BMJ Open

Refining Ovarian Cancer Test accuracy Scores (ROCkeTS) - protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-010333
Article Type:	Protocol
Date Submitted by the Author:	21-Oct-2015
Complete List of Authors:	Sundar, Sudha; University of Birmingham, Cancer Sciences; City Hospital, Pan BirminghamGynaecological Cancer Centre Rick, Caroline; University of Birmingham, BCTU Dowling, Francis; University of Birmingham, BCTU Au, Pui; University of Birmingham, BCTU Snell, Kym; University of Birmingham, Public Health, Epidemiology & Biostatistics Rai, Nirmala; University of Birmingham, Cancer Sciences Champaneria, Rita; University of Birmingham, BCTU Stobart, Hilary; Patient and Public Involvement, Neal, Richard; Bangor University, North Wales Centre for Primary Care Research Davenport, Clare; University of Birmingham, Epidemiology ,Public Health and Biostatistics Mallett, Susan; University of Birmingham, School of Health & Population Sciences Sutton, Andrew; University of Birmingham, Health Economics Kehoe, Sean; University of Birmingham, of Cancer Studies Timmerman, Dirk; KU Leuven, Department of Obstetrics and Gynaecology Bourne, Tom; Imperial College London, Queen Charlottes and Chelsea Hospital; KU Leuven, Department of Obstetrics and Gynaecology Van Calster, Ben; KU Leuven, ; KU Leuven, Department of Obstetrics and Gynaecology Gentry-Maharaj, Aleks; Universlty College London, Women's Cancer Menon, Usha; University College London, Women's Cancer Deeks, Jon; University of Birmingham, Public Health, Epidemiology and Biostatistics; University of Birmingham, BCTU
Primary Subject Heading:	Obstetrics and gynaecology
Secondary Subject Heading:	Diagnostics, Evidence based practice, Oncology, Radiology and imaging
Keywords:	Gynaecological oncology < GYNAECOLOGY, Ultrasound < RADIOLOGY & IMAGING, GYNAECOLOGY

Refining Ovarian Cancer Test accuracy Scores (ROCkeTS) - protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer

Sudha Sundar,[1,2] Caroline Rick,[3] Francis Dowling,[3] Pui Au,[3] Kym Snell,[4] Nirmala Rai ,[1,4] Rita Champaneria,[4] Hilary Stobart,[5] Richard Neal,[6] Clare Davenport,[4] Susan Mallett,[4] Andrew Sutton,[7] Sean Kehoe,[1] Dirk Timmerman[8], Tom Bourne.[8,9] Ben Van Calster.[8] Aleksandra Gentry-Maharaj,[10] Usha Menon,[10] Jon Deeks.[3,4]

1. School of Cancer Sciences, University of Birmingham, UK
2. Pan Birmingham Gynaecological Cancer Centre, City Hospital, Birmingham, UK
3. Birmingham Clinical Trials Unit, University of Birmingham, UK
4. Diagnostic Test Accuracy Group, University of Birmingham, UK
5. Patient Representative
6. Primary Care, Bangor University, North Wales, UK
7. Health Economics, University of Birmingham, UK
8. Department of Obstetrics and Gynaecology, KU Leuven, Belgium
9. Queen Charlotte's and Chelsea Hospital, Imperial College, London, UK

Gynaecological Cancer Research Centre, Department of Women's Cancer, Institute for Women's Health, UCL, UK

Author for correspondence

Sudha Sundar

Senior Lecturer in Gynaecological Oncology

School of Cancer Sciences and Pan Birmingham Gynaecological Cancer Centre, City Hospital and

University of Birmingham,

Edgbaston,

Birmingham.

B15 2TT

s.s.sundar@bham.ac.uk

Abstract

Introduction: Ovarian cancer (OC) is associated with nonspecific symptoms such as bloating, making accurate diagnosis challenging: only 1 in 3 women with OC present through primary care referral. NICE guidance recommends sequential testing with CA125 and routine ultrasound in primary care. However these diagnostic tests have limited sensitivity or specificity. Improving accurate triage in women with vague symptoms is likely to improve mortality by streamlining referral and care pathways. The Refining Ovarian Cancer Test Accuracy Scores (ROCkeTS; HTA 13/13/01) project will derive and validate new tests/risk prediction models that estimate the probability of having OC in women with symptoms. This protocol refers to the prospective study only (phase 3).

Methods and Analysis:– ROCkeTS comprises 4 parallel Phases. The full ROCkeTS protocol can be found at http://www.birmingham.ac.uk/ROCKETS. Phase 3 – prospective test accuracy study. The study will recruit 2,450 patients from 15 UK sites. Recruited patients complete symptom and anxiety questionnaires, donate a serum sample and undergo ultrasound scored per IOTA criteria. Recruitment is at rapid access clinics, emergency departments and elective clinics. Models to be evaluated include the IOTA ultrasound simple rules, ADNEX and novel models derived from analysis of existing datasets.

Estimates of sensitivity, specificity, c-statistic (area under Receiver operating curve curve), Positive Predictive Value and Negative Predictive value of diagnostic tests evaluated and a calibration plot for models will be evaluated.

ROCkeTS has received ethical approval from the NHS West Midlands REC (14/WM/1241) and is registered on the Controlled trials website (ISRCTN17160843) and the NIHR Cancer and Reproductive Health portfolios.

Strengths

- ROCkeTS conforms to PROBE design for biomarker evaluation and STARD criteria for test accuracy studies
- Stringent efforts to collect outcome data to prevent ascertainment bias
- Inbuilt ultrasound quality control.

Limitations

- Recruitment is at secondary care so the population will be less heterogenous than that seen in primary care.

Introduction

Ovarian cancer (OC) is the seventh most common cancer in women worldwide, with 239,000 new cases diagnosed in 2012.[1] In the UK, OC has an annual incidence of 7,116 women and causes 4,271 deaths; the lifetime risk of developing OC is 1 in 54.[2] 70% of patients will present at an advanced stage (stages III/IV). The International Cancer Benchmarking Project showed that the reason that OC survival in the UK is significantly lower than other western countries seems to be related to a lower proportion of patients receiving treatment and surviving the first year after cancer diagnosis and is likely due to a delay in diagnosis.[3] 5-year survival rates are 43% overall but over 90% for early stage tumours.[4] High grade serous is the most common histotype (80%). Worryingly, long term survival from OC has remained static over the last decades at 30%.[2]

Previously considered a 'silent killer', it is now recognised that patients with OC suffer from a number of nonspecific symptoms. These include abdominal bloating, distension, feeling full quickly and/or loss of appetite, pelvic/abdominal pain, increased urinary urgency and/or frequency, unexplained weight loss, fatigue or changes in bowel habit. These symptoms are very common.[5 6] Interrogation of United Kingdom General Practice databases suggest that on average 1 in 2 women between the ages of 45 and 74 present once a year to their General practitioner (GP)/primary care doctor with these symptoms. Abdominal bloating alone,[5 6] is documented in 16-30% of women presenting to GPs.[7] Diagnostic challenges are considerable given (1) the low incidence of OC (a GP sees a woman with OC once in 3-5 years) and (2) the low positive predictive value (PPV) of symptoms (only 1 in 400-600 symptomatic women have OC).[8 9] Unfortunately, these diagnostic challenges result in nearly 36% of women, subsequently diagnosed with OC, presenting to the GP with symptoms three or more times prior to diagnosis.[10] Two large prospective studies of symptom triggered testing for Ovarian Cancer suggest that symptom triggered testing using CA125 is likely to result in referral of a higher proportion of patients with resectable disease. [11 12]

The UK introduced symptom-triggered testing for Ovarian Cancer in women with vague symptoms in 2011. National Institute for Health and Care Excellence (NICE) guidelines recommend sequential testing using serum CA125 followed by pelvic ultrasound (USS) in women (particularly aged ≥50) presenting to primary care with symptoms –such as persistent abdominal distension/'bloating', feeling full and/or loss of appetite, pelvic/abdominal pain,

increased urinary urgency and/or frequency, unexplained weight loss, fatigue or changes in bowel habit on a persistent or frequent basis.[13] However, the NICE guidance does not specify the type of ultrasound abnormalities that should prompt referral. Current tests have limited sensitivity with CA125 being elevated only in 50% of women with early stage OC. Once referred to secondary care, women with complex masses considered benign can undergo laparoscopic or conservative management, whereas women with malignancy who undergo surgery by trained gynaecological oncologists have the best outcome.[14 15] Use of NICE guidance in practice is extremely variable. A survey of 258 GPs reported that the majority would refer patients on the basis of raised CA125 even if the USS was normal.[16] A recent audit revealed that the majority of referrals (90%) for suspected OC did not follow guidance. Referrals were heterogeneous with regard to which symptoms prompted these, what GPs considered to be abnormal CA125 level or abnormal USS.[17] Two thirds of women referred were premenopausal.[17]

Therefore, improved diagnostic tools in primary care and better pre-surgical triage in secondary care is likely to improve mortality in OC patients. Optimal diagnostic pathways for premenopausal women with suspected OC/complex ovarian mass and raised CA125 also need to be defined. Ovarian cysts in premenopausal women are extremely common, however it is important to recognise that about 1,000 women under 50 will be diagnosed with OC in the UK annually.[2]

The National Institute of Health Research (NIHR) Health Technology Assessment (HTA) programme commissioned **ROCkeTS – '<u>R</u>efining <u>o</u>varian <u>c</u>ancer <u>t</u>est accuracy scores'** 13/13/01 (1st October 2014 and 30th September 2018) which aims to identify, refine and validate tests and clinical risk scores (risk prediction models) that estimate the probability of having OC in post- and premenopausal women with symptoms, in primary and secondary care. The project comprises 4 phases/work packages, phase 1- systematic reviews of existing risk prediction algorithms and biomarkers to detect ovarian cancer, phase 2 - Interrogation of datasets from 2 large OC trials; the screening trial UKCTOCS and the pre-operative detection International Ovarian tumour analysis (IOTA) studies to derive or refine risk scores with biomarkers from the systematic review, phase 3 – a prospective study to validate new scores in women newly referred to secondary care with symptoms of OC, phase 4 – a model of the

diagnostic pathway across primary and secondary care. A cost consequence analysis of testing pathways will be delivered as part of ROCkeTS. The full ROCkeTS protocol can be found at http://www.birmingham.ac.uk/ROCKETS

The ROCkeTS project evaluates risk prediction models derived from conventional tests – symptoms, ultrasound variables and commercially available serum blood tests.

The Phase 3 prospective study will recruit 2,450 premenopausal and postmenopausal symptomatic women over 23 months (recruitment started in June 2015). Patients will be recruited through rapid access clinics, emergency departments or through routine clinic referrals, reflecting the heterogeneity of diagnostic routes for OC in the UK.

Recruited patients consent to complete a validated symptom questionnaire, donate a serum sample and undergo an ultrasound scan scored as per the IOTA criteria. Standard care pathways are followed beyond this point. Patients are triaged to receive surgery or conservative management based on standard of care model Risk of malignancy index (RMI) scores for postmenopausal women, using a threshold of 250 in postmenopausal women or as per local unit practice in premenopausal women. Participants and treating doctors will be blinded to the new biochemical serum test results which will be performed in batches as these are not part of standard clinical care. Outcome data will be collected from histology reports in women undergoing surgery or biopsy and from follow-up over 12 months of those managed conservatively. Symptom scores, serum biomarker tests and ultrasound data will be analysed at the end of follow-up to validate risk prediction models derived in phases 1 and 2.

ROCKETS STUDY DESIGN

Aim of the study

- To derive and validate risk prediction models that estimate the probability of having OC for women with symptoms suggestive of OC for postmenopausal and premenopausal women.
- To identify optimal risk thresholds for the models that can guide patient management.

Design

The ROCkeTS study is a single arm prospective cohort diagnostic test accuracy study to evaluate existing and novel risk prediction models for pre and post-menopausal women with symptoms.

A test accuracy study is different to an effectiveness study in that randomisation of subjects is not involved. It is designed to generate a comparison of measurements obtained by index tests with those obtained by reference standards. In this way the accuracy of index tests can be estimated. A reference standard is a test that confirms or refutes the presence or absence of disease beyond reasonable doubt. Therefore it is sometimes also known as the "gold standard".

Here, the reference standard will be histology of tissues taken from patients who proceed to surgery or biopsy or follow-up using structured templates at a minimum of 12 months after presentation. The accuracy of the index test will be compared against that of the comparator test, the existing standard risk prediction score RMI. In ROCkeTS, the index test (novel risk prediction models) will be derived in phases 1 and 2 and validated in phase 3. We will identify biochemical markers, symptom indices and USS as likely components of a novel risk prediction model, as these may be implemented across primary and secondary care. Therefore we will collect symptom questionnaires, blood and USS data in the study to be analysed and validated at the end of the study.

Possible components of the new risk prediction model/s

Symptoms

Case-control studies demonstrate that symptom questionnaires have good diagnostic accuracy; however, they need to be refined for use by patients in primary care[18 19] as the duration of symptoms preceding diagnosis is uncertain.[19] Symptom questionnaires may help triage patients prior to referral and would help standardise symptoms for any prediction model. This is particularly important, given the subjective nature of eliciting symptoms through unstructured clinical history taking and the existing audit evidence that they are interpreted variably by GPs who will only see few cases of OC in their practice. A robust symptom score that can triage referral based on a questionnaire may be very useful.

Biochemical markers

A number of serum biomarker tests and multiple-marker based algorithms (ROMA, OVA1) have been identified in the last decade. Abnormal Human Epididymis 4 (HE4) biomarker levels may improve risk stratification for OC. A recent systematic review reports that HE4 may improve the diagnostic performance of CA125, however studies showed considerable heterogeneity.[20]

Ultrasound based models –IOTA risk prediction models

After publication of an agreement on terms, definitions and measurement methods to describe adnexal masses,[21] the IOTA collaboration set out multicentre studies on large cohorts of patients presenting with an adnexal mass. The IOTA database has enabled both previously developed prediction models to be tested and novel prediction models and rules to characterise ovarian pathology prior to surgery to be developed and validated.[22,23] In a systematic review with meta-analysis, IOTA algorithms such as the simple rules were identified to be the best pre-surgical diagnostic tools to characterise adnexal masses, with improved performance over the RMI.[24] Although the RCOG included the simple rules in their green top guideline for evaluating ovarian pathology in premenopausal women, IOTA models are not commonly used in NHS clinical practice. [25] Recently the ADNEX (Assessment of Different NEoplasias in the adneXa) model was published. As a multiclass prediction model, it differs from all other models by differentiating between malignant and benign masses, and also discriminating between 4 types of malignant tumours (borderline ovarian tumours, stage I ovarian cancer, stage II-IV ovarian cancer and metastatic tumours of other primary origin). ADNEX still needs an extensive external validation, but is considered to be promising. [22 26-28]

Target Population

1. Post and premenopausal women with symptoms of suspected OC. Symptoms are as defined by NICE including but not restricted to persistent or frequent abdominal distension, feeling full (early satiety) and/or loss of appetite, pelvic or abdominal pain, increased urinary urgency and/or frequency. Symptoms listed here are not an exhaustive list.

2. Patients referred with symptoms from GP as suspected OC.

Comparator

Risk of Malignancy Index cut-off of 250.[29]

Source of potential participants

Patients referred to secondary care as Outpatients; either as urgent 2 week or routine referrals, USS clinics, Inpatient and emergency presentation.

Inclusion and Exclusion Criteria

Multicentre trial across the UK who meet the eligibility criteria described below.

Inclusion criteria:

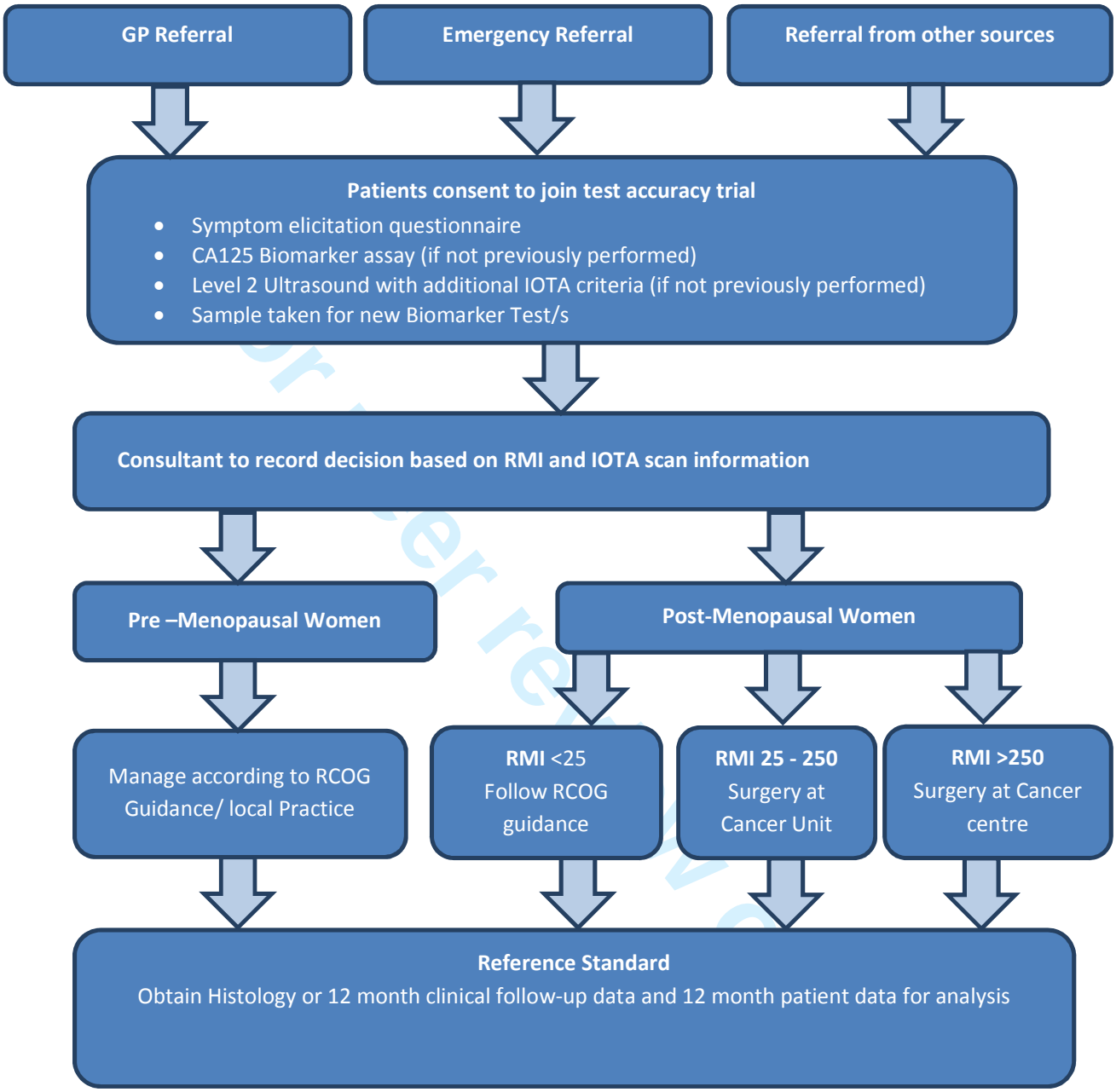
- Pre and postmenopausal women with symptoms of suspected OC and either raised CA125 or abnormal USS. Menopause is defined as >12 months of amenorrhoea.
- Aged between 16 and 90 years.
- Patients able to provide informed consent.

NB – Patients with >120 days delay between initial registration IOTA scan and surgery will need a repeat IOTA scan prior to surgery

Exclusion criteria:

- Ultrasound (USS) reveals non-ovarian pathology e.g. fibroids or simple ovarian cysts < 5cm in size (very low risk of malignancy).
- Patients with normal pelvis USS are excluded from the study.
- Patients who decline transvaginal scan.
- Previous ovarian malignancy
- Pregnant patients
- Patients with previous bilateral oophorectomy
- Active non ovarian malignancy – Women with a past history of cancer are only eligible if there are no documented persistent or recurrent disease and have not received treatment for this in the last 12 months. This exclusion does not apply to patients with premalignant disease eg cervical intra-epithelial neoplasia or patients receiving Tamoxifen/other drugs to prevent breast cancer recurrence

Figure 1: Flow chart



STUDY PROCEDURES AND TESTS

The Index tests

There are three index tests that will be performed in the prospective single arm test accuracy study.

1. Participants entering the study will complete a symptom elicitation questionnaire and anxiety questionnaire (STA6) and Impact of event score.

2. Ultrasonographers will record the USS variables and score the USS using IOTA simple rules and ADNEX models.[22 26 27] For most women in the trial, this will only mean some additional data being collected during their USS appointment. For a small number of women, this may mean an additional USS scan after consent.

3. Participants will have an additional blood sample taken at baseline for biomarker assessment at the end of the study. Serum will be banked for testing at a later point. Treating doctors will be blinded to any new biomarker assessments done as part of this trial. Details of blood sample collection will be provided in a lab manual.

Apart from the study tests, all other aspects of participant management are entirely at the discretion of the local doctors and as per the Royal College of Gynaecologists (RCOG) guidelines for management of these participants.[25 29] Treating clinicians will be asked to record their treatment recommendations as per standard care and after any additional USS information in order to assess the impact of this test on care pathways.

Quality assurance of index tests

The performance of USS is subjective and operator dependent. Therefore sonographers/doctors carrying out USS will undergo a face-to-face IOTA training course provided as part of their participation in the ROCkeTS trial. Sites will commit to undergoing quality assurance during the ROCkeTS trial. Online USS training materials will be developed during this project for future use by the UK NHS. Quality assurance of testing will begin with a clearly documented staff training programme. A register of staff that have been trained, and had their competence assessed will be maintained, and only staff whose names appear on this list will be permitted to undertake scans within the ROCkeTS prospective trial. Staff will also

receive a site visit and assessment of their competence. Competence will be assessed by those authorised by the IOTA team.

Reference standard/Follow-up Schedule

Reference standard for the study will be histology of tissue taken at surgery or biopsy in women who are managed surgically following study enrolment. Outcome of participants referred for suspected OC that do not undergo surgery will be assessed by a follow-up visit at 12 months or by a telephone call or a questionnaire from the research nurse at 12 months, as per the local investigators' discretion and clinical assessment. Wellbeing will be ascertained at this follow-up. A structured template will be used.

Sample Size

The sample size has a 90% power to detect with 5% significance difference in accuracy between the existing test (RMI) and the new model. Due to the expected difference in performance in pre and postmenopausal women, separate sample sizes have been calculated. Sample sizes are calculated assuming independence of test errors and interim analyses will confirm parameter assumptions.

For postmenopausal women: Performance of RMI is assumed to be 70% sensitive and 90% specific.[29] 1,400 participants will be required to detect an increase in sensitivity of 10% (to 80%) and in specificity of 5% (to 95%). Based on a prevalence of 30% of OC in referred women (local audit figures), with sensitivity and specificity of RMI assumed to be 70% and 90%, a sample size of 1,333 provides 90% power to detect an increase of sensitivity to 80% and specificity to 95% in paired data (conservatively assuming independence of test errors). Allowing for a loss to follow-up of up to 5%, this gives a final sample size of 1,400.

For premenopausal women: Performance of RMI is assumed to have a sensitivity of 72% and a specificity of 46% (local audit figures). The trial is powered to detect an increase in sensitivity of 10% (to 82%) and in specificity of 10% (to 56%). Prevalence of OC in premenopausal women referred to secondary care is around 10%.[17] A sample size of 1,000 will provide 100 OC events in which to build new models combining symptom and test data (adequate events to model 10 predictor variables), and will provide 90% power to detect an increase in specificity of 8% (from 46% for RMI to 54%). With a predicted loss to follow-up of up to 5%, the final sample size required is 1,050 women.

Study Duration

We anticipate recruitment of 2,450 participants within 23 months, with a minimum of 12 months follow-up from the last participant entering the study.

Data collection

All information will be collected on standard proformas (Case report forms; CRF) and identified by study number. Information will be collated on paper forms and then either copied and sent to the coordinating centre for input or entered directly into the study database via a web interface. A dataset including age, ethnicity, parity, GP details and significant medical/surgical history will be collected. We aim to use the NHS number as the primary identifier when linking to national registries and to track individuals throughout the NHS. Additional data on outcomes such as cancer or non-cancereous conditions will be collected at follow-up.

Data will be collected on relevant medical, obstetric and gynaecological, surgical history, emotional impact as well as information on the symptoms that prompted GP referral or investigation. USS information will be collected. Data on the reference diagnosis will be obtained from the histopathology report and a structured template to assess wellbeing for participants who do not undergo surgery will be collected directly from the participants. Importantly, outcomes collected will include all conditions/diagnoses in women with nonspecific symptoms.

Outcome measures and costs

Validation of a risk prediction model and tests for estimating the probability of OC in women with suspected OC; the key outcome measures are the accuracy of the tests and models in terms of their discrimination ability (e.g. sensitivity, specificity) and calibration (observed versus predicted probabilities), and the identification of thresholds to guide patient management decisions.

Trial data collection will be undertaken prospectively for all participants in order to inform the costs for each pathway.

Analysis plan - Test accuracy

We will report estimates of sensitivity, specificity, c-statistic (area under Receiver operating curve (ROC) curve), PPV and Negative Predictive value (NPV) of tests evaluated and a calibration plot for models evaluated. Also, in terms of our new model derived from phase 2 of the wider ROCkeTS project, its improvement over existing models will be summarised by comparing the c-statistic and the calibration. We will also summarise the net-reclassification index for each new predictor that existing models omitted.

The risk prediction models derived in phases 1 and 2 of the ROCkeTS project will each produce a predicted risk of OC by 12 months for all the individuals in our study. Therefore, we will compare the observed outcome at 12 months with this predicted risk. The calibration (in terms of calibration slope) and discrimination (e.g. c-statistic) will be evaluated for the models derived and identified in phases 1 and 2, and their performance compared to the existing RMI model. The calibration will be shown visually by grouping women into deciles ordered by predicted risk and considering the agreement between the mean predicted risk and the observed events in each decile.

Analysis plan - Cost consequence analysis

Resource use for each of the diagnostic tests will be broken down and displayed along with their unit costs alongside the outcomes for each pathway. The resource usage will include the types of tests administered, the number of inpatient and outpatient consultations, and any operative procedures undertaken. This approach will help to show which are the major cost drivers for each of the diagnostic pathways and will be collected as part of the clinical CRF.

Study Conduct

The conduct of the study will be in accordance with the Research Governance Framework for Health and Social Care and/or the Research Governance Framework for Health and Community Care. The participant's written informed consent to participate in the trial will be obtained before any trial procedures or questionnaires are completed. The women's GP will be notified of her participation in the study with her consent (GP letter).

Participants will be free to withdraw from the study at any time without any effect on their standard of care, data and samples provided up to the point a participant withdraws will be

retained unless the participant expressly requests their removal. This is because analysis will be based on all recruited participants and per protocol.

Ethics and dissemination: This study has received ethics permission from the NHS West Midlands REC (Ref14/WM/1241) and has no specific safety considerations. Outputs will be disseminated through open access publications, on the website www.birmingham.ac.uk/ROCKETS

Funding Source

This study is funded by a grant from National Institute of Heath Research, Health Technology assessment HTA 13/13/01

Disclaimer

The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HTA, NIHR, NHS or the Department of Health.

Authors' contributions

SS, CR, FD,PA, KS, NR, RC, HS, CD, SM,AS, SK & JD conceptualised and designed and statistics for the study and writing the paper, RN, DT, TB, BC, AM and UM critical input into study design and writing of the manuscript.

Competing Interests statement

All authors reports grants from NIHR HTA during the conduct of the study. UM reports other from Abcodia, outside the submitted work. TB reports other from Samsung Electronics, other from Roche Diagnostics, outside the submitted work.

Acknowledgements

We thank the ROCkeTS Project Oversight committee for their kind input and guidance. Chair – Prof Peter Sasieni, Members – Mr Nigel Acheson, Mr Andy Nordin, Prof Peter Murchie, Dr Michael Weston, Simon Newman (Target Ovarian Cancer)

1. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F., GLOBOCAN 2012 v1.0 CIaMWICNI, Lyon FIAfRoCAfhgif.
2. CRUK. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/ovary/incidence/uk-ovarian-cancer-incidence-statistics, 2013.
3. Maringe C, Walters S, Butler J, Coleman MP, Hacker N, Hanna L, et al. Stage at diagnosis and ovarian cancer survival: evidence from the International Cancer Benchmarking Partnership. *Gynecol Oncol* 2012;127(1):75-82.
4. Cancer Research UK. Ovarian cancer statistics, 2014.
5. Wheatley G. Ovarian cancer. Usefulness of abdominal symptoms in early diagnosis. *Bmj* 2009;339:b3954.
6. Pitts MK, Heywood W, Ryall R, Smith AM, Shelley JM, Richters J, et al. High prevalence of symptoms associated with ovarian cancer among Australian women. *Aust N Z J Obstet Gynaecol*;51(1):71-8.
7. Agrawal A, Whorwell PJ. Review article: abdominal bloating and distension in functional gastrointestinal disorders--epidemiology and exploration of possible mechanisms. *Aliment Pharmacol Ther* 2008;27(1):2-10.
8. Sharma A, Menon,U., Jacobs, IJ. Symptoms in the diagnosis of Ovarian Cancer In: Hillard T, editor. *The Yearbook of Obstetrics and Gynaecology*, 2007:159-70
9. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* 2009;339:b2998.
10. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;13(4):353-65.
11. Gilbert L, Basso O, Sampalis J, Karp I, Martins C, Feng J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *Lancet Oncol* 2012;13(3):285-91.
12. Andersen MR, Lowe KA, Goff BA. Value of symptom-triggered diagnostic evaluation for ovarian cancer. *Obstet Gynecol* 2014;123(1):73-9.
13. NICE. Ovarian cancer: The recognition and initial management of ovarian cancer. In: CG122) NIoHaCEg, editor, 2011.
14. Vernooij F, Heintz AP, Witteveen PO, van der Heiden-van der Loo M, Coebergh JW, van der Graaf Y. Specialized care and survival of ovarian cancer patients in The Netherlands: nationwide cohort study. *J Natl Cancer Inst* 2008;100(6):399-406.
15. Vernooij F, Heintz P, Witteveen E, van der Graaf Y. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. *Gynecol Oncol* 2007;105(3):801-12.
16. Moss EL, Moran A, Reynolds TM, Stokes-Lampard H. Views of general practitioners on the role of CA125 in primary care to diagnose ovarian cancer. *BMC Womens Health* 2013;13(1):8.
17. Rai N, Nevin J, Downey G, Abedin P, Balogun M, Kehoe S, et al. Outcomes following implementation of symptom triggered diagnostic testing for ovarian cancer. *Eur J Obstet Gynecol Reprod Biol* 2015;187:64-9.
18. Rufford BD, Jacobs IJ, Menon U. Feasibility of screening for ovarian cancer using symptoms as selection criteria. *BJOG* 2007;114(1):59-64.
19. Lim AW, Mesher D, Gentry-Maharaj A, Balogun N, Jacobs I, Menon U, et al. Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst* 2012;104(2):114-24.
20. Wu L, Dai ZY, Qian YH, Shi Y, Liu FJ, Yang C. Diagnostic value of serum human epididymis protein 4 (HE4) in ovarian carcinoma: a systematic review and meta-analysis. *Int J Gynecol Cancer* 2012;22(7):1106-12.
21. 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 in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2000;16(5):500-5.

22. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010;341:c6839.
23. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. *J Clin Oncol* 2005;23(34):8794-801.
24. Kaijser J, Sayasneh A, Van Hoorde K, Ghaem-Maghami S, Bourne T, Timmerman D, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. *Hum Reprod Update* 2014;20(3):449-62.
25. RCOG/BSGE. Green top guideline No: 62, Management of suspected ovarian masses in premenopausal women. London: RCOG, November 2011.
26. 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.
27. Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, et al. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2010;36(2):226-34.
28. Sayasneh A, Kaijser J, Preisler J, Johnson S, Stalder C, Husicka R, et al. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. *Gynecol Oncol* 2013.
29. RCOG GN. Ovarian cysts in postmenopausal women,

October 2003

Reviewed 2010.

BMJ Open

Refining Ovarian Cancer Test accuracy Scores (ROCkeTS) - protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-010333.R1
Article Type:	Protocol
Date Submitted by the Author:	12-Apr-2016
Complete List of Authors:	Sundar, Sudha; University of Birmingham, Cancer Sciences; City Hospital, Pan BirminghamGynaecological Cancer Centre Rick, Caroline; University of Birmingham, BCTU Dowling, Francis; University of Birmingham, BCTU Au, Pui; University of Birmingham, BCTU Snell, Kym; University of Birmingham, Public Health, Epidemiology & Biostatistics Rai, Nirmala; University of Birmingham, Cancer Sciences Champaneria, Rita; University of Birmingham, BCTU Stobart, Hilary; Patient and Public Involvement, Neal, Richard; Bangor University, North Wales Centre for Primary Care Research Davenport, Clare; University of Birmingham, Epidemiology ,Public Health and Biostatistics Mallett, Susan; University of Birmingham, School of Health & Population Sciences Sutton, Andrew; University of Birmingham, Health Economics Kehoe, Sean; University of Birmingham, of Cancer Studies Timmerman, Dirk; KU Leuven, Department of Obstetrics and Gynaecology Bourne, Tom; Imperial College London, Queen Charlottes and Chelsea Hospital; KU Leuven, Department of Obstetrics and Gynaecology Van Calster, Ben; KU Leuven, Department of Obstetrics and Gynaecology; KU Leuven, Gentry-Maharaj, Aleks; Universlty College London, Women's Cancer Menon, Usha; University College London, Women's Cancer Deeks, Jon; University of Birmingham, Public Health, Epidemiology and Biostatistics; University of Birmingham, BCTU
Primary Subject Heading:	Obstetrics and gynaecology
Secondary Subject Heading:	Diagnostics, Evidence based practice, Oncology, Radiology and imaging
Keywords:	Gynaecological oncology < GYNAECOLOGY, Ultrasound < RADIOLOGY & IMAGING, GYNAECOLOGY

Refining Ovarian Cancer Test accuracy Scores (ROCkeTS) - protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer

Sudha Sundar,[1,2] Caroline Rick,[3] Francis Dowling,[3] Pui Au,[3] Kym Snell,[4] Nirmala Rai ,[1,4] Rita Champaneria,[4] Hilary Stobart,[5] Richard Neal,[6] Clare Davenport,[4] Susan Mallett,[4] Andrew Sutton,[7] Sean Kehoe,[1] Dirk Timmerman[8], Tom Bourne.[8,9] Ben Van Calster.[8] Aleksandra Gentry-Maharaj,[10] Usha Menon,[10] Jon Deeks.[3,4] on behalf of the ROCkeTS study group

1. Institute of Cancer and Genomic Sciences , University of Birmingham, UK
2. Pan Birmingham Gynaecological Cancer Centre, City Hospital, Birmingham, UK
3. Birmingham Clinical Trials Unit, University of Birmingham, UK
4. Diagnostic Test Accuracy Group, University of Birmingham, UK
5. Patient Representative
6. Primary Care, Bangor University, North Wales, UK
7. Health Economics, University of Birmingham, UK
8. Department of Obstetrics and Gynaecology, KU Leuven, Belgium
9. Queen Charlotte's and Chelsea Hospital, Imperial College, London, UK

Gynaecological Cancer Research Centre, Department of Women's Cancer, Institute for Women's Health, UCL, UK

Author for correspondence

Sudha Sundar

Senior Lecturer in Gynaecological Oncology

Institute of Cancer and Genomic Sciences and Pan Birmingham Gynaecological Cancer Centre, City Hospital and

University of Birmingham,

Edgbaston,

Birmingham.

B15 2TT

s.s.sundar@bham.ac.uk

Abstract

Introduction: Ovarian cancer (OC) is associated with nonspecific symptoms such as bloating, making accurate diagnosis challenging: only 1 in 3 women with OC present through primary care referral. NICE guidance recommends sequential testing with CA125 and routine ultrasound in primary care. However these diagnostic tests have limited sensitivity or specificity. Improving accurate triage in women with vague symptoms is likely to improve mortality by streamlining referral and care pathways. The Refining Ovarian Cancer Test Accuracy Scores (ROCkeTS; HTA 13/13/01) project will derive and validate new tests/risk prediction models that estimate the probability of having OC in women with symptoms. This protocol refers to the prospective study only (phase 3).

Methods and Analysis:– ROCkeTS comprises 4 parallel Phases. The full ROCkeTS protocol can be found at http://www.birmingham.ac.uk/ROCKETS. Phase 3 – prospective test accuracy study. The study will recruit 2,450 patients from 15 UK sites. Recruited patients complete symptom and anxiety questionnaires, donate a serum sample and undergo ultrasound scored per IOTA criteria. Recruitment is at rapid access clinics, emergency departments and elective clinics. Models to be evaluated include ultrasound models derived by the International Ovarian Tumour analysis (IOTA) group and novel models derived from analysis of existing datasets.

Estimates of sensitivity, specificity, c-statistic (area under Receiver operating curve curve), Positive Predictive Value and Negative Predictive value of diagnostic tests evaluated and a calibration plot for models will be evaluated.

ROCkeTS has received ethical approval from the NHS West Midlands REC (14/WM/1241) and is registered on the Controlled trials website (ISRCTN17160843) and the NIHR Cancer and Reproductive Health portfolios.

Strengths

- ROCkeTS conforms to PROBE design for biomarker evaluation and STARD criteria for test accuracy studies
- Stringent efforts to collect outcome data to prevent ascertainment bias
- Inbuilt ultrasound quality control.

Limitations

- Recruitment is at secondary care so the population will be less heterogenous than that seen in primary care.

Introduction

Ovarian cancer (OC) is the seventh most common cancer in women worldwide, with 239,000 new cases diagnosed in 2012.[1] In the UK, OC has an annual incidence of 7,116 women and causes 4,271 deaths; the lifetime risk of developing OC is 1 in 54.[2] 70% of patients will present at an advanced stage (stages III/IV). The International Cancer Benchmarking Project showed that the reason that OC survival in the UK is significantly lower than other western countries seems to be related to a lower proportion of patients receiving treatment and surviving the first year after cancer diagnosis and is likely due to a delay in diagnosis.[3] 5-year survival rates are 43% overall but over 90% for early stage tumours.[4] High grade serous is the most common histotype (80%). Worryingly, long term survival from OC has remained static over the last decades at 30%.[2]

Previously considered a 'silent killer', it is now recognised that patients with OC suffer from a number of nonspecific symptoms. These include abdominal bloating, distension, feeling full quickly and/or loss of appetite, pelvic/abdominal pain, increased urinary urgency and/or frequency, unexplained weight loss, fatigue or changes in bowel habit. These symptoms are very common.[5 6] Interrogation of United Kingdom General Practice databases suggest that on average 1 in 2 women between the ages of 45 and 74 present once a year to their General practitioner (GP)/primary care doctor with these symptoms. Abdominal bloating alone,[5 6] is documented in 16-30% of women presenting to GPs.[7] Diagnostic challenges are considerable given (1) the low incidence of OC (a GP sees a woman with OC once in 3-5 years) and (2) the low positive predictive value (PPV) of symptoms (only 1 in 400-600 symptomatic women have OC).[8 9] Unfortunately, these diagnostic challenges result in nearly 36% of women, subsequently diagnosed with OC, presenting to the GP with symptoms three or more times prior to diagnosis.[10] Two large prospective studies of symptom triggered testing for Ovarian Cancer suggest that symptom triggered testing using CA125 is likely to result in referral of a higher proportion of patients with resectable disease. [11 12]

The UK introduced symptom-triggered testing for Ovarian Cancer in women with vague symptoms in 2011. National Institute for Health and Care Excellence (NICE) guidelines recommend sequential testing using serum CA125 followed by pelvic ultrasound (USS) in women (particularly aged ≥50) presenting to primary care with symptoms –such as persistent abdominal distension/'bloating', feeling full and/or loss of appetite, pelvic/abdominal pain,

increased urinary urgency and/or frequency, unexplained weight loss, fatigue or changes in bowel habit on a persistent or frequent basis.[13] However, the NICE guidance does not specify the type of ultrasound abnormalities that should prompt referral. Current tests have limited sensitivity with CA125 being elevated only in 40- 50% of women with stage 1 OC in both screening and presurgical studies [14,15] Once referred to secondary care, women with complex masses considered benign can undergo laparoscopic or conservative management, whereas women with malignancy who undergo surgery by trained gynaecological oncologists have the best outcome.[16,17] Use of NICE guidance in practice is extremely variable. A survey of 258 General Practitioners (GP's) reported that the majority would refer patients on the basis of raised CA125 even if the USS was normal.[18] A recent audit revealed that the majority of referrals (90%) for suspected OC did not follow guidance. Referrals were heterogeneous with regard to which symptoms prompted these, what GPs considered to be abnormal CA125 level or abnormal USS.[19] Two thirds of women referred were premenopausal.[19]

Therefore, improved diagnostic tools in primary care and better pre-surgical triage in secondary care is likely to improve mortality in OC patients. Optimal diagnostic pathways for premenopausal women with suspected OC/complex ovarian mass and raised CA125 also need to be defined. Ovarian cysts in premenopausal women are extremely common, however it is important to recognise that about 1,000 women under 50 will be diagnosed with OC in the UK annually.[2]

The National Institute of Health Research (NIHR) Health Technology Assessment (HTA) programme commissioned **ROCkeTS – 'Refining ovarian cancer test accuracy scores'** 13/13/01 (1st October 2014 and 30th September 2018) which aims to identify, refine and validate tests and clinical risk scores (risk prediction models) that estimate the probability of having OC in post- and premenopausal women with symptoms, in primary and secondary care. The project comprises 4 phases/work packages, phase 1- systematic reviews of existing risk prediction algorithms and biomarkers to detect ovarian cancer, phase 2 - Interrogation of datasets from 2 large OC trials; the screening trial UKCTOCS and the pre-operative detection International Ovarian tumour analysis (IOTA) studies to derive or refine risk scores with biomarkers from the systematic review, phase 3 – a prospective study to validate new scores in women newly referred to secondary care with symptoms of OC, phase 4 – a model of the

diagnostic pathway across primary and secondary care. A cost consequence analysis of testing pathways will be delivered as part of ROCkeTS.

The full ROCkeTS protocol can be found at http://www.birmingham.ac.uk/ROCKETS

The ROCkeTS project evaluates risk prediction models derived from conventional tests – symptoms, ultrasound variables and commercially available serum blood tests.

Phase 3 of ROCKETS is a cross sectional study aimed at establishing the accuracy of tests and prediction models for the diagnosis of prevalent ovarian cancer. However of necessity the reference standard for a correct diagnosis is follow up to 12 months, as the study is not restricted to women scheduled for immediate surgery. This is a diagnostic trial (not a screening trial) as women are symptomatic and referred to secondary care.

The Phase 3 prospective study will recruit 2,450 premenopausal and postmenopausal symptomatic women over 23 months (recruitment started in June 2015). Patients will be recruited through rapid access clinics, emergency departments or through routine clinic referrals, reflecting the heterogeneity of diagnostic routes for OC in the UK.

Recruited patients consent to complete a validated symptom questionnaire, donate a serum sample and undergo an ultrasound scan scored as per the IOTA criteria. Standard care pathways are followed beyond this point. Patients are triaged to receive surgery or conservative management based on standard of care model Risk of malignancy index (RMI) scores for postmenopausal women, using a threshold of 250 in postmenopausal women or as per local unit practice in premenopausal women. This threshold was selected by the funder as comparator as part of the commissioning brief as this is current standard of care in the UK. [13] Participants and treating doctors will be blinded to the new biochemical serum test results which will be performed in batches as these are not part of standard clinical care. Outcome data will be collected from histology reports in women undergoing surgery or biopsy and from follow-up over 12 months of those managed conservatively. Symptom scores, serum biomarker tests and ultrasound data will be analysed at the end of follow-up to validate risk prediction models derived in phases 1 and 2.

ROCKETS STUDY DESIGN

Aim of the study

- To derive and validate risk prediction models that estimate the probability of having OC for women with symptoms suggestive of OC for postmenopausal and premenopausal women.
- To identify optimal risk thresholds of best models for the diagnosis of OC that can guide patient management.

Design

The ROCkeTS study is a single arm prospective cohort diagnostic test accuracy study to evaluate existing and novel risk prediction models for pre and post-menopausal women with symptoms.

A test accuracy study is different to an effectiveness study in that randomisation of subjects is not involved. It is designed to generate a comparison of measurements obtained by tests under investigation (index tests) with those obtained from current standard of care tests (comparator test) against a reference standard. In this way the accuracy of any new 'index' tests can be estimated. A reference standard is a test that confirms or refutes the presence or absence of disease beyond reasonable doubt. Therefore it is sometimes also known as the "gold standard". In this study, as per standards of diagnostic test evaluation, estimates of sensitivity, specificity, c-statistic (area under Receiver operating curve (ROC) curve), PPV and Negative Predictive value (NPV) of tests evaluated and a calibration plot for models evaluated will be presented.

Here, the reference standard will be histology of tissues taken from patients who proceed to surgery or biopsy or follow-up using structured templates at a minimum of 12 months after presentation. The accuracy of the index test will be compared against that of the comparator test, the existing standard risk prediction score Risk of malignancy index (RMI). This index, is an algorithm combining menopausal status, CA125 and ultrasound features and is standard of care in the UK. [20] In ROCkeTS, the index test (novel risk prediction models) will be derived in phases 1 and 2 and validated in phase 3. We will identify biochemical markers, symptom indices and USS as likely components of a novel risk prediction model, as these may be implemented across primary and secondary care. Therefore we will collect symptom

questionnaires, blood and USS data in the study to be analysed and validated at the end of the study.

Possible components of the new risk prediction model/s

Symptoms

Case-control studies demonstrate that symptom questionnaires have good diagnostic accuracy; however, they need to be refined for use by patients in primary care[21 22] as the duration of symptoms preceding diagnosis is uncertain.[22] Symptom questionnaires may help triage patients prior to referral and would help standardise symptoms for any prediction model. This is particularly important, given the subjective nature of eliciting symptoms through unstructured clinical history taking and the existing audit evidence that they are interpreted variably by GPs who will only see few cases of OC in their practice. A robust symptom score that can triage referral based on a questionnaire may be very useful.

Biochemical markers

A number of serum biomarker tests (multiplex testing - OVA1) and multiple-marker based algorithms (Risk of Ovarian malignancy algorithm, ROMA,) have been identified in the last decade. Abnormal Human Epididymis 4 (HE4) biomarker levels may improve risk stratification for OC. A recent systematic review reports that HE4 may improve the diagnostic performance of CA125, however studies showed considerable heterogeneity.[23]

Ultrasound based models –IOTA risk prediction models

After publication of an agreement on terms, definitions and measurement methods to describe adnexal masses,[24] the IOTA collaboration set out multicentre studies on large cohorts of patients presenting with an adnexal mass. The IOTA database has enabled both previously developed prediction models to be tested and novel prediction models and rules to characterise ovarian pathology prior to surgery to be developed and validated. [25 26] In a systematic review with meta-analysis, IOTA algorithms such as the simple rules were identified to be the best pre-surgical diagnostic tools to characterise adnexal masses, with improved performance over the RMI.[27] Although the Royal college of Obstetricians and Gynaecologists (RCOG) included the simple rules in their guidance for evaluating ovarian

pathology in premenopausal women, IOTA models are not commonly used in NHS clinical practice. [28] Recently the ADNEX (Assessment of Different NEoplasias in the adneXa) model was published. As a multiclass prediction model, it differs from all other models by differentiating between malignant and benign masses, and also discriminating between 4 types of malignant tumours (borderline ovarian tumours, stage I ovarian cancer, stage II-IV ovarian cancer and metastatic tumours of other primary origin). ADNEX still needs an extensive external validation, but is considered to be promising. [29]

Target Population

1. Post and premenopausal women with symptoms of suspected OC. Symptoms are as defined by NICE including but not restricted to persistent or frequent abdominal distension, feeling full (early satiety) and/or loss of appetite, pelvic or abdominal pain, increased urinary urgency and/or frequency, postmenopausal bleeding. Symptoms listed here are not an exhaustive list and this will be updated from any symptoms identified through Phase 1 systematic reviews of literature.
2. Patients referred with symptoms from GP as suspected OC.

Comparator

Risk of Malignancy Index cut-off of 250.[30]

Source of potential participants

Patients referred to secondary care as Outpatients; either as urgent 2 week or routine referrals, USS clinics, Inpatient and emergency presentation.

Inclusion and Exclusion Criteria

Multicentre trial across the UK who meet the eligibility criteria described below.

Inclusion criteria:

- Newly presenting Pre and postmenopausal women with symptoms of suspected OC and either raised CA125 or abnormal USS or both. Menopause is defined as >12 months of amenorrhoea.
- Aged between 16 and 90 years.
- Patients able to provide informed consent.

NB – Patients with >120 days delay between initial registration IOTA scan and surgery will need a repeat IOTA scan prior to surgery

Exclusion criteria:

- Ultrasound (USS) reveals non-ovarian pathology e.g. fibroids
- Patients who decline transvaginal scan.
- Previous ovarian malignancy
- Pregnant patients
- Patients with previous bilateral oophorectomy
- Active non ovarian malignancy – Women with a past history of cancer are only eligible if there are no documented persistent or recurrent disease and have not received treatment for this in the last 12 months. This exclusion does not apply to patients with premalignant disease eg cervical intra-epithelial neoplasia or patients receiving Tamoxifen/other drugs to prevent breast cancer recurrence

See figure 1 for participant flow through the trial.

Figure 1: Flow chart

STUDY PROCEDURES AND TESTS

The Index tests

There are three index tests that will be performed in the prospective single arm test accuracy study.

1. Participants entering the study will complete a symptom elicitation questionnaire and anxiety questionnaire (STA6) and Impact of event score.

2. Ultrasonographers will record the USS variables and score the USS using IOTA simple rules and ADNEX models.[29 31 32] For most women in the trial, this will only mean some additional data being collected during their USS appointment. For a small number of women, this may mean an additional USS scan after consent.

3. Participants will have an additional blood sample taken at baseline for biomarker assessment at the end of the study. Serum will be banked for testing at a later point. Treating doctors will be blinded to any new biomarker assessments done as part of this trial. Details of blood sample collection will be provided in a lab manual.

Apart from the study tests, all other aspects of participant management are entirely at the discretion of the local doctors and as per the RCOG guidelines for management of these participants.[28 30] Treating clinicians will be asked to record their treatment recommendations as per standard care and after any additional USS information in order to assess the impact of this test on care pathways.

Quality assurance of index tests

The performance of USS is subjective and operator dependent. Therefore sonographers/doctors carrying out USS will undergo a face-to-face IOTA training course provided as part of their participation in the ROCkeTS trial. Sites will commit to undergoing quality assurance during the ROCkeTS trial. Online USS training materials will be developed during this project for future use by the UK NHS. Quality assurance of testing will begin with a clearly documented staff training programme. A register of staff that have been trained, and had their competence assessed will be maintained, and only staff whose names appear on this list will be permitted to undertake scans within the ROCkeTS prospective trial. Staff will also

receive a site visit and assessment of their competence. Competence will be assessed by those authorised by the IOTA team.

Reference standard/Follow-up Schedule

Reference standard for the study will be histology of tissue taken at surgery or biopsy in women who are managed surgically following study enrolment. Histology data collected will include information on whether the tumour was considered invasive cancer or borderline. Outcome of participants referred for suspected OC that do not undergo surgery will be assessed by a follow-up visit at 12 months or by a telephone call or a questionnaire from the research nurse at 12 months, as per the local investigators' discretion and clinical assessment. Wellbeing will be ascertained at this follow-up. A structured template will be used. Women will also be asked to give permission for data to be linked with crosschecked against national cancer registry data so that we are able to identify the false negative rate for any diagnostic tests.

Sample Size

The diagnostic accuracy from models will be compared to the current tests recommended by NICE for primary and secondary care. The sample size has a 90% power to detect with 5% significance difference in accuracy between the existing test (RMI threshold 250) and the new model. Due to the expected difference in performance in pre and postmenopausal women, separate sample sizes have been calculated. Sample sizes are calculated assuming independence of test errors and interim analyses will confirm parameter assumptions.

For postmenopausal women: Performance of RMI at threshold 250 is assumed to be 70% sensitive and 90% specific.[30] 1,400 participants will be required to detect an increase in sensitivity of 10% (to 80%) and in specificity of 5% (to 95%). Based on a prevalence of 30% of OC in referred women (local audit figures), with sensitivity and specificity of RMI assumed to be 70% and 90%, a sample size of 1,333 provides 90% power to detect an increase of sensitivity to 80% and specificity to 95% in paired data (conservatively assuming independence of test errors). Allowing for a loss to follow-up of up to 5%, this gives a final sample size of 1,400.

For premenopausal women: Performance of RMI is assumed to have a sensitivity of 72% and a specificity of 46% (local audit figures). The trial is powered to detect an increase in sensitivity of 10% (to 82%) and in specificity of 10% (to 56%). Prevalence of OC in

premenopausal women referred to secondary care is around 10%.[19] A sample size of 1,000 will provide 100 OC events in which to build new models combining symptom and test data (adequate events to model 10 predictor variables), and will provide 90% power to detect an increase in specificity of 8% (from 46% for RMI to 54%). With a predicted loss to follow-up of up to 5%, the final sample size required is 1,050 women.

Study Duration

We anticipate recruitment of 2,450 participants within 23 months, with a minimum of 12 months follow-up from the last participant entering the study.

Data collection

All information will be collected on standard proformas (Case report forms; CRF) and identified by study number. Information will be collated on paper forms and then either copied and sent to the coordinating centre for input or entered directly into the study database via a web interface. A dataset including age, ethnicity, parity, GP details and significant medical/surgical history will be collected. We aim to use the NHS number as the primary identifier when linking to national registries and to track individuals throughout the NHS. Additional data on outcomes such as cancer or non-cancereous conditions will be collected at follow-up.

Data will be collected on relevant medical, obstetric and gynaecological, surgical history, emotional impact as well as information on the symptoms that prompted GP referral or investigation. USS information will be collected. Data on the reference diagnosis will be obtained from the histopathology report and a structured template to assess wellbeing for participants who do not undergo surgery will be collected directly from the participants. Importantly, outcomes collected will include all conditions/diagnoses in women with nonspecific symptoms.

Outcome measures and costs

Validation of a risk prediction model and tests for estimating the probability of OC in women with suspected OC; the key outcome measures are the accuracy of the tests and models in terms of their discrimination ability (e.g. sensitivity, specificity) and calibration (observed

versus predicted probabilities), and the identification of thresholds to guide patient management decisions.

Trial data collection will be undertaken prospectively for all participants in order to inform the costs for each pathway.

Analysis plan - Test accuracy

We will report estimates of sensitivity, specificity, c-statistic (area under Receiver operating curve (ROC) curve), PPV and Negative Predictive value (NPV) of tests evaluated and a calibration plot for models evaluated. Also, in terms of our new model derived from phase 2 of the wider ROCkeTS project, its improvement over existing models will be summarised by comparing the c-statistic and the calibration. We will also summarise the net-reclassification index for each new predictor that existing models omitted.

The risk prediction models derived in phases 1 and 2 of the ROCkeTS project will each produce a predicted risk of OC by 12 months for all the individuals in our study. Therefore, we will compare the observed outcome at 12 months with this predicted risk. The calibration (in terms of calibration slope) and discrimination (e.g. c-statistic) will be evaluated for the models derived and identified in phases 1 and 2, and their performance compared to the existing RMI model. The calibration will be shown visually by grouping women into deciles ordered by predicted risk and considering the agreement between the mean predicted risk and the observed events in each decile. The aim is to use predefined models on the phase 3 data, so the bulk of phase 3 analysis will be external validation of predefined models. Where the dataset is used to derive a new model, optimism will be reduced using shrinkage methods through internal validation.

Generalisability of results to primary care.

We appreciate that there may be recalibration required for any models validated within ROCkeTS in the primary care setting. However, we stress that to conduct a study that recruited women in primary care to validate diagnostic models would need to be extremely large and prohibitively expensive. Furthermore, our work and others have shown that GP's

are likely to refer on the basis of a raised Ca125 or abnormal scan rather than follow the NICE suggested referral pathway which is based on sequential Ca125 and USS and referral only if both are abnormal. [18 19] Thus we believe that the population of patients referred through rapid access clinics with symptoms will be more heterogenous than anticipated from the NICE guidance and therefore maybe more applicable to a primary care population.

Analysis plan - Cost consequence analysis

Resource use for each of the diagnostic tests will be broken down and displayed along with their unit costs alongside the outcomes for each pathway. The resource usage will include the types of tests administered, the number of inpatient and outpatient consultations, and any operative procedures undertaken. This approach will help to show which are the major cost drivers for each of the diagnostic pathways and will be collected as part of the clinical CRF.

Study Conduct

The conduct of the study will be in accordance with the Research Governance Framework for Health and Social Care and/or the Research Governance Framework for Health and Community Care. The participant's written informed consent to participate in the trial will be obtained before any trial procedures or questionnaires are completed. The women's GP will be notified of her participation in the study with her consent (GP letter).

Participants will be free to withdraw from the study at any time without any effect on their standard of care, data and samples provided up to the point a participant withdraws will be retained unless the participant expressly requests their removal. This is because analysis will be based on all recruited participants and per protocol.

Ethics and dissemination: This study has received ethics permission from the NHS West Midlands REC (Ref14/WM/1241) and has no specific safety considerations. Outputs will be disseminated through open access publications, on the website www.birmingham.ac.uk/ROCKETS

Funding Source

This study is funded by a grant from National Institute of Heath Research, Health Technology assessment HTA 13/13/01

Disclaimer

The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HTA, NIHR, NHS or the Department of Health.

Authors' contributions

All authors contributed to writing of the protocol.

Competing Interests statement

All authors reports grants from NIHR HTA during the conduct of the study. UM reports other from Abcodia, outside the submitted work. TB reports other from Samsung Electronics, other from Roche Diagnostics, outside the submitted work.

Acknowledgements

We thank the ROCkeTS Project Oversight committee for their kind input and guidance. Chair – Prof Peter Sasieni, Members – Mr Nigel Acheson, Mr Andy Nordin, Prof Peter Murchie, Dr Michael Weston, Simon Newman and Annwen Jones (Target Ovarian Cancer)

ROCkeTS Study Group

Site	Principle Investigator	Imaging Lead
Belfast City Hospital	Dr Hans Nagar	Dr Hans Nagar
Birmingham Women's Hospital	Dr Parveen Abedin	Dr Moji Balogun
City Hospital Birmingham	Dr Sudha Sundar	Dr Humaira Khan
East Surrey Hospital	Dr Karen Jermy	Dr Karen Jermy
Manor Hospital	Mr Fateh Ghazal	Dr Harinder Rai
Norfolk & Norwich University Hospital	Dr Tim Duncan	Dr Victoria Ames
Princess Anne Hospital	Dr Susanne Johnson	Dr Susanne Johnson
Queen Elizabeth Hospital, Gateshead	Dr Tracey Hughes	Dr Tracey Hughes
Royal Blackburn Hospital	Mr Mark Willett	Dr Eswarappa Nagaraju
Royal Hallamshire Hospital	Dr Shahram Abdi	Dr Shahram Abdi
Royal Preston/Sharoe Green Hospitals	Mr Patrick Keating	Dr Tracy Butcher
Royal Sussex County Hospital	Dr Sonali Kaushik	Dr Lavanya Vitta
The Royal Victoria Infirmary	Dr Mark Roberts	Yasmin McGrady
University Hospital Of North Durham	Dr Partha Sengupta	Mr Robert Kent
University Hospital Of Wales	Dr Aarti Sharma	Dr Anju Sinha

1. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F., GLOBOCAN 2012 v1.0 ClaMWICNI, Lyon FIAfRoCAfhgif.

2. CRUK. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/ovary/incidence/uk-ovarian-cancer-incidence-statistics, 2015.
3. Maringe C, Walters S, Butler J, Coleman MP, Hacker N, Hanna L, et al. Stage at diagnosis and ovarian cancer survival: evidence from the International Cancer Benchmarking Partnership. *Gynecol Oncol* 2012;127(1):75-82.
4. Cancer Research UK. Ovarian cancer statistics, 2014.
5. Wheatley G. Ovarian cancer. Usefulness of abdominal symptoms in early diagnosis. *Bmj* 2009;339:b3954.
6. Pitts MK, Heywood W, Ryall R, Smith AM, Shelley JM, Richters J, et al. High prevalence of symptoms associated with ovarian cancer among Australian women. *Aust N Z J Obstet Gynaecol*;51(1):71-8.
7. Agrawal A, Whorwell PJ. Review article: abdominal bloating and distension in functional gastrointestinal disorders--epidemiology and exploration of possible mechanisms. *Aliment Pharmacol Ther* 2008;27(1):2-10.
8. Sharma A, Menon,U., Jacobs, IJ. Symptoms in the diagnosis of Ovarian Cancer In: Hillard T, editor. *The Yearbook of Obstetrics and Gynaecology*, 2007:159-70
9. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* 2009;339:b2998.
10. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;13(4):353-65.
11. Gilbert L, Basso O, Sampalis J, Karp I, Martins C, Feng J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *Lancet Oncol* 2012;13(3):285-91.
12. Andersen MR, Lowe KA, Goff BA. Value of symptom-triggered diagnostic evaluation for ovarian cancer. *Obstet Gynecol* 2014;123(1):73-9.
13. NICE. Ovarian cancer: The recognition and initial management of ovarian cancer, 2011.
14. Jacobs I, Bast RC, Jr. The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 1989;4(1):1-12.
15. Jacobs I, Davies AP, Bridges J, Stabile I, Fay T, Lower A, et al. Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. *BMJ* 1993;306(6884):1030-4.
16. Vernooij F, Heintz AP, Witteveen PO, van der Heiden-van der Loo M, Coebergh JW, van der Graaf Y. Specialized care and survival of ovarian cancer patients in The Netherlands: nationwide cohort study. *J Natl Cancer Inst* 2008;100(6):399-406.
17. Vernooij F, Heintz P, Witteveen E, van der Graaf Y. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. *Gynecol Oncol* 2007;105(3):801-12.
18. Moss EL, Moran A, Reynolds TM, Stokes-Lampard H. Views of general practitioners on the role of CA125 in primary care to diagnose ovarian cancer. *BMC Womens Health* 2013;13(1):8.
19. Rai N, Nevin J, Downey G, Abedin P, Balogun M, Kehoe S, et al. Outcomes following implementation of symptom triggered diagnostic testing for ovarian cancer. *Eur J Obstet Gynecol Reprod Biol* 2015;187:64-9.
20. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol* 1990;97(10):922-9.
21. Rufford BD, Jacobs IJ, Menon U. Feasibility of screening for ovarian cancer using symptoms as selection criteria. *BJOG* 2007;114(1):59-64.
22. Lim AW, Mesher D, Gentry-Maharaj A, Balogun N, Jacobs I, Menon U, et al. Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst* 2012;104(2):114-24.

23. Wu L, Dai ZY, Qian YH, Shi Y, Liu FJ, Yang C. Diagnostic value of serum human epididymis protein 4 (HE4) in ovarian carcinoma: a systematic review and meta-analysis. *Int J Gynecol Cancer* 2012;22(7):1106-12.
24. 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 in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2000;16(5):500-5.
25. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. *J Clin Oncol* 2005;23(34):8794-801.
26. Sayasneh A, Kaijser J, Preisler J, Johnson S, Stalder C, Husicka R, et al. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. *Gynecol Oncol* 2013.
27. Kaijser J, Sayasneh A, Van Hoorde K, Ghaem-Maghami S, Bourne T, Timmerman D, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. *Hum Reprod Update* 2014;20(3):449-62.
28. RCOG/BSGE. Green top guideline No: 62, Management of suspected ovarian masses in premenopausal women. London: RCOG, November 2011.
29. 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.
30. RCOG GN. Ovarian cysts in postmenopausal women,

October 2003

Reviewed 2010.
31. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010;341:c6839.
32. Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, et al. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2010;36(2):226-34.

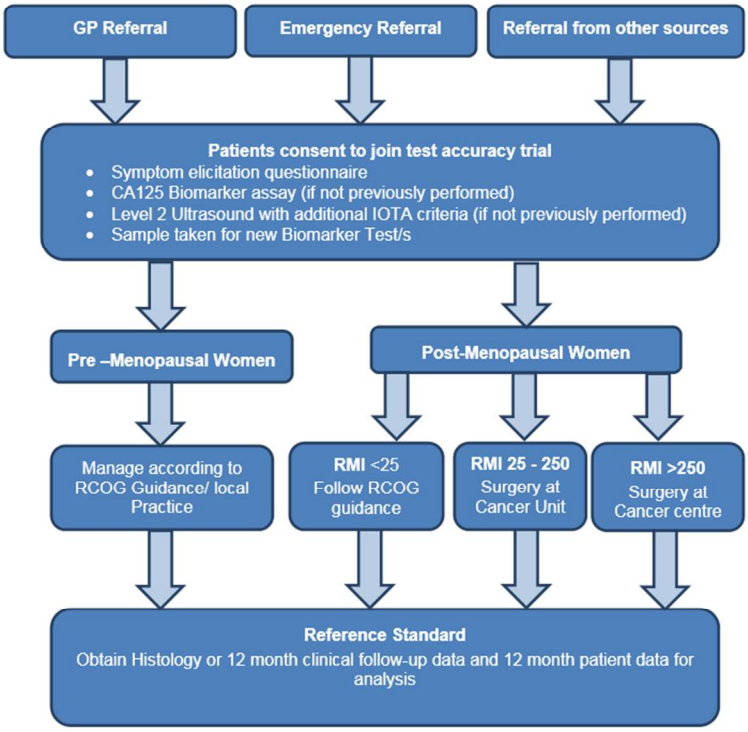


Figure 1

67x94mm (300 x 300 DPI)