comparisons. Where sufficient data are available, clinically relevant subgroup analysis will be considered (e.g. age, pre- versus post menopause, people with versus without pelvic mass, people with versus without a family history of ovarian cancer). Similarly, if sufficient data are available, different target conditions (e.g. early stage versus early stage plus borderline, different histotypes of epithelial carcinoma [serous, mucinous, clear cell, endometrial], non-epithelial ovarian cancer).

Where meta-analysis is considered unsuitable for some or all of the data identified (e.g. due to the heterogeneity and/or small numbers of studies), we will employ a narrative synthesis. Typically, this will involve the use of text and tables to summarise data. These will allow the reader to consider any outcomes in the light of differences in study designs and potential sources of bias for each of the studies being reviewed. Studies will be organised by research question addressed and by intervention evaluated. A detailed commentary on the major methodological problems or biases that affected the studies will also be included, together

with a description of how this may have affected the individual study results. Recommendations for further research will be made based on any gaps in the evidence or methodological flaws.

4 Methods for synthesising evidence of cost-effectiveness

4.1 Identifying and reviewing published cost-effectiveness studies

Search strategy

A systematic literature search will be performed to identify published full economic evaluations by searching the following databases:

- NHS Economic Evaluation Database (NHS EED) (Internet)
- MEDLINE (Ovid)
- MEDLINE In-Process Citations and Daily Update (Ovid)
- EMBASE (Ovid)
- EconLit (EBSCO)
- CEA Registry (<http://www.cearegistry.org/>)
- Research Papers in Economics (RePEc) (<http://repec.org/>)

A summary with the results and the methodological quality of the selected studies will be reported. Methodological quality will be assessed using the Drummond checklist.⁴⁶ Data extraction will focus on technologies compared, indicated population, main results in terms of costs, consequences and the incremental cost-effectiveness of the alternatives compared, but also on modelling methods and the sources of input parameters used.

Exploration of the literature regarding published utility and cost studies will be performed. The intention of this explorative review is to identify studies that can be used to support the development of a health economic model, and to estimate the model input parameters, that will aim to answer the research questions of this assessment, but not to perform a systematic review.

4.2 Evaluation of costs, quality of life and cost-effectiveness

Decision analytic modelling will be undertaken to determine the cost-effectiveness of alternative risk scores (including alternative RMI score thresholds), which include HE4, CA125 or ultrasound, compared to the RMI score with a referral threshold of ≥ 250 (current practice), when routinely used, in secondary care, to guide decisions about referral to a specialist MDT, for people with suspected ovarian cancer. More specifically, this population will consist of people with pelvic masses on ultrasound and/or elevated CA125 levels, assessed in primary care. The systematic review component assessment will include data for people of any age. The inclusion of studies conducted in people under 18 years old is intended to inform possible research recommendations for this group, in whom the intervention technologies have not been validated. For this reason, and because we anticipate that insufficient data will be available, no cost-effectiveness modelling will be undertaken for this population.

Diagnosis and treatment strategies

The analysis will consider the long-term consequences of using different risk scores to guide referral decisions. For risk scores for which performance is unclear, when feasible, assumptions will be made to provide some indication of the (range) of cost-effectiveness outcomes. For instance, published studies that report on the value of risk scores from initial diagnosis through to intermediate and final health outcomes may not be available for all risk scores listed in the scope.

Model structure

The economic model compares different risk scoring strategies and will consist of a decision tree to estimate the short-term outcomes (including test strategy results) followed by a long-term state-transition (i.e. Markov) model to estimate long-term consequences in terms of costs and QALYs. In order to be consistent with earlier related assessments, the economic model used in NICE clinical guideline 122 will be used as a starting point.¹ A simple draft model structure of the state-transition model, based on Appendix 1 of clinical guideline 122, is presented (Appendix 2). This model structure may be developed/expanded as necessary choices and definitions regarding the final structure of the model will depend on the findings from the literature review and consultation with clinical experts. In addition, the existence/availability of any other electronic models that reflect the cost-effectiveness of diagnosis and treatment pathways for these patients, and are representative of current care within the NHS, will be determined for useful information and/or methods.

Issues relevant to analyses:

- Longer term costs and consequences will be discounted using the UK discount rates of 3.5% of both costs and effects.

- Probabilistic sensitivity analyses will be performed using parameter distributions instead of fixed values.
- Decision uncertainty regarding mutually exclusive alternatives will be reflected using cost-effectiveness planes and cost-effectiveness acceptability curves.

Health outcomes

Utility values, based on literature or other sources, will be incorporated in the economic model. QALYs will be calculated from the economic modelling.

Costs

Resource utilisation will be estimated for the diagnostic tests strategies and treatments. Data for the cost analyses will be drawn from routine NHS sources (e.g. NHS reference costs, Personal Social Services Research Unit (PSSRU), British National Formulary (BNF)), discussions with clinical experts and with the manufacturers of the comparators.

5 Handling of information from the companies

All data submitted by the manufacturers/sponsors will be considered if received by the EAG no later than 20/02/2017. Data arriving after this date will be considered if practicable and at the discretion of the EAG. If the data meet the inclusion criteria for the review, they will be extracted and quality assessed in accordance with the procedures outlined in this protocol.

Any 'commercial in confidence' data provided by manufacturers, and specified as such, will be highlighted in blue and underlined in the assessment report (followed by company name in parentheses). Any 'academic in confidence' data provided by manufacturers, and specified as such, will be highlighted in yellow and underlined in the assessment report. Any confidential data used in the cost-effectiveness models will also be highlighted.

6 Competing interests of authors

None

7 Timetable/milestones

Milestones	Completion data
Draft protocol	25/10/2016
Final protocol	18/11/2016
Progress report	20/02/2017
Draft assessment report	19/04/2017
Final assessment report	18/05/2017

8 References

- [1] National Collaborating Centre for Cancer. *Ovarian cancer: the recognition and initial management of ovarian cancer [Internet]*. Manchester: NICE, 2011 [accessed 22.9.16] Available from: <https://www.nice.org.uk/guidance/cg122/evidence/full-guideline-181688797>
- [2] Myers ER, Bastian LA, Havrilesky LJ, Kulasingam SL, Terplan MS, Cline KE, et al. Management of adnexal mass. *Evid Rep Technol Assess (Full Rep)* 2006(130):1-145.
- [3] Prat J. FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication. *J Gynecol Oncol* 2015;26(2):87-9.
- [4] Office for National Statistics. *Cancer registration statistics, England: 2014. Cancer diagnoses and age-standardised incidence rates for all cancer sites by age, sex and region [Internet]*. Newport: ONS, 2016 [accessed 22.9.16]. 17 p. Available from: <http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2014/pdf>
- [5] Cancer Research UK. Ovarian cancer (C56-C57.4): 2013. Number of new cases, crude and European age-standardised (AS) incidence rates per 100,000 population, females, UK [Internet]. 2015 [accessed 22.9.16].
- [6] Cancer Research UK. *Ovarian cancer (C56-C57.4): 2012-2014. Average number of deaths per year and age-specific mortality rates per 100,000 population, UK [Internet]*. London: Cancer Research UK, 2015 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/deaths_crude_ovary_M14.pdf
- [7] Cancer Research UK. *Ovarian cancer (C56-C57.4): 2011-2013. Average number of new cases per year and age-specific incidence rates per 100,000 population, females, UK [Internet]*. London: Cancer Research UK, 2015 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/cases_crude_ovary_1.pdf
- [8] Cancer Research UK. *Ovarian cancer (C56): 2014. Proportion of cancers diagnosed at each stage, all ages, England [Internet]*. London: Cancer Research UK, 2016 [accessed 22.9.16] Available from: http://www.cancerresearchuk.org/sites/default/files/cstream-node/inc_stage_ovarian_0.pdf
- [9] Wang J, Gao J, Yao H, Wu Z, Wang M, Qi J. Diagnostic accuracy of serum HE4, CA125 and ROMA in patients with ovarian cancer: a meta-analysis. *Tumour Biol* 2014;35(6):6127-38.
- [10] British Society for Gynaecological Endoscopy, Royal College of Obstetricians & Gynaecologists. *Management of suspected ovarian masses in premenopausal women*

[Internet]. London: RCOG, 2011 [accessed 22.9.16] Available from: <http://www.ncin.org.uk/view?rid=1740>

[11] Khoja L, Nolan K, Mekki R, Milani A, Mescallado N, Ashcroft L, et al. Improved survival from ovarian cancer in patients treated in phase III trial active cancer centres in the UK. *Clin Oncol (R Coll Radiol)* 2016;28(12):760-5.

[12] Anton C, Carvalho FM, Oliveira EI, Maciel GA, Baracat EC, Carvalho JP. A comparison of CA125, HE4, risk ovarian malignancy algorithm (ROMA), and risk malignancy index (RMI) for the classification of ovarian masses. *Clinics (Sao Paulo)* 2012;67(5):437-41.

[13] Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;112(1):40-6.

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

[15] Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008;31(6):681-90.

[16] Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, et al. Predicting the risk of malignancy in adnexal masses based on the Simple Rules from the International Ovarian Tumor Analysis group. *Am J Obstet Gynecol* 2016;214(4):424-37.

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

[18] Van Calster B, Van Hoorde K, Froyman W, Kaijser J, Wynants L, Landolfo C, et al. Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors. *Facts Views Vis Obgyn* 2015;7(1):32-41.

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

[20] Olaitan A. NICE on ovarian cancer. Recommendations for detection in primary care are flawed. *BMJ* 2011;342:d3022.

[21] National Institute for Health and Care Excellence. *Ovarian cancer: the recognition and initial management of ovarian cancer. Evidence review [Internet]*. Manchester: NICE, 2016

[accessed 22.9.16] Available from:
<https://www.nice.org.uk/guidance/cg122/evidence/evidence-review-181688798>

[22] Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. Accuracy of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. *Eur J Obstet Gynecol Reprod Biol* 2009;142(2):99-105.

[23] Zurawski VR, Jr., Orjaseter H, Andersen A, Jellum E. Elevated serum CA 125 levels prior to diagnosis of ovarian neoplasia: relevance for early detection of ovarian cancer. *Int J Cancer* 1988;42(5):677-80.

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

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

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

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

[28] Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007;7:3.

[29] National Institute for Health and Care Excellence. *Olaparib for maintenance treatment of relapsed, platinum-sensitive, BRCA mutation-positive ovarian, fallopian tube and peritoneal cancer after response to second-line or subsequent platinum-based chemotherapy: technology appraisal 381* [Internet]. London: NICE, 2016 [accessed 12.10.16] Available from: <http://nice.org.uk/guidance/ta381>

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

[31] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care* [Internet]. York: University of York, 2009 [accessed

23.3.11] Available from:
<http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm>

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

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

[34] Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol* 2011;64(6):602-7.

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

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

[37] Waffenschmidt S. Assessing the completeness of systematic reviews via the “related articles” function and/or a simple structured Boolean search in PubMed – a pilot study (B202) [Oral presentation]. *19th Cochrane Colloquium*. Madrid, 19-22 October 2011.

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

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

[40] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-36.

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

[42] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

- [43] Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PMM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58(10):982-90.
- [44] Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;61(11):1095-103.
- [45] Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8(2):239-51.
- [46] Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford: Oxford University Press, 2005.

Protocol Appendix 1: Clinical effectiveness search

Embase: 1974 to 16 November 2016

Searched: 17.11.16

- 1 exp ovary cancer/ (96907)
- 2 uterine tube tumor/ (1263)
- 3 (AOSCa\$ or HGSC or EOC or HGSOC or LGSC or LGSOC or OVCA\$ or dysgerminom\$).ti,ab,ot. (9147)
- 4 ((ovar\$ or high-grade serous or low-grade serous or sertoli-leydig cell or fallopian or oviduct or uterine or uterus or tubal) adj5 (Cancer\$ or adenocarcin\$ or adeno-carcin\$ or tumo?r\$ or sarcoma\$ or neoplas\$ or metasta\$ or meta-sta\$ or carcino\$ or oncogenesis or malignan\$ or choriocarcinom\$ or teratoma\$ or cystadenocarcin\$ or rhabdomyosarcom\$ or rhabdo-myosarcom\$ or rhabdosarcom\$ or leiomyosarcoma\$ or leio-myosarcom\$ or androblastom\$ or arrhenoblastom\$ or adenoma\$ or lesion\$ or oncolo\$)).ti,ab,ot. (135513)
- 5 peritoneum cancer/ (3860)
- 6 (peritoneum or borderline or epithelial or primary peritoneal).ti,ab,ot. (405383)
- 7 or/5-6 (407687)
- 8 ovar\$.ti,ab,ot. (278490)
- 9 7 and 8 (29906)
- 10 1 or 2 or 3 or 4 or 9 (167512)
- 11 ((risk adj4 malignan\$ adj4 index) or (risk adj4 malignan\$ adj4 indice\$) or RMI).ti,ab,ot. (1385)
- 12 (menopau\$ or perimenopaus\$ or premenopaus\$ or postmenopaus\$ or POF or climacteric or (change adj2 life)).ti,ab,ot. or menopause/ (136215)
- 13 (Ultraso\$ or phonophoresis or sonication or sonification or ultra sound or ultrashell or sonograph\$ or doptone\$ or echograph\$ or echogram\$ or echosound\$).ti,ab,ot. or ultrasound/ or sonography/ (591434)
- 14 (CA125\$ or CA 125\$ or ca 12-5\$ or (antigen adj2 "125") or (mucin adj1 "16") or mucin16 or (muc adj1 "16") or muc16).ti,ab,ot. (11890)
- 15 CA 125 antigen/ (13565)
- 16 14 or 15 (16868)
- 17 12 and 13 and 16 (616)
- 18 11 or 17 (1867)
- 19 ovarian malignancy algorithm/ (1)
- 20 (ROMA or (Ovar\$ adj5 Algor\$)).ti,ab,ot. (2492)
- 21 (human epididymis protein 4 or human epididymal protein 4 or WAP four disulfide core domain protein 2 or wap 4 disulfide core domain protein 2 or WFCD2 or EDDM4 or WAP5 or wap four disulfide core domain 2 or wap 4 disulfide core domain 2 or HE 4 or HE4).ti,ab,ot. (943)
- 22 human epididymis protein 4/ (502)
- 23 or/21-22 (1023)
- 24 12 and 16 and 23 (234)
- 25 19 or 20 or 24 (2582)
- 26 (IOTA or international ovarian tumo?r analysis).ti,ab,ot. (844)
- 27 ((Simple adj3 rules) or (simple adj3 descriptors) or SRrisk or b-rules or m-rules).ti,ab,ot. (1794)
- 28 13 or 26 (592162)
- 29 27 and 28 (66)
- 30 (adnex\$ adj8 (model\$ or score\$ or assess\$)).ti,ab,ot. (461)
- 31 (ova2 or overa).ti,ab,ot. (78)
- 32 follitropin/ (56471)

- 33 (Follicle stimulating hormone\$ or FSH or follitropin or fertiline fertinom p or follicotropin folliculostimulating hormone\$ or follitrophin or follitropin\$ or folltropin\$ or 9002-68-0).ti,ab,ot,rn. (67373)
- 34 or/32-33 (67533)
- 35 apolipoprotein A1/ (16266)
- 36 (apolipoprotein A1 or apo a1 or apo hdl 3 or apo hdl iii or apo high density lipoprotein 3 or apolipoprotein a 1 or apolipoprotein a i or apoprotein a1 or apoprotein ai or apoprotein a 1 or apoprotein a i).ti,ab,ot. (9151)
- 37 or/35-36 (18709)
- 38 transferrin/ (27773)
- 39 (transferrin or siderophilin or transferrin?emia or transferrins or trf or 82030-93-1).ti,ab,ot,rn. (42230)
- 40 or/38-39 (42315)
- 41 16 and 23 and 34 and 37 and 40 (3)
- 42 31 or 41 (81)
- 43 18 or 25 or 29 or 30 or 42 (4849)
- 44 10 and 43 (1167)

Protocol Appendix 2: Draft model structure

