Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)

Usha Menon, Aleksandra Gentry-Maharaj, Rachel Hallett, Andy Ryan, Matthew Burnell, Aarti Sharma, Sara Lewis, Susan Davies, Susan Philpott, Alberto Lopes, Keith Godfrey, David Oram, Jonathan Herod, Karin Williamson, Mourad W Seif, Ian Scott, Tim Mould, Robert Woolas, John Murdoch, Stephen Dobbs, Nazar N Amso, Simon Leeson, Derek Cruickshank, Alistair Mcguire, Stuart Campbell, Lesley Fallowfield, Naveena Singh, Anne Dawnay, Steven J Skates, Mahesh Parmar, Ian Jacobs

Summary

Background Ovarian cancer has a high case–fatality ratio, with most women not diagnosed until the disease is in its advanced stages. The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is a randomised controlled trial designed to assess the effect of screening on mortality. This report summarises the outcome of the prevalence (initial) screen in UKCTOCS.

Methods Between 2001 and 2005, a total of 202 638 post-menopausal women aged 50–74 years were randomly assigned to no treatment (control; n=101 359); annual CA125 screening (interpreted using a risk of ovarian cancer algorithm) with transvaginal ultrasound scan as a second-line test (multimodal screening [MMS]; n=50 640); or annual screening with transvaginal ultrasound (USS; n=50 639) alone in a 2:1:1 ratio using a computer-generated random number algorithm. All women provided a blood sample at recruitment. Women randomised to the MMS group had their blood tested for CA125 and those randomised to the USS group were sent an appointment to attend for a transvaginal scan. Women with abnormal screens had repeat tests. Women with persistent abnormality on repeat screens underwent clinical evaluation and, where appropriate, surgery. This trial is registered as ISRCTN22488978 and with ClinicalTrials.gov, number NCT00058032.

Findings In the prevalence screen, 50 078 (98·9%) women underwent MMS, and 48 230 (95·2%) underwent USS. The main reasons for withdrawal were death (two MMS, 28 USS), non-ovarian cancer or other disease (none MMS, 66 USS), removal of ovaries (five MMS, 29 USS), relocation (none MMS, 39 USS), failure to attend three appointments for the screen (72 MMS, 757 USS), and participant changing their mind (483 MMS, 1490 USS). Overall, 4355 of 50 078 (8.7%) women in the MMS group and 5779 of 48 230 (12·0%) women in the USS group required a repeat test, and 167 (0·3%) women in the MMS group and 1894 (3·9%) women in the USS group required clinical evaluation. 97 of 50 078 (0·2%) women from the MMS group and 845 of 48 230 (1·8%) from the USS group underwent surgery. 42 (MMS) and 45 (USS) primary ovarian and tubal cancers were detected, including 28 borderline tumours (eight MMS, 20 USS). 28 (16 MMS, 12 USS) of 58 (48·3%; 95% CI 35·0–61·8) of the invasive cancers were stage I/II, with no difference (p=0·396) in stage distribution between the groups. A further 13 (five MMS, eight USS) women developed primary ovarian cancer during the year after the screen. The sensitivity, specificity, and positive-predictive values for all primary ovarian and tubal cancers were 89·4%, 99·8%, and 43·3% for MMS, and 84·9%, 98·2%, and 5·3% for USS, respectively. For primary invasive epithelial ovarian and tubal cancers, the sensitivity, specificity, and positive-predictive values were 89·5%, 99·8%, and 35·1% for MMS, and 75·0%, 98·2%, and 2·8% for USS, respectively. There was a significant difference in specificity (p<0·0001) but not sensitivity between the two screening groups for both primary ovarian and tubal cancers as well as primary epithelial invasive ovarian and tubal cancers.

Interpretation The sensitivity of the MMS and USS screening strategies is encouraging. Specificity was higher in the MMS than in the USS group, resulting in lower rates of repeat testing and surgery. This in part reflects the high prevalence of benign adnexal abnormalities and the more frequent detection of borderline tumours in the USS group. The prevalence screen has established that the screening strategies are feasible. The results of ongoing screening are awaited so that the effect of screening on mortality can be determined.

Funding Medical Research Council, Cancer Research UK and the Department of Health, UK; with additional support from the Eve Appeal, Special Trustees of Bart's and the London, and Special Trustees of University College London Hospital.

Lancet Oncol 2009; 10: 327–40

Published Online March 11, 2009 DOI:10.1016/S1470-2045(09)70026-9

See Reflection and Reaction page 308

Gynaecological Oncology, University College London Elizabeth Garrett Anderson Institute for Women's Health, London, UK (U Menon MD, A Gentry-Maharaj PhD, R Hallett PhD, A Ryan PhD, M Burnell PhD, A Sharma MRCOG, S Lewis MSc, S Davies RGN, S Philpott MSc, Prof I Jacobs FRCOG); **Northern Gynaecological Oncology Centre, Queen Elizabeth Hospital, Gateshead, UK** (A Lopes FRCOG, K Godfrey FRCOG); **Department of Gynaecological Oncology, St Bartholomew's Hospital, London, UK** (D Oram FRCOG); **Department of Gynaecology, Liverpool Women's Hospital, Liverpool, UK** (J Herod MRCOG); **Department of Gynaecological Oncology, Nottingham City Hospital, Nottingham, UK** (K Williamson FRCOG); **Academic Unit of Obstetrics and Gynaecology, St Mary's Hospital, Manchester, UK** (M W Seif FRCOG); **Department of Gynaecological Oncology, Derby City Hospital, Derby, UK** (I Scott FRCOG); **Department of Gynaecological Oncology, Royal Free Hospital, London, UK** (T Mould FRCOG); **Department of Gynaecological Oncology, St Mary's Hospital, Portsmouth, UK** (R Woolas MD); **Department of Gynaecological Oncology, St Michael's Hospital, Bristol, UK** (J Murdoch FRCOG); **Department of Gynaecological Oncology, Belfast City Hospital,**

Belfast, UK (S Dobbs MRCOG); **Department of Obstetrics and Gynaecology, Wales College of Medicine, Cardiff University, Cardiff, UK** (N N Amso PhD); **Department of Gynaecological Oncology, Llandudno Hospital, North Wales, UK** (S Leeson FRCOG); **Department of Gynaecological Oncology, James Cook University Hospital, Middlesbrough, UK** (D Cruickshank FRCOG); **Department of Social Policy, London School of Economics, London, UK** (Prof A Mcguire PhD); **Create Health Clinic, London, UK** (Prof S Campbell DSc Lond); **Cancer Research UK Sussex Psychosocial Oncology Group at Brighton & Sussex Medical School, University of Sussex, Falmer, UK** (Prof L Fallowfield DPhil); **Barts and the London, London, UK** (N Singh FRCPath); **Clinical Biochemistry, University College London Hospitals, London, UK** (A Dawnay PhD); **Department of Medicine, Harvard Medical School, Boston, MA, USA** (S J Skates PhD); **and Cancer Group, Medical Research Council Clinical Trials Unit, London, UK** (Prof M Parmar PhD)

Correspondence to:
Dr Usha Menon, University College London Elizabeth Garrett Anderson Institute for Women's Health , Maple House, 149 Tottenham Court Road, London, W1T 7DN, UK
u.menon@ucl.ac.uk

Introduction

Ovarian cancer is a disease with a poor prognosis. Although advances in therapy have improved median survival during the past decade, there has been little or no change in the overall mortality rate.[1,2] Women are commonly diagnosed with stage III/IV disease, for which 5-year survival rates are around 27% and 16%, respectively.[3–5] This has led to efforts over the past two decades to develop early detection strategies using serum CA125 and ultrasound.[6,7] Preliminary evidence from a previous randomised controlled trial suggests that screening sequentially with CA125 and ultrasound (multimodal screening) can result in a survival benefit.[8] Median survival was significantly increased in women who developed ovarian cancer in the screened group compared with the control group (72·9 *vs* 41·8 months, p=0·0112). Improved survival has also been reported in a single-arm ultrasound-based study.[9]

Refinements have been made to screening since the two previous studies, including the introduction of transvaginal ultrasound,[10] improvements in the interpretation of ultrasound findings using morphology-based indices,[9,11–13] and the development of a risk of ovarian cancer algorithm for the interpretation of serial CA125 results.[14,15]

The multicentre United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is a randomised controlled trial designed to provide definitive data on the effect of ovarian cancer screening on mortality, as well as comprehensively addressing the cost, acceptance, physical and psychosocial morbidity, and performance characteristics of multimodal screening and ultrasound-based screening.

Methods

Participants

Women were recruited through 13 regional trial centres located in National Health Service (NHS) Trusts in England, Wales, and Northern Ireland.[16] All women provided written consent. Eligibility criteria included age 50–74 years, and postmenopausal status defined as greater than and including 12 months, amenorrhoea following a natural or surgical menopause, or greater than and including 12 months of hormone-replacement therapy commenced for menopausal symptoms.[17] Women were excluded if they had a history of bilateral oophorectomy, active malignancy (women with a past history of malignancy were eligible if they had no documented persistent or recurrent disease and had not received treatment for >12 months), previous history of ovarian cancer, participation in other ovarian cancer screening trials, or increased risk of familial ovarian cancer. High-risk women were eligible for a separate trial: the United Kingdom Familial Ovarian Cancer Screening Study (UKFOCSS).[18]

The study was approved by the UK North West Multicentre Research Ethics Committee (00/8/34), with site-specific approval from the local regional ethics committees and the Caldicott guardians (data controllers) of the participating Primary Care Trusts.

***Panel 1:* Classification of ovarian morphology on ultrasound**

Normal

- Ovary of uniform hypoechogenicity and with a smooth outline with or without a single inclusion cyst or spots of calcifications
- Inclusion cyst must be single, less than 10 mm in diameter and not distort the outline of the ovary

Simple cyst

- A single, thin walled, anechoic cyst with no septa or papillary projections

Complex

- Any case in which the ovary has any non-uniform ovarian echogenicity, excluding single simple or inclusion cysts

Procedures

Invitations to participate in the trial were sent to women aged 50–74 years whose details were obtained from the age and sex registers of the participating 27 Primary Care Trusts.[16] On enrolment at the trial centres, women viewed an information video and participated in a group discussion before completing a datasheet, consent forms, and undergoing venepuncture. The recruitment questionnaires were sent to the coordinating centre, where they were scanned electronically using computerised intelligent character-reading and optical-mark-reading software (Teleform Elite version 8.1.1, Cardiff Software Inc, Vista, CA, USA). Any inconsistency or information not recognised by the data-capture software was verified manually by trained data-entry staff, who validated the computer-interpreted data. Once the custom-built trial-management system confirmed eligibility, participants were randomly assigned to either no treatment (control); annual CA125 screening (interpreted using a patented risk-of-ovarian-cancer algorithm) with transvaginal ultrasound scan as a second-line test (multimodal screening; MMS); or annual screening with transvaginal ultrasound (USS) alone in a 2:1:1 ratio with a computer-generated random number algorithm.

Randomisation was done as follows: first, the trial management system allocated a set of 32 random numbers to each trial centre; second, the lowest eight were allocated to the MMS group, the next eight to the USS group, and the remaining 16 to the control group; third, each successive volunteer within the trial centre was randomly allocated one of the random numbers and so randomly assigned a group; and finally, when all 32 random numbers had been used up a further set of 32 was generated. Randomisation was accomplished by the trial-management system using the visual basic randomisation statement and the Rnd function.

Following randomisation, letters were automatically printed and sent to each woman and their general practitioner confirming eligibility and randomisation status. Between April, 2001, and October, 2005, 202 638 women were enrolled in the trial. The women in the MMS and USS groups will be screened until Dec 31, 2011. All women will be followed up until Dec 31, 2014.

Screening tests

Two screening tests were used. The first was measurement of serum CA125. Blood samples were taken in gel tubes (8 mL gel separation serum tubes; Greiner Bio-One 455071, Stonehouse, UK) at the trial centres and transported overnight at ambient temperature to the central laboratory. All blood samples received more than 56 h after venepuncture were discarded and repeat samples requested. The blood was centrifuged at 1500*g* for 10 min and the serum separated. Excess serum was stored in aliquots. Serum CA125 concentrations were determined by electro-chemiluminescence sandwich immunoassay on an Elecsys 2010 (Roche Diagnostics, Burgess Hill, UK) using two monoclonal antibodies (OC125 and M11; Fujirebio Diagnostics AB, Göteborg, Sweden).

The other screening test was transvaginal ultrasound or, where this was not acceptable to a participant, transabdominal scan of the pelvis. The first scan (level 1 scan), and any repeat level 1 scans needed because of an unsatisfactory first scan, were done by type 1 sonographers, who were certified sonographers, trained midwives, or doctors with experience in gynaecological scanning who were working in the NHS. All staff who administered scans underwent additional training for assessment of postmenopausal ovaries. Level 2 scans were done by type 2 sonographers following the detection of an abnormality. Type 2 sonographers were experienced gynaecologists, radiologists, or senior sonographers (usually at superintendent grade in the NHS) with particular expertise in gynaecological scanning. Most scans were done on a dedicated Kretz SA9900 ultrasound machine (Medison, Seoul, South Korea).

Ovarian morphology and dimensions were assessed and volume determined using the formula for an ovoid ($d_1 \times d_2 \times d_3 \times 0 \cdot 532$). Ovarian morphology was classified as shown in panel 1. Detailed description of all features—the number and size of cysts, wall regularity, presence and thickness of septae, size of papillations, and echogenicity of the fluid contents—were recorded. Cysts and complex morphology were classified pictorially, initially using the format reported in the Kentucky screening trial,[19] and from 2003 onwards using the International Ovarian Tumour Analysis (IOTA) classification.[13]

Measurement of the ovary was important to confirm that it had been visualised, and for audit purposes. The volume of any ovarian cyst visualised was also measured using the above formula. Only cyst volume

***Panel 2:* Possible outcomes after level 1 and level 2 screens in the MMS group**

Level 1 screen

- Normal risk of ovarian cancer score: women returned to annual screening, with the next level 1 blood test scheduled on the next anniversary of the randomisation date
- Intermediate risk of ovarian cancer score: women were recalled for a repeat CA125 measurement 12 weeks after the screen. The risk of ovarian cancer was recalculated and triaged as for the level 1 screen. Any women whose risk of ovarian cancer remained intermediate after three CA125 tests were referred for a level 2 screen
- Elevated risk of ovarian cancer score: women were recalled for a level 2 screen in 6–8 weeks, with earlier screens arranged where there was a high index of suspicion

Level 2 screen

- Women with a normal transvaginal ultrasound scan result and normal or intermediate risk of ovarian cancer returned to annual screening, with the next level 1 test on the next anniversary of the randomisation date
- Women with a normal transvaginal ultrasound scan result but an elevated risk of ovarian cancer, or an unsatisfactory scan irrespective of risk of ovarian cancer status, underwent a repeat level 2 screen in 6 weeks and were triaged again on the basis of the results to annual screening or clinical assessment
- Women with an abnormal transvaginal ultrasound scan were referred for clinical assessment irrespective of their risk of ovarian cancer status

***Panel 3:* Possible outcomes after level 1 and level 2 screens in the USS group**

Level 1 screen

- Women with a normal scan returned to annual screening, with the next level 1 transvaginal ultrasound scan on the next anniversary of the randomisation date
- Women with an unsatisfactory scan result attended a repeat level 1 scan in 12 weeks. Women were returned to annual screening following two unsatisfactory scans
- Women with an abnormal level 1 scan were referred for a level 2 scan in 6–8 weeks, with earlier scans arranged where there was a high index of suspicion

Level 2 screen

- Women with a normal level 2 scan returned to annual screening, with the next level 1 transvaginal ultrasound scan on the next anniversary of the randomisation date
- An unsatisfactory level 2 scan led to a repeat level 2 scan in 6 weeks or earlier, and women were triaged on the basis of the findings to annual screening or clinical assessment
- Women with an abnormal scan were referred for clinical assessment

was used for scan classification. When ovaries were not visualised, the sonographers specified whether a good view of iliac vessels had been obtained or a poor view owing to obstruction by the bowel, fibroids, pelvic varicosities, or for other reasons. Ascites was defined as a maximum vertical pool measurement of greater than or equal to 10 mm.

Based on the visualisation and morphology of the ovaries, the scan was classified as either a normal scan, in which both ovaries had normal morphology or simple cysts less than 60 cm^3, or were not visualised but a good view of the iliac vessels was obtained; an_unsatisfactory scan, in which one or both ovaries were not visualised owing to a poor view; or an abnormal scan, in which one or both ovaries had complex morphology or simple cysts greater than 60 cm^3, or ascites.

Scan images were transferred weekly on magneto-optical discs for central archiving. Trial centres were able to request central review of ultrasound images.

Screening strategies

In the level 1 screen in the MMS group, women underwent venepuncture and serum CA125 measurement. The assay results were uploaded directly into the trial-management system, which calculated the risk of ovarian cancer using an algorithm developed previously.[15,17] The first risk of ovarian cancer determination was based on a single measurement of CA125 and the woman's age-specific incidence of ovarian cancer. Subsequent calculations of the risk of ovarian cancer were based on both the absolute CA125 concentration and the rate of change in CA125 concentration. The risk of ovarian cancer summarises, in one number, the information about risk of ovarian cancer, therefore simplifying the practical implementation of the screening protocol. Women were triaged into three risk groups on the basis of their risk of ovarian cancer, which determined whether they returned to annual screening or went on to have a repeat CA125 measurement or level 2 screen (panel 2). Level 2 screening involved venepuncture for repeat CA125 assay and a transvaginal ultrasound scan. The results of the level 2 screen triggered three possible courses of action, as shown in panel 2.

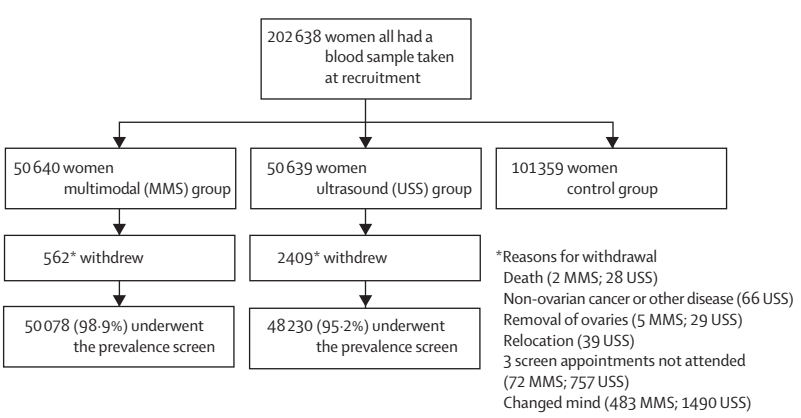

Figure 1: **Randomisation and initial (prevalence) screen**

The initial cutoffs used for intermediate and elevated risk of ovarian cancer were greater than or equal to 1/1818 and greater than or equal to 1/500, respectively. As the proportion of women classified into these risk categories were less than the proposed 15% and 2%, respectively, the cutoffs were revised on April 1, 2005, to greater than or equal to 1/3500 and greater than or equal to 1/1000 after extensive data review by the independent data monitoring and ethics and trial steering committees. 86·4% of the prevalence screen CA125 concentrations were classified using the pre-2005 cutoffs.

Women randomly assigned to the USS group had a transvaginal ultrasound scan at their regional trial centre. There were three possible courses of action depending on the results of the level 1 scan, including referral for a level 2 scan, the results of which triggered one of a further three courses of action, as shown in panel 3.

All clinicians were provided written information on the risk estimates for malignancy associated with the various morphological classifications from the IOTA series once the estimates had been presented at the annual European Society of Gynaecological Oncology meeting in 2003. Additionally, clinicians were made aware that women who had previously undergone a hysterectomy had an increased incidence of adhesions and peritoneal pseudocysts that may be reported as multilocular adnexal cysts.

Clinical assessment

This was undertaken by a designated clinician and included clinical evaluation and investigations as appropriate. These included serum CA125 in women in the ultrasound group, repeat transvaginal scans and doppler studies, CT/MRI of the abdomen and pelvis, and occasionally assessment of other tumour markers. For women in the MMS group with a normal transvaginal ultrasound scan but elevated risk of ovarian cancer, clinical assessment included ruling out other causes of increased CA125 concentrations. The management plan took the views of the individual into account, and also accounted for any significant comorbidity, the specific morphological features of the detected lesion, and history of a previous hysterectomy or major pelvic surgery that could be responsible for false-positive ultrasound appearances.

For women who underwent surgery, the recommendation was removal of both ovaries and fallopian tubes for histopathological examination, even if the ovaries appeared macroscopically normal. Where pelvic adhesions were present and there was an increased risk of complications, the clinician could opt to remove only the ovary found to have an abnormality on ultrasound and not proceed to remove the contralateral ovary.

	MMS group (N=50 078)	USS group (N=48 230)	Overall (N=98 308)
Age (years) at randomisation	60·6 (56·1–66·2)	60·5 (56·0–66·0)	60·6 (56·0–66·1)
Years since last menstruation at randomisation	11·4 (5·3–18·5)	11·2 (5·2–19·0)	11·3 (5·3–18·4)
Duration of HRT use (years) at randomisation if applicable	8·1 (4·6–12·0)	8·2 (4·6–12·1)	8·1 (4·6–12·1)
Duration of OCP use (years) if applicable	5 (2–10)	5 (2–10)	5 (2–10)
Miscarriages (pregnancies <6 months)	0 (0–1)	0 (0–1)	0 (0–1)
Number of children (pregnancies >6 months)	2 (2–3)	2 (2–3)	2 (2–3)
Height (cm)	162·6 (157·5–165·1)	162·6 (157·5–165·1)	162·6 (157·5–165·1)
Weight (kg)	67·6 (60·3–76·2)	67·1 (60·3–76·2)	67·6 (60·3–76·2)
Ethnic origin			
White	48 340 (96·5)	46 509 (96·4)	94 849 (96·5)
Black	655 (1·3)	678 (1·4)	1333 (1·4)
Asian	431 (0·9)	419 (0·9)	850 (0·9)
Other	417 (0·8)	384 (0·8)	801 (0·8)
Missing	235 (0·5)	240 (0·5)	475 (0·5)
Hysterectomy	9620 (19·2)	9078 (18·8)	18 698 (19·0)
Ever use of OCP	29 743 (59·4)	29 048 (60·2)	58 791 (59·8)
Use of HRT at recruitment	9379 (18·7)	9046 (18·8)	18 425 (18·7)
Personal history of cancer*	2936 (5·9)	2819 (5·8)	5755 (5·9)
Personal history of breast cancer	1831 (3·7)	1793 (3·7)	3624 (3·7)
Maternal history of ovarian cancer	789 (1·6)	747 (1·5)	1536 (1·6)
Maternal history of breast cancer	3128 (6·2)	3070 (6·4)	6198 (6·3)

Data are median (IQR) or number (%). HRT=hormone-replacement therapy. OCP=oral contraceptive pill. *Includes women who have previously had breast cancer.

Table 1: **Baseline characteristics of UKCTOCS participants who underwent the prevalence screen**

Hysterectomy was only undertaken where there were clear clinical indications. The approach to surgery depended on the results of the preoperative investigations. The primary intervention in most cases was laparoscopy with the intention of performing a laparoscopic bilateral salpingo-oophorectomy. A laparotomy was undertaken if clinical findings or laparoscopy led to a strong suspicion of ovarian cancer, or if a laparoscopic procedure was not felt to be appropriate for other reasons. Women found to have ovarian or tubal cancer at a primary laparoscopic procedure underwent a subsequent staging procedure.

A follow-up plan was drawn up if, after clinical assessment, investigation, and discussion with the woman, a decision was made to manage the findings conservatively. Most women were followed up with a transvaginal ultrasound scan and a serum CA125 assessment at 3 months with a possible repeat at 6 months, and returned to annual screening if the findings were unchanged at this review.

Follow-up

All participants are being followed up through a flagging study with the NHS Information Centre for Health and Social Care (formerly Office for National Statistics, ONS) in England and Wales, and with the Central Services Agency and Cancer Registry in Northern Ireland. This provides regular notification of any cancer registrations or deaths in the cohort. For the purpose of this analysis, up-to-date cancer registration data was obtained from the agencies on June 13, 2008. Additionally, women continued to attend for subsequent annual screens, and those who had been in the trial for 3·5 years following randomisation were sent follow-up questionnaires.

Confirmation of diagnosis

Medical records of procedures undertaken after an abnormal screen result were obtained. Operative notes and histopathology and cytology reports of women who underwent surgery were reviewed to confirm the diagnosis. In the case of a diagnosis of cancer, further information was obtained, including the discharge summary, multidisciplinary team meeting notes, and other correspondence. This was also done for all cases of women found to have ovarian, tubal, peritoneal, or disseminated cancer of unknown origin (International Classification of Disease [ICD]-10 codes C56, C57.0, C48.2, and C80), through screening, flagging for cancer registration, follow-up questionnaire, or directly from the participants. The final diagnosis including the primary site, stage, and grade of any cancer was made by an independent outcomes committee. Where documentation was insufficient to arrive at a definitive conclusion, further information and clarification was sought from the team who treated the patient. Complex cases were discussed by the whole committee and a consensus diagnosis reached. Histology review was undertaken only for those cases

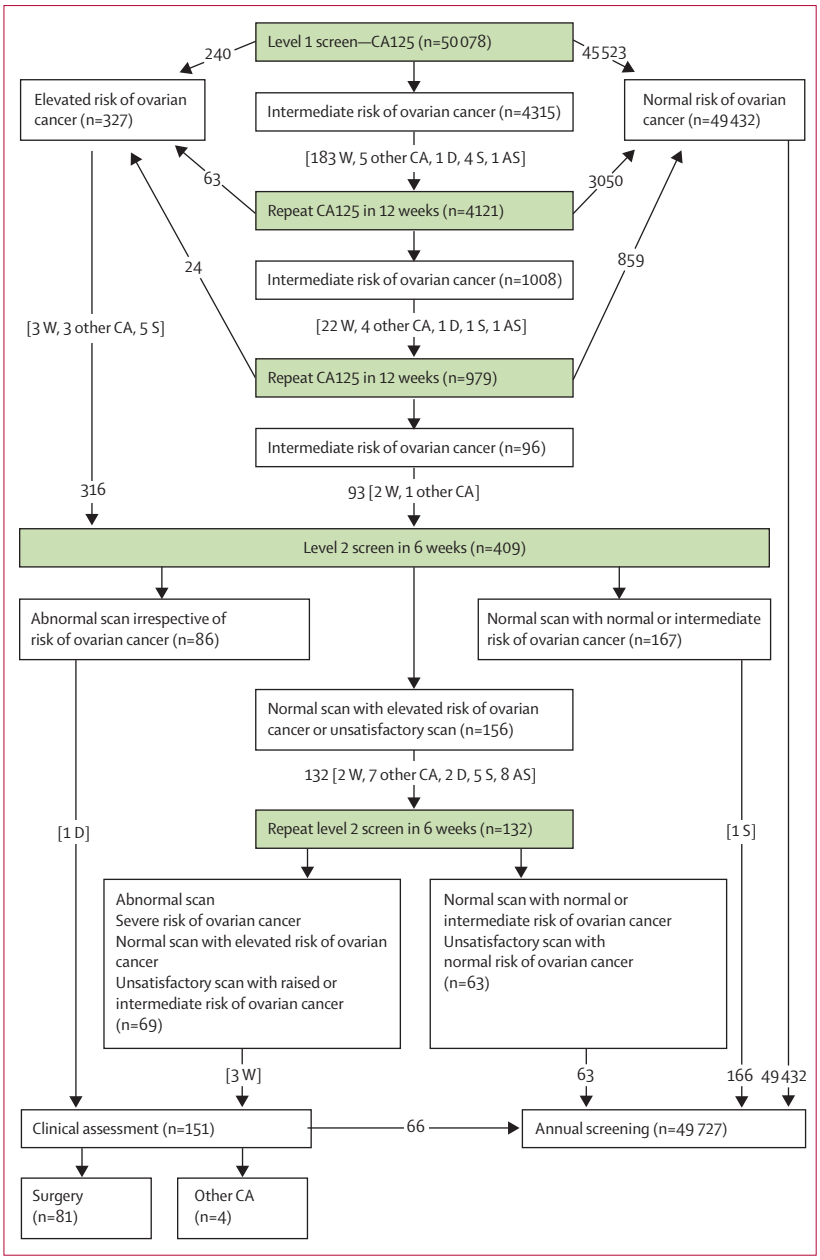

Figure 2: **Multimodal screening (MMS) algorithm and outcome of initial screen**
Boxes represent tests (green) or results. Numbers inside boxes indicate the number of volunteers undergoing a specific test or having a certain result. Where a test or result can occur via multiple routes the numbers of volunteers per route are indicated on the arrows. Numbers in square brackets indicate volunteers who deviated from the protocol and the reason. AS=annual screening. CA=diagnosed with other cancer. D=died. S=surgery. W=withdrew.

where the pathology report showed ambiguity in relation to the site of origin, staging, and grading of the cancer, or where there was a discrepancy between the pathology report and the national cancer registration.

The hospital notes of women who underwent surgery were obtained. All surgical complications were confirmed by review of the surgical and clinical notes by a senior trial gynaecological oncologist blinded to the randomisation group.

Statistical analysis

The primary outcome measure in UKCTOCS is ovarian cancer mortality and the primary comparison is based on an intention-to-treat analysis between the control group and both screened groups combined (MMS plus USS). However, as the operating characteristics of the two screening groups are different a comparison between the control group and an individual screen group (MMS or USS) is of equal interest. Randomisation in UKCTOCS was completed in October, 2005. Women in the MMS and USS groups will be screened until Dec 31, 2011, and all women will be followed up until Dec 31, 2014. The design provides greater than 90% power to detect a 30% reduction in ovarian cancer mortality between the control and combined screening groups, and greater than 80% power to detect a 30% reduction in mortality between the control and either one of the individual screening groups, with both comparisons tested at a significance level of 0·05. It is important to note that if one of the comparisons (control *vs* MMS or USS) is significant and the other is not, the result would not necessarily imply that one method is significantly better than the other. Only the direct comparison between the two methods will address this issue. If, as anticipated, the difference in ovarian cancer mortality between the two screened groups is modest, then this study will have limited power to detect such differences. The choice of screening strategy will then be based on other outcome measures such as sensitivity, positive-predictive value, morbidity, quality of life, and health economics.

This paper presents the outcome of the prevalence screen in women randomly assigned to either MMS or USS. The prevalence screen was defined as a single or series of serum CA125 assays with or without transvaginal ultrasound scan (MMS) or transvaginal ultrasound scan alone (USS) culminating in surgery or a return to annual screening. The screen was considered positive (screen positive) if the woman was referred for surgery and negative (screen negative) if the woman was returned to annual screening. All women in the cohort were censored 1 year after their last transvaginal ultrasound scan or CA125 serum assay in the prevalence screen. The primary outcome measure was primary ovarian or fallopian tube cancer (ICD-10 code C56 and C57.0, respectively) diagnosed within 12 months of the last scan or serum CA125 test in the prevalence screen. Women with primary peritoneal cancer (ICD-10 code C48.2) and those with ovarian neoplasms of uncertain behaviour (ICD-10 code D39.1) were not included in the outcome measure. Sensitivity, specificity, and positive-predictive values were calculated for the MMS and USS screens separately and in combination. Subgroup analyses of primary invasive epithelial ovarian cancers (excluding borderline malignancies) were also undertaken. The data was analysed with STATA version 10.0. No other analysis of these data has been undertaken at the time of writing.

This trial is registered as ISRCTN22488978 and with ClinicalTrials.gov, number NCT00058032.

Role of the funding source
The funding bodies had no role in the study design, data collection, analysis, interpretation or writing of the report. The corresponding author had full access to all data in the study. The UKCTOCS investigators had final responsibility for the decision to submit the report for publication.

Results

The trial profile is shown in figure 1. The most common reasons for withdrawal were death (two MMS; 28 USS), non-ovarian cancer or other disease (66 USS), removal of ovaries (five MMS; 29 USS), relocation (39 USS), failure to attend three appointments for the screen (72 MMS; 757 USS), and participant changing her mind (483 MMS; 1490 USS). The baseline characteristics of women in the prevalence screen are shown in table 1. In accordance with good practice for randomised controlled trials we did not statistically compare the baseline characteristics of the women assigned to the two groups;[20,21] the groups were well balanced (table 1).

Of the 50 078 women who underwent a prevalence screen in the MMS group, 45 523 (90·9%) were classified as low risk by the risk of ovarian cancer algorithm and returned to annual screening. 240 (0·5%) had an elevated risk and were referred for a level 2 screen. 4315 (8·6%) women had intermediate risk leading to a recommendation for repeat CA125 testing in 3 months. 169 of these intermediate-risk women were referred for a level 2 screen. 409 (0·8%) women underwent a level 2 screen, after which 167 (0·3%) women were referred for clinical assessment and 81 proceeded to surgery. Additionally, 16 women had clinical assessment and surgery following an abnormal screen without additional tests, as per protocol. Overall, 4555 (9·1%) women required a repeat test and 97 (0·2%) had surgery (figure 2). In the course of the screen, five women in the MMS group died from unrelated causes, 24 were diagnosed with cancers other than ovarian cancer, and 215 withdrew from the trial.

Of the 48 230 women randomly assigned to the USS group, 42 416 (87·9%) had transvaginal ultrasound scans, 4325 (9·0%) had transabdominal ultrasound scans, and 1489 (3·1%) had both. 42 451 (88·0%) women had normal scans and were returned to annual screening. 2774 (5·8%) women were found to have abnormalities and referred directly for a level 2 screen. 3005 (6·2%) women had unsatisfactory scans necessitating a repeat level 1 scan, and 110 women from this group were referred for a level 2 screen. Overall, 5779 (12·0%) women in the USS group required a repeat test. 2785 (5·8%) women underwent a level 2 screen. Of these women, 1894 (3·9%) were referred for clinical assessment, and 775 women proceeded to surgery. Additionally, 70 women had clinical assessment and surgery following an abnormal screen without additional screens, as per protocol. Overall, 845 (1·8%) women in the USS group had surgery (figure 3). Six women in the USS group died from unrelated causes during the course of screening, seven were diagnosed with cancers other than ovarian cancer, and 252 withdrew from the trial.

Overall, 942 (0·95%) of the 98 308 women screened underwent surgery as a result of the prevalence screen, with 8·7 women in the USS group undergoing surgery for every one woman from the MSS group who underwent surgery. The number of operations in the MMS group was significantly lower than in the USS group (p<0·005). There was a difference in surgical approach between the two groups, with 75 of 97 (77·3%) operations in the MMS group involving laparotomy versus 397 of 845 (46·9%) in the USS group (table 2). 834 (47 MMS, 787 USS) women who underwent surgery were found to have benign pathology or normal ovaries (table 3), of whom 24 (2·9%; 95% CI 1·7–4·0) experienced a major complication. This included two of 47 (4·3%; 95% CI 0·0–10·3) women in MMS and 22 of 787 (2·8%; 95% CI 1·65–3·95) women in the USS group (table 2). The complications were six cases of perforation of a hollow viscus, two cases of excessive haemorrhage requiring further surgery, one re-admission for portal site pain and surgery to remove an endometriotic

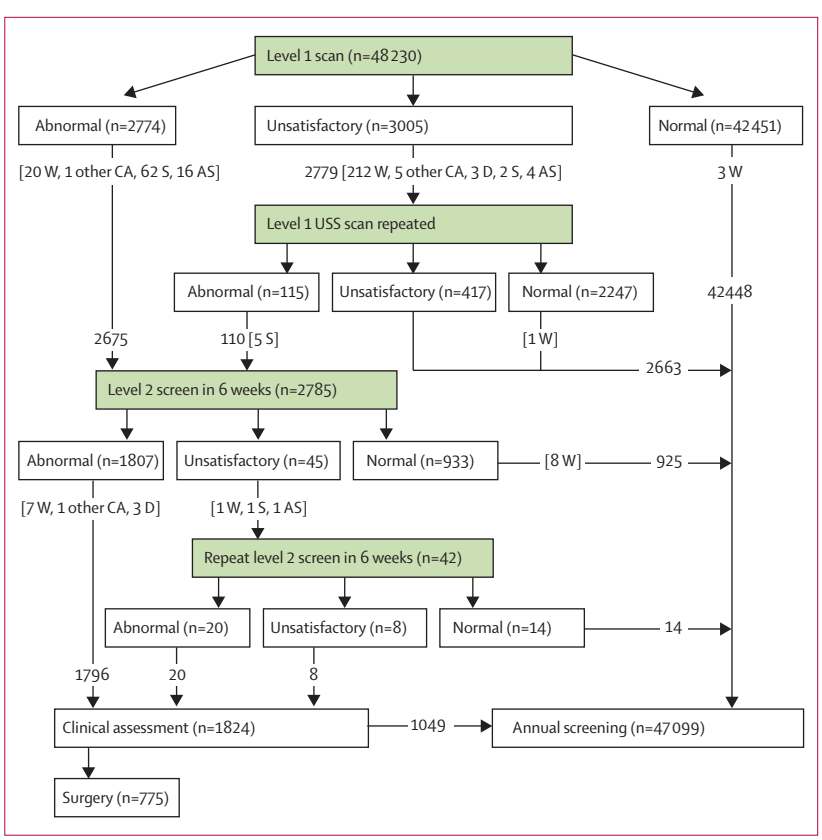

Figure 3: **Ultrasound screening (USS) algorithm and outcome of initial screen**
Boxes represent tests (green) or results. Numbers inside boxes indicate the number of volunteers undergoing a specific test or having a certain result. Where a test or result can occur via multiple routes the numbers of volunteers per route are indicated on the arrows. Numbers in square brackets indicate volunteers who deviated from the protocol and the reason. AS=annual screening. CA=diagnosed with other cancer. D=died. S=surgery. W=withdrew.

	MMS (N=50 640)	USS (N=50 639)	Overall (N=101 279)
Number who underwent the prevalence (first) level 1 screen*	50 078 (98·9%)	48 230 (95·2%)	98 308 (97·1%)
Normal	45 523 (90·9%)	42 451 (88·0%)	87 974 (89·5%)
Intermediate risk of ovarian cancer and unsatisfactory scans	4315 (8·6%)	3005 (6·2%)	7320 (7·4%)
Elevated risk of ovarian cancer and abnormal scan	240 (0·5%)	2774 (5·8%)	3014 (3·1%)
Number women who underwent (first) level 2 screen†	409 (0·8%)	2785 (5·8%)	3194 (3·2%)
Normal	167 (40·8%)	933 (33·5%)	1100 (34·4%)
Abnormal	86 (21·0%)	1807 (64·9%)	1893 (59·3%)
Unsatisfactory	156 (38·1%)	45 (1·6%)	201 (6·3%)
Number of women referred for clinical evaluation	167 (0·3%)	1894 (3·9%)	2070 (2·1%)
Number of women who underwent surgery‡	97 (0·2%)	845 (1·8%)	942 (1·0%)
Diagnostic laparoscopy	6 (6·2%)	34 (4·0%)	40 (4·2%)
Operative laparoscopy	16 (16·5%)	413 (48·9%)	429 (45·5%)
Diagnostic laparoscopy and laparotomy	6 (6·2%)	80 (9·5%)	86 (9·1%)
Laparotomy	69 (71·1%)	317 (37·5%)	386 (41·0%)
Unknown	0 (0·0%)	1 (0·1%)	1 (0·1%)

*Fisher's exact test, significant difference (p<0·0001) between MMS and USS in results of level 1 screen. †Fisher's exact test, significant difference (p<0·0005) between MMS and USS in results of level 2 screen. ‡Fisher's exact test, significant difference (p<0·0001) between MMS and USS in surgical approach. Owing to very large sample sizes, the p values tend to imply statistical difference where clinically meaningful difference is minimal.

Table 2: **Outcome of prevalence screen in UKCTOCS**

	MMS	USS	Overall
Total surgeries	97	845*	942
Denied access to notes	0	1	1
Diagnostic laparoscopy, ovary normal, not removed	6	34†	40
Normal ovaries	0	15	15
Benign ovarian neoplasm	40	732	772
Ovarian neoplasm of uncertain behaviour (ICD-10 D39.1)	1‡	5	6
Primary peritoneal cancer (ICD-10 C48.2)	1	1	2
Other non-ovarian cancer	4§	7¶	11
Metastatic ovarian cancer	3‖	5**	8
Non-epithelial neoplasm of ovary (ICD-10 C56)	0	1	1
Primary borderline epithelial neoplasm of ovary (ICD-10 C56)	8	20	28
Primary invasive epithelial neoplasm of ovary (ICD-10 C56)	32	23	55
Primary invasive epithelial neoplasm of fallopian tube (ICD-10 C57.0)	2	1	3
Total malignant neoplasms of ovary (ICD-10 C56) and fallopian tube (ICD-10 C57.0)	42	45	87
Screen-negative cancers within 1 year of screen			
Borderline epithelial neoplasm of ovary (ICD-10 C56)	1	0	1
Primary invasive epithelial neoplasm of ovary (ICD-10 C56)	4	8	12
Total malignant neoplasm of ovary (ICD-10 C56) and fallopian tube (ICD-10 C57.0)	5	8	13

*One participant refused access to notes, at the time of writing there is no ONS registration of a cancer for this case. †One woman was diagnosed with ovarian cancer at a second operation undertaken 22 months after the prevalence screen. ‡Patient developed postmenopausal bleeding while waiting for a repeat CA125 test and was diagnosed to have synchronous endometrial cancer and ovarian granulosa cell tumour. §Two endometrial cancers, one stomach cancer, one follicular lymphoma. ¶Three endometrial cancers, one cervical cancer, one anal cancer, one lymphoma, and one multiple myeloma. ||One pancreatic cancer, one colorectal cancer, and one cancer of the appendix. **Three breast cancers, one endometrial cancer, and one cancer of the appendix.

Table 3: **Histology in women who underwent surgery as a result of screening (screen positives)**

nodule and residual ovary in left pelvic side wall, one pulmonary embolism, two cases of deep-vein thrombosis, four cases of wound dehiscence, one wound haematoma, two hernias, one significant ileus, one bowel obstruction, one bowel fistula, and two cases of significant infection.

Ovarian or tubal malignancies were detected in 87 women: 42 in the MMS group and 45 in the USS group (table 3). Of these malignancies, eight in the MMS group and 20 in the USS group were borderline epithelial neoplasms (p=0·013). There was no significant difference (Fisher's exact test p=0·229) in the number of stage III borderline cancers in the MMS (two of eight) compared with the USS group (one of 20). Fewer primary invasive epithelial cancers (24 *vs* 34) were detected in the USS than in the MMS group. Overall, 28 of the 58 (48·3%; 95% CI 35·0–61·8) primary invasive epithelial cancers detected were stage I/II. There was no difference (Fisher's exact test p=0·396) in the stage distribution between the two groups (table 4).

In the MMS group, 33 (78·6%) of the 42 women with ovarian or tubal malignancies had ovarian cancer detected as a result of an elevated risk of ovarian cancer on the level 1 screen (first blood test). The median time to surgery from the level 1 screen in this group was 75·0 days (IQR 55·8–114·8). In the USS group, all 45 (100%) women had ovarian cancer detected as a result of an abnormal scan on the level 1 screen (first scan). The median time to surgery from level 1 scan in these women was similar to that for women in the MMS group: 81·5 days (IQR 60·3–112·5). This period included the level 2 screen within 42 days, and then referral, clinical assessment, and often further imaging before surgery. However, nine (21·4%) of the women in the MMS group with ovarian or tubal malignancies had ovarian cancer detected after an intermediate risk of ovarian cancer at the level 1 screen that led to repeat tests. The median time to surgery from the level 1 screen in these women was 273·9 days (IQR 220·0–331·0), since the protocol for managing intermediate risk was to repeat CA125 tests

over a period of 8–10 months (figure 2). Of the nine women with ovarian cancer detected after an intermediate risk of ovarian cancer at the level 1 scan, seven had abnormal level 2 scans. One woman initially had a normal level 2 scan, and one woman had two normal level 2 scans but was operated on based on a severe risk of ovarian cancer classification.

Four ovarian cancers have not been included in the analysis of performance characteristics as they were diagnosed more than 1 year after the last test on the prevalence screen. They include one woman in the MMS group and two from the USS group who withdrew from the trial after their level 1 screen and had an ovarian cancer diagnosed more than 1 year later. Additionally, a fourth woman had an abnormal prevalence screen in the USS group and underwent diagnostic laparoscopy. She was thought to have an ovarian fibroma, and her ovaries were not removed (table 3). At an incident screen 22 months later, an increase in size of the mass on ultrasound prompted bilateral salpingo-oophorectomy, and she was diagnosed with stage IC papillary serous cystadenocarcinoma.

Median follow-up from the last test on the prevalence screen to cancer registration update was 4·57 years (IQR 3·68–5·54). As information on cancers can take up to 3 years to be recorded by the national cancer registries, we explored in detail the other sources of follow-up data in the 12 658 women for whom time from censorship (1 year from the date of the last scan or CA125 test on the prevalence screen) to cancer registry follow up on June 13, 2008, was less then 3 years. In this cohort, after the censorship date, we had additional confirmation of ovarian cancer status in 11 336 women, as they had attended for further screening, and in an additional 17 women through returned follow-up questionnaires. The source of verification of ovarian cancer status was limited to cancer registry follow up that was less than 3 years from the date of censorship in only 1275 women (1·3% of the entire cohort; table 5).

13 additional (interval) ovarian and tubal cancers were diagnosed clinically within 1 year of a normal prevalence screen result. This included one primary borderline and four invasive epithelial ovarian cancers in the MMS group, and eight primary invasive epithelial ovarian cancers in the USS group (table 3). The CA125 concentrations at the prevalence screen ranged from 7 to 24 IU/L in the five women with ovarian and tubal cancers in the MMS group, and the cancers were diagnosed at 92, 204, 254, 294, and 329 days after the screen. All of the prevalence scans were normal in the USS group, and the cancers were diagnosed at 30, 203, 255, 267, 278, 293, 301, and 341 days after the screen.

For all primary ovarian and tubal cancers, the sensitivity, specificity, and positive-predictive values for MMS and USS screening are shown in table 6. There were 2·3 (42 of 97) operations per case of ovarian cancer in the MMS and 18·8 (45 of 845) operations per case of ovarian cancer in the USS group. There was a significant difference in specificity between the two groups (p<0·0001), but no difference in sensitivity (p=0·564). When the analysis was restricted to primary invasive epithelial ovarian and tubal cancers, sensitivity was, compared with values when all cancers were included (table 6), much the same in the MMS group, but lower in the USS group, although the difference in sensitivity between MMS and USS was still not statistically significant (p=0·126). 2·9 (34 of 97) operations were done per case in the MMS and 35·2 (24 of 845) operations per case in the USS group (table 6).

	Screen positive			Screen negative		
	MMS	USS	Overall	MMS	USS	Overall
Stage						
I	14	10	24	3	0	3
II	2	2	4	0	0	0
III	18	10	28	1	7*	8
IV	0	2	2	0	1	1
Early (I/II) stage cancers (%)	47·1%	50·0%	48·3%	75·0%	0·0%	25·0%
Lower 95% CI	29·8%	29·1%	35·0%	19·4%	0·0%	5·5%
Upper 95% CI	64·9%	70·9%	61·8%	99·4%	41·0%	57·2%
Morphology						
Serous	21	14	35	0	2	2
Endometrioid	5	3	8	1	0	1
Clear cell	0	5	5	1	0	1
Carcinosarcoma	1	0	1	1	0	1
Adenocarcinoma	7	2	9	1	6	7
Grade						
1	3	2	5	0	0	0
2	6	2	8	2	0	2
3	24	14	38	2	6	8
Not graded	1	6	7	0	2	2

*In two cases a diagnosis was made on the basis of ascitic fluid cytology, omental biopsy, and imaging: primary surgery was not undertaken.

Table 4: **Characteristics of primary invasive epithelial ovarian and tubal cancers (ICD-10 C56 and C57.0)**

	MMS (N=50 078)	USS (N=48 230)	Overall (N=98 308)
ONS follow-up >3 years from censorship date or when death certificate was available	45 544 (90·9%)	40 106 (83·2%)	85 650 (87·1%)
Number of women who have had an appointment after censorship date*	4071 (8·1%)	7295 (18·2%)	11 366 (11·6%)
Number of women who have completed a follow-up questionnaire after censorship date*	4 (0·01%)	13 (0·03%)	17 (0·02%)
Remaining*	459 (0·9%)	816 (1·7%)	1275 (1·3%)

*In women with ONS follow-up <3 years from censorship date. Censorship date is 1 year from the date of the last scan or CA125 in the prevalence screen.

Table 5: **Details of follow-up of women who underwent screening**

Discussion

Both a CA125-based and an ultrasound-based screening strategy are feasible on a large scale. On the initial screen, almost half of the cancers detected were in stage I/II.

	MMS	USS	Overall	p value*
Total				
Number of women	50 078	48 230	98 308	..
Number of surgeries	97	845	942	..
Primary ovarian and tubal malignancies (ICD-10 C56 and C57.0) within 1 year of prevalence screen†				
Screen positives	42	45	87	..
Screen negatives	5	8	13	..
Sensitivity	89·4%	84·9%	87·0%	0·564
95% CI	76·9–96·5	72·4–93·3	78·8–92·9	..
Specificity	99·8%	98·2%	99·0%	<0·0001‡
95% CI	99·8–99·8	98·1–98·4	99·0–99·1	..
Positive-predictive value	43·3%	5·3%	9·2%	..
95% CI	33·3–53·8	3·9–7·1	7·5–11·3	..
Number of operations per screen positive	2·3	18·8	10·8	..
Primary invasive epithelial ovarian and tubal malignancies within 1 year of prevalence screen§				
Screen positives	34	24	58	..
Screen negatives	4	8	12	..
Sensitivity	89·5%	75·0%	82·9%	0·126
95% CI	75·2–97·1	56·6–88·5	72·0–90·8	..
Specificity	99·8%	98·2%	99·0%	<0·0001‡
95% CI	99·8–99·8	98·1–98·4	99·0–99·1	..
Positive-predictive value	35·1%	2·8%	6·2%	..
95% CI	25·6–45·4	1·8–4·2	4·7–7·9	..
Number of operations per screen positive	2·9	35·2	16·2	..

*Fisher's exact test. †Includes borderline and ovarian neoplasm of uncertain behaviour. ‡Due to very large sample sizes the p values tend to imply statistically significant difference where clinically meaningful difference is minimal. §Borderline epithelial ovarian cancers and ovarian neoplasms of uncertain behaviour treated as false positives.

Table 6: **Performance characteristics for detection of malignant ovarian and tubal neoplasms (ICD-10 C56 and C57.0) in the prevalence screen**

Specificity was higher in the MMS than in the USS group, resulting in fewer repeat tests and almost nine times fewer operations per cancer detected. The stage distribution of the screen-detected primary invasive cancers was similar in both groups. However, more borderline epithelial ovarian neoplasms were detected in the USS group than in the MMS group. Among missed cancers, there were slightly more stage I cancers in the MMS than in the USS group.

To our knowledge, this is the largest randomised controlled trial of ovarian cancer screening to date. It is also the first ovarian cancer screening trial to randomly assign women to two screening strategies, enabling the performance of both strategies to be compared directly. The prevalence screen involved almost 100 000 women, with systematic follow-up to detect missed cancers. The main strengths of the study are recruitment by random invitation using the health-authority registers of 27 Primary Care Trusts, the multicentre design involving recruitment and screening through 13 NHS Trusts, the scale of the trial, high compliance with screening, randomisation to two well-defined screening strategies, and an independent review of surgical outcomes and detailed follow-up of the entire cohort. These factors provide confidence about the validity of the findings.

ICD-10 codes were used when reporting outcomes so that comparisons across studies can be made.[22] Many previous studies have included primary peritoneal cancers (ICD-10 C48.2) as true positives,[23] while others have included ovarian granulosa cell tumours under primary ovarian cancers, and have not classified them separately as ovarian neoplasm of uncertain behaviour (ICD-10 D39.1).[9] We have excluded both from our primary outcome measure. Although primary peritoneal cancers are treated similarly to advanced primary ovarian carcinomas, neither the MMS nor the USS screening strategies were developed to detect them. However, data on these cancers would be interesting, so primary peritoneal cancers are listed separately in table 3. By contrast, granulosa cell tumours are unlikely to contribute significantly to mortality, and their inclusion makes comparison with national statistics difficult. However, we have included primary borderline (low malignant potential) ovarian neoplasms, since they share the same ICD-10 code (C56) as primary invasive epithelial ovarian cancers. Their inclusion is therefore the only way to compare trial data with national and international incidence and mortality statistics. However, a subgroup analysis excluding these cancers has also been undertaken.

84 primary ovarian (ICD-10 C56) and three tubal cancers (ICD-10 C57.0) were detected on screening, with a further 13 primary ovarian cancers diagnosed clinically in the ensuing year. Overall, the total numbers of cases were similar, but there was a difference in the number of borderline cancers between the two groups. 19% (eight of 42) of the primary ovarian cancers detected were borderline in the MMS group, compared with 15% reported in clinical series.[24] However, 44% (20 of 45) of the primary ovarian cancers detected in the USS group were borderline. Borderline ovarian tumours have 10-year survival rates in excess of 95%.[25] Of the 28 borderline tumours detected, only three were stage III. This highlights an issue that has already become a significant problem in other cancer-screening strategies—the detection of cancers that may never have been diagnosed in an individual's lifetime had they not been screened. Overdiagnosis or pseudodisease may be thought of as pathological diagnoses detected by screening or autopsy, but which have little clinical relevance. It could be argued that these cases would be best classified as false positives. In cancer screening, estimates of overdetection range from 3 to 50% of cases in breast cancer screening[26] and 22 to 34% in prostate cancer screening.[27] A similar rate of 25% overdetection has been reported with chest radiography for lung cancer, with the use of CT expected to result in higher rates.[28] In ovarian cancer, such false positives may include ovarian neoplasms of uncertain behaviour (ICD-10 D39.1) and borderline disease. Once borderline cancers are detected during screening, it is difficult not to operate given that borderline and stage I invasive

ovarian cancers share common morphological features on ultrasound imaging.[29] The novel design of UKCTOCS, which randomly assigns women to two very different screening strategies, provides some insight into the extent of overdiagnosis inherent in the different screening strategies. The results indicate that pseudodisease will be less apparent with a serum CA125-based ovarian cancer screening strategy than with USS screening. This accords with results from the prevalence screen of the Prostate Lung Colon Ovarian Cancer (PLCO) screening trial, in which only one of nine borderline ovarian neoplasms were detected by CA125 screening, whereas all nine were detected by ultrasound.[23] There is a possibility that some of these borderline tumours may progress to invasive cancers if undetected, although there is little evidence to support this. Differences between the screening groups on later follow-up should help to elucidate the issue, and allow definite estimates of overdiagnosis to be calculated.

Given the issues surrounding borderline disease, a separate analysis was done excluding borderline lesions, effectively treating such lesions as false positives. This resulted in a fall in the sensitivity of the USS screen from about 85% to 75%, with an attendant increase in the ratio of operations per true positive from 19:1 to 35:1. The detection rates in the MMS group remained at around 89%, with a small increase in the ratio of operations per true positive from 2·3:1 to 2·9:1. Unlike specificity, the difference in sensitivity between the two screening groups was not statistically significant (table 6). However, the clinical effect of this difference will depend on the mortality effect on follow up, and on issues such as patient satisfaction and acceptability.

Various factors may contribute to the detection of more primary invasive epithelial cancers in the MMS group than in the USS group. Most notably, the MMS strategy incorporates follow-up of women initially classified as intermediate risk with repeat CA125 tests for a period of 8–10 months. 21·4% (nine of the 34) of the cancers in the MMS group were identified via this pathway. It is possible that a transvaginal ultrasound scan done at the time of the level 1 screen, 7·5 months earlier than when the transvaginal ultrasound scan was actually done in these women, would not have detected an abnormality. Although an unsatisfactory scan in the USS group does lead to a repeat level 1 scan in 3 months, the follow-up is not equivalent to that of an intermediate risk of ovarian cancer, as women are returned to annual screening after two unsatisfactory scans but have continued follow-up with CA125 after two intermediate risk of ovarian cancer results. Ultrasound, unlike CA125, has a subjective element, and it is possible that the heightened awareness of a sonographer in view of rising CA125 concentrations contributed to a positive diagnosis. However, it is noteworthy that two of the nine women in the MMS group who were diagnosed after initially being classified as intermediate risk had normal scans initially. In one of these women surgery was done on the basis of a severe risk of ovarian cancer classification despite a repeat normal ultrasound scan. A detailed analysis of all the cancers detected in both groups is underway, and might shed more light on this issue.

	Van Nagell and colleagues,[9] 2007	Kobayashi and colleagues,[22] 2008	PLCO,[23] 2005	UKCTOCS USS group	UKCTOCS MMS group
Study design	Single-arm prospective study	RCT with one screening strategy in study group	RCT with one screening strategy in study group	RCT with two screening strategies in the study group	RCT with two screening strategies in the study group
Screening strategy	Ultrasound	Physical exam, ultrasound, and CA125	Ultrasound and CA125	Ultrasound	CA125 interpreted by ROC algorithm
Number of women screened	25 327	41 688	28 816	48 227	50 078
Mean number of screens per women	4·8	5·4	1 (first)	1 (first)	1 (first)
Number of women who had surgery	364	903	570	845	97
Primary epithelial ovarian and tubal cancers (ICD-10 C56, C57.0)*					
Number of women with primary epithelial ovarian and tubal cancers	39	27	27	45	42
Interval (missed) cancers diagnosed within 1 year of screen	9	8	†	8	5
Apparent sensitivity	81%	77%	‡	85%	89%
Operations per cancers listed above detected	9·3	33·0	21·1	18·8	2·3
Number of borderlines or low malignant potential tumours	10	†	9	20	8
Outcome measure: primary invasive epithelial ovarian and tubal cancers (within 1 year of screen)					
Apparent sensitivity	76·3%	‡	‡	75·0%	89·5%
Specificity	98·7%	‡	98·4%	98·3%	99·9%
Number of operations per cancer detected	9·3	‡	21·1	35·2	2·9
% of stage I/II cancers among screen-detected cancers	82·1%	‡	22·2%	50·0%	47·1%

*Excludes primary peritoneal cancers and non-epithelial ovarian neoplasms to allow comparison with Kobayashi et al, 2008. †Cannot be calculated owing to an absence of data. ‡Absent data.

Table 7: **Comparison of outcomes of different ovarian cancer screening strategies in recent general population trials**

Differences between the tests themselves could also lead to differences in the detection of primary invasive epithelial cancers. Serum CA125 is a highly reproducible assay, and in this trial all measurements were done in one central accredited laboratory, which was subject to external quality control. By contrast, ultrasound has a subjective element, and accuracy is correlated with the experience of the sonographer.[30,31] There are no described quality-assurance measures for scanning postmenopausal ovaries in asymptomatic women. Measures have been developed in the course of the trial based on visualisation rates and ovarian volume, and validation of these measures is underway. More than 100 sonographers were required to deliver the scan load of 55 000 scans per year; these individuals were fully certified staff working in ultrasound departments in the UK NHS. They were required to undertake regular training, audit, and feedback as part of the trial. However, there will be a degree of heterogeneity between sonographers given the size of the trial. This is being analysed and will be reported elsewhere. Until then, it should be noted that the number of interval cancers and sensitivity in the USS group is in keeping with other large single-centre and multicentre series (table 7) for which systematic follow-up and tracing of interval cancers has been undertaken. In the MMS group, all scans were undertaken by very experienced type 2 specialists.

The higher overall rates of surgery in the USS group primarily reflect the high prevalence of benign adnexal lesions in postmenopausal women. A previous ultrasound and autopsy study found that 15·4% of 234 postmenopausal women who had died from non-gynaecological diseases had ovarian cysts.[32] Here, 1894 (3·9%) of women were found to have abnormal scans. Clinical assessment in the USS group, which included the use of morphological features detected during ultrasound, serum CA125, and other imaging modalities, decreased surgical rates to 44·6% (845 of 1894) of those found to have abnormalities compared with 58·1% (97 of 167) of those with abnormalities in the MMS group (table 2). The lack of follow-up data on the outcome of pelvic masses with benign ultrasound morphology[33] means that a proportion of women and clinicians will opt for surgery once a complex adnexal lesion is detected, even if it is more likely to be benign. This is exemplified in this series, in which a lesion detected on ultrasound was not removed on laparoscopy because it was thought to be an ovarian fibroma (table 2). On follow-up 22 months later it had increased in size and a stage IC papillary serous cystadenocarcinoma was diagnosed at surgery. The higher proportion of laparoscopic procedures in the USS group than in the MMS group (62·4% *vs* 28·6%) reflects the lower suspicion of malignancy among clinicians for certain ultrasound-detected lesions. However, such decisions are not straightforward, as the risk of malignancy associated with lesions such as multilocular cysts in clinical series is 18%, rising to 49% if a solid component is also present.[29] During further rounds of screening there is likely to be a substantial fall in the number of women undergoing surgery for benign lesions in the USS group, as most will have been removed or detected and managed conservatively during the prevalence screen. It is therefore important to wait for the results of incidence screening before drawing definite conclusions about the positive-predictive value and specificity characteristics of the two screening strategies. In the ultrasound screening trial by Van Nagell and colleagues,[9] the number of operations per case of invasive epithelial ovarian cancer detected was 9·3:1 (table 7) after a mean of 4·8 annual screens. 2·9% of women undergoing surgery which resulted in benign pathology or normal ovaries being detected, experienced a major complication involving injury to a hollow viscus or significant haemorrhage. There was no difference in the complication rates in the two screening groups.

The proportion of primary invasive ovarian and fallopian tube cancers diagnosed with stage I/II disease (48%) was encouraging compared with the 26% rate in the clinical series[34] and 22% in the prevalence screen of the PLCO screening trial in the USA (table 7).[23] The highest reported proportion of early-stage cancers detected on screening is from the University of Kentucky screening programme,[9] where 82% of primary invasive epithelial ovarian cancers detected were stage I/II. However, these were the combined results of prevalence and incidence screening, with each woman in the study having had a mean of 4·8 scans. The overall number of ovarian cancers reported in the University of Kentucky study was also lower (table 7).[9] The effect of this apparent stage shift on mortality will not be known until sufficient events have accrued for a comparison with mortality in the control group, when the trial is completed in December, 2014. The false negatives were mostly diagnosed with stage I/II disease in the MMS group, and all with stage III/IV disease in the USS group. Numbers are too small at present to draw any meaningful conclusions, but with data from incidence screening it should be possible to investigate this further.

The differences in the uptake of the initial screen between women randomly assigned to the MMS group and USS group (figure 1) must be interpreted with care. It is important not to interpret this as indicating that women preferred a blood test to a scan, as a significant proportion of this difference reflects trial design. At recruitment, all women donated a blood sample. In women randomly assigned to the MMS group, this was assayed for CA125. However, women randomly assigned to the USS group had to attend again for their scan, and therefore had an opportunity to withdraw. Analysis of the psychosocial data and compliance with annual screening will provide better measures of womens' preferences.

A limitation of trial design could be that the criteria used to classify scans did not incorporate one of the many weighted morphological indices that have been

proposed to improve discrimination between benign and malignant masses.[29,35] However, it is important to note that most of these indices were derived from clinical series in symptomatic patients, in whom advanced cancers are the norm. Data on morphological characteristics of early ovarian cancers in asymptomatic patients and outcome on long-term follow-up of ultrasound-detected lesions are limited. Simple unilocular cysts are an exception, for which long-term follow-up of women has shown that they are invariably benign.[36] This was incorporated into the UKCTOCS trial design, with simple cysts less than 60 cm^3 in size classified as normal and the women returned to annual screening with no further review. In the absence of robust data on the long-term outcomes of all other complex lesions, any woman with a complex lesion was assessed individually. All clinicians were provided with pictorial depictions of complex morphology, initially using the Kentucky format[19] until the adoption in 2003 of the IOTA format,[13] and all clinicians were aware of the risk of malignancy associated with the features in the clinical series.

Another issue often raised is the current relevance of the screening strategies, which were designed in 1999. Serum CA125 and transvaginal ultrasound remain at the core of all new screening and diagnostic strategies being proposed for ovarian cancer, and although many new markers have been discovered since 1999, none have so far been validated in a prospective screening trial. However, it is hoped that this will change in the next few years. The trial serum bank, which currently has over 350 000 samples, will make the retrospective testing of new markers possible.

A final limitation of this report is that no data are available on cancers in the control group. However, in line with other ovarian cancer screening randomised controlled trials,[23] it was felt by the overseeing committees that the release of such information when screening is ongoing would compromise the overall outcome of the trial. These data should be available soon after screening is completed in 2011, ahead of the mortality analysis, which will require follow up until 2014.

Women were invited from the age and sex registers of participating Primary Care Trusts. Although this maximises external validity by excluding biases related to advertisement and self-referral, the cohort is still likely to be healthier than the general UK population because of the characteristics of the women who chose to respond. The healthy-volunteer effect is likely to result in a lower incidence of ovarian cancer deaths in the study population than in the entire UK population. However, there is no reason to believe that the ovarian cancers that occur in this cohort would be any different to those in the general population. The multicentre design, NHS hospital setting, use of standard tests (CA125 and transvaginal ultrasound) done by staff similar to those who would deliver a national programme, and the management of women with abnormalities in NHS clinics using national guidelines ensures that the findings of the trial are applicable to the wider population.

The results of the prevalence screen of the UKCTOCS show that both screening strategies are feasible. There are inherent differences between the two strategies being tested, with a more subjective element inherent in the ultrasound-based strategy than with the CA125 test, for which it is easy to implement stringent quality control. However, both screening strategies have encouraging performance characteristics. MMS has significantly better specificity than does USS, resulting in fewer repeat tests and less surgery; sensitivity for the detection of primary epithelial cancers of the ovaries and fallopian tubes seems better with MMS than with USS, although the difference is not statistically significant. Overdiagnosis of borderline cancers seems to be less of a problem with MMS than with USS. Analysis of the psychosocial effect and cost-effectiveness of these strategies is currently underway. The results of ongoing screening are required before a conclusion can be drawn regarding the effect of screening on mortality.

Contributors

IJ, UM, MP, SJS, SC, LF, and AM were involved in study design. All authors were involved in running the trial. UM, AGM, and RH undertook the literature search for this manuscript. UM, AGM, RH, and AR drafted the manuscript. UM, AGM, RH, AR, and MB prepared the figures and tables. UM, MB, and AGM did the statistical analysis. All authors critically revised the manuscript and approved the final version. UM is the guarantor.

Conflicts of interest

IJ has consultancy arrangements with Vermillion and Becton Dickinson, both of which have an interest in tumour markers and ovarian cancer. They have provided consulting fees, funds for research, and staff, but none of which were directly related to this study. SJS has received research support from Fujirebio Diagnostics, but not in relation to this trial. IJ and SJS are co-inventors of the risk of ovarian cancer algorithm, patent number 5800347. None of the other authors declared any conflicts of interest.

Acknowledgments

We are particularly grateful to the women throughout the UK who are participating in the trial, to the entire medical, nursing, and administrative staff who work on the UKCTOCS and to the independent members of the numerous oversight committees. The trial was core funded by the Medical Research Council, Cancer Research UK, and the Department of Health, with additional support from the Eve Appeal, Special Trustees of Bart's and the London, and Special Trustees of University College London Hospital (UCLH). A large portion of this work was done at UCLH/UCL within the "women's health theme" of the National Institute for Health Research UCLH/UCL Comprehensive Biomedical Research Centre supported by the Department of Health. SS has received research support from the National Cancer Institute (grant numbers CA086381 and CA083639). The researchers are independent from the funders.

References

1 Office for National Statistics. Mortality statistics cause: review of the Registrar General on deaths by cause, sex and age, England and Wales. Series DH2 32 2005; 275.

2 Berrino F, De Angelis R, Sant M, et al. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995–99: results of the EUROCARE-4 study. *Lancet Oncol* 2007; **8:** 773–83.

3 Stearns AT, Hole D, George WD, Kingsmore DB. Comparison of breast cancer mortality rates with those of ovarian and colorectal carcinoma. *Br J Surg* 2007; **94:** 957–65.

4 Engel J, Eckel R, Schubert-Fritschle G, et al. Moderate progress for ovarian cancer in the last 20 years: prolongation of survival, but no improvement in the cure rate. *Eur J Cancer* 2002; **38:** 2435–45.

5 CRUK. UK Ovarian cancer survival statistics. http://info.cancerresearchuk.org/cancerstats/types/ovary/survival/ (accessed Nov 15, 2008).

6 Jacobs I, Stabile I, Bridges J, et al. Multimodal approach to screening for ovarian cancer. *Lancet* 1988; **1:** 268–71.

7 Campbell S, Bhan V, Royston P, Whitehead MI, Collins WP. Transabdominal ultrasound screening for early ovarian cancer. *BMJ* 1989; **299:** 1363–67.

8 Jacobs IJ, Skates SJ, MacDonald N, et al. Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 1999; **353:** 1207–10.

9 van Nagell JR Jr, DePriest PD, Ueland FR, et al. Ovarian cancer screening with annual transvaginal sonography: findings of 25,000 women screened. *Cancer* 2007; **109:** 1887–96.

10 DePriest PD, van Nagell JR Jr. Transvaginal ultrasound screening for ovarian cancer. *Clin Obstet Gynecol* 1992; **35:** 40–44.

11 DePriest PD, Shenson D, Fried A, et al. A morphology index based on sonographic findings in ovarian cancer. *Gynecol Oncol* 1993; **51:** 7–11.

12 Menon U, Talaat A, Jeyarajah AR, et al. Ultrasound assessment of ovarian cancer risk in postmenopausal women with CA125 elevation. *Br J Cancer* 1999; **80:** 1644–47.

13 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:** 500–05.

14 Skates SJ, Pauler D, Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical change point and mixture models of longitudinal markers. *J Am Stat Assoc* 2001; **96:** 429–39.

15 Skates SJ, Menon U, MacDonald N, et al. Calculation of the risk of ovarian cancer from serial CA-125 values for preclinical detection in postmenopausal women. *J Clin Oncol* 2003; **21** (10 suppl)**:** 206s–10s.

16 Menon U, Gentry-Maharaj A, Ryan A, et al. Recruitment to multicentre trials—lessons from UKCTOCS: descriptive study. *BMJ* 2008; **337:** 1283–86.

17 Menon U, Skates SJ, Lewis S, et al. Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. *J Clin Oncol* 2005; **23**: 7919–26.

18 Rufford B, Menon U, Jacobs I. Screening for familial ovarian cancer. In: Morrison PJ, Hodgson SV, Haites NE, eds. Familial breast and ovarian cancer—genetics, screening and management. Cambridge: Cambridge University Press, 2002: 220–33.

19 DePriest PD, Varner E, Powell J, et al. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. *Gynecol Oncol* 1994; **55:** 174–78.

20 Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 1989; **8:** 467–75.

21 Senn S. Testing for baseline balance in clinical trials. *Stat Med* 1994; **13:** 1715–26.

22 Kobayashi H, Yamada Y, Sado T, et al. A randomized study of screening for ovarian cancer: a multicenter study in Japan. *Int J Gynecol Cancer* 2008; **18:** 414–20.

23 Buys SS, Partridge E, Greene MH, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: findings from the initial screen of a randomized trial. *Am J Obstet Gynecol* 2005; **193:** 1630–39.

24 Skirnisdottir I, Garmo H, Wilander E, Holmberg L. Borderline ovarian tumors in Sweden 1960–2005: trends in incidence and age at diagnosis compared to ovarian cancer. *Int J Cancer* 2008; **123:** 1897–901.

25 Sherman ME, Mink PJ, Curtis R, et al. Survival among women with borderline ovarian tumors and ovarian carcinoma: a population-based analysis. *Cancer* 2004; **100:** 1045–52.

26 Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol* 2007; **8:** 1129–38.

27 Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics* 2008; **64:** 10–19.

28 Reich JM. A critical appraisal of overdiagnosis: estimates of its magnitude and implications for lung cancer screening. *Thorax* 2008; **63**: 377–83.

29 Valentin L, Ameye L, Testa A, et al. Ultrasound characteristics of different types of adnexal malignancies. *Gynecol Oncol* 2006; **102**: 41–48.

30 Timmerman D, Verrelst H, Bourne TH, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound Obstet Gynecol* 1999; **13:** 17–25.

31 Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999; **14:** 338–47.

32 Dorum A, Blom GP, Ekerhovd E, Granberg S. Prevalence and histologic diagnosis of adnexal cysts in postmenopausal women: an autopsy study. *Am J Obstet Gynecol* 2005; **192:** 48–54.

33 Valentin L. Use of morphology to characterize and manage common adnexal masses. *Best Pract Res Clin Obstet Gynaecol* 2004; **18**: 71–89.

34 Ries LAG, Eisner MP, Kosary CL, eds. SEER cancer statistics review, 1975–2002. National Cancer Institute: Bethesda, MD, USA, 2005.

35 Ueland FR, DePriest PD, Pavlik EJ, Kryscio RJ, van Nagell JR Jr. Preoperative differentiation of malignant from benign ovarian tumors: the efficacy of morphology indexing and Doppler flow sonography. *Gynecol Oncol* 2003; **91**: 46–50.

36 Bailey CL, Ueland FR, Land GL, et al. The malignant potential of small cystic ovarian tumors in women over 50 years of age. *Gynecol Oncol* 1998; **69**: 3–7.