Ultrasound Obstet Gynecol 2012; **40**: 245–254
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.12281

Editorial

Ovarian cancer: role of ultrasound in preoperative diagnosis and population screening

STUART CAMPBELL

Create Health Clinic, London, UK
(e-mail: profscampbell@hotmail.com)

Four papers in this issue of the Journal are concerned with the role of ultrasound in the diagnosis and early detection of ovarian cancer. Three of them[1–3] address the problem of the accurate diagnosis of ovarian cancer in women who have a presumed ovarian mass identified by ultrasound. Since the seminal paper by Granberg *et al.*[4] on morphological characterization of ovarian cysts by transvaginal scanning there has been an explosion of interest in this subject, with ultrasound algorithms based on morphological indices and Doppler being produced on an industrial scale. Geomini *et al.*[5] reviewed 109 studies on 83 different prediction models but did not include in their analysis any of the well-known International Ovarian Tumor Analysis (IOTA) studies which have evaluated a further 11 logistic regression models in 28 papers. There have also been models developed in non-English language journals[6] that have somehow crept under the radar. Not to mention the vast hinterland of literature concerned with modifying, evaluating and comparing these algorithms in different populations and subsets of these populations. This Editorial is an attempt to put our three new papers into some kind of context and also to address the question as to where we go from here.

Running parallel with diagnostic studies, there have been numerous papers addressing the role of ultrasound in detecting ovarian cancer in an unselected population of women, and the fourth paper in this issue[7] analyzes the risk of malignancy in masses detected during an ovarian cancer screening program in a general population of women over the age of 50 years. The aim of these screening studies is to detect ovarian cancer at an early treatable stage and reduce mortality, but, as ovarian cysts are common in postmenopausal women, ultrasound has a dual role in detecting ovarian cysts and then making an accurate diagnosis of malignancy in these cysts. Population screening studies began in the early 1980s with programs based on abdominal scanning[8] and, as there is now a large amount of data on screening by transvaginal ultrasound in healthy volunteers over the age of 50, it is appropriate at this juncture to try to evaluate the role of ultrasound in early cancer detection.

The Problem

Ovarian cancer is one of the greatest health problems in gynecology. In developed countries it is the most common genital tract malignancy, with women having a 1–2% life-time risk of developing the disease[9]. It is also the most lethal gynecological malignancy, with an overall 5-year survival of 45%[10]. For example, in the United States approximately 21 550 women develop ovarian cancer each year and 14 600 women die from the disease[10]. In Europe, the corresponding figures are 66 700 and 41 900, respectively[11]. Over 90% of ovarian cancers are sporadic and occur in the general population, mainly in women over 50 years of age. Familial predisposition has been described in 5–10% of a younger subset of women who develop ovarian cancer and most of these cases are associated with mutations in the *BRCA1, BRCA2* and *MMR* genes[12,13]. Between 80 and 85% of cancers are epithelial in origin (EOC), the most common histological subtype being serous ovarian cancer, which usually presents at advanced stages and has the poorest outcome[14]. Ovarian cancer presents late as early symptoms are often vague and the condition is usually first identified as abdominal distension, a feeling of bloatedness[15] or as an abdominal mass. Sixty per cent of women are diagnosed at an advanced stage, which has a 5-year survival as low as 10%. When the disease is diagnosed at Stage 1 (i.e. confined to the ovaries), the 5-year survival is in excess of 90%[16]. This forms the rationale for ovarian cancer screening programs, the premise being that early detection may affect long-term survival.

Recent studies on the origin and pathogenesis of ovarian cancer may have implications for the screening and diagnosis of this condition[17]. EOC presents as a heterogeneous group of tumors that can be classified on a morphologic and molecular genetic basis into two types. Type I are slow-growing cancers with good prognosis, such as low-grade serous, low-grade endometrioid, clear cell, mucinous and Brenner carcinomas and borderline tumors. They are easily detected by pelvic examination and/or transvaginal ultrasound; however, they constitute only 25% of ovarian cancers and account for approximately 10% of ovarian cancer deaths. Type-II tumors are more aggressive and include high-grade serous, high-grade endometrioid and undifferentiated tumors and carcinosarcomas. Type-II tumors represent approximately 75% of all ovarian carcinomas and are responsible for 90% of ovarian cancer deaths. They are more difficult to detect due to their rapid growth and dissemination. They display p53 mutations in over 80% of cases and rarely harbor the mutations that are found in the Type I-tumors. Recent advances in our understanding of the cell of origin of ovarian cancer may help us to explain the biological differences between Type-I and Type-II cancers.

The traditional view of ovarian carcinogenesis has been that the various different tumors arise '*de novo*' from the single layer surface epithelium (mesothelium) of the ovary and that metaplastic changes occur following proliferation to repair the defect in the damaged epithelium following ovulation[18]. This theory goes some way to explaining why the suppression of ovulation by oral contraceptives reduces the risk of developing EOC. However, recent studies[19,20] on the origin of ovarian cancer have identified a precursor *in-situ* lesion called serous intraepithelial tubal carcinoma (STIC) in the Fallopian tube that, morphologically and molecularly, resembles high-grade ovarian serous carcinoma. Thus, rather than developing *de novo* from the ovary, as previously proposed, the majority of Type-II tumors appear to arise from a STIC in the fimbriated end of the Fallopian tube. In other words, the majority of high-grade serous cancers are of Mullerian not mesothelial origin and arise in the Fallopian tube before spreading to the ovary. Implantation of fimbrial cells could also occur at the time of ovulation, so the protective effect of ovulation suppression could still be explained in this scenario.

Preoperative diagnosis of malignancy in ovarian cysts

Ovarian cysts are common in the postmenopausal woman. In the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial of women over the age of 50[21], 21% had an ovarian cyst, of which 5.5% were complex. Even without a screening program, the greater use of transvaginal ultrasound scanning for the investigation of abnormal bleeding, pelvic pain or fertility problems means that an increasing number of cysts will come to the attention of gynecologists. The risk of ovarian cancer in these cysts is low, but much unnecessary anxiety can be caused and unnecessary intervention undertaken if a wrong diagnosis is made. Following expert sonography, the majority of ovarian cysts can be managed conservatively[22] once the nature and benignity of the cyst has been confirmed. Not infrequently, however, inappropriate surgical intervention will be undertaken even for non-ovarian masses such as pedunculated fibroids, hydrosalpinges or pseudocysts of peritoneal fluid. The importance of expert transvaginal ultrasound diagnosis of adnexal masses, therefore, cannot be overestimated.

Historical perspective

Although attempts had been made to classify ovarian cysts by transabdominal scanning[23], it was Granberg *et al.*[4] who introduced the classification which is the basis for morphological assessment today. They performed transvaginal scans on 1016 adnexal masses on the day before surgery. If the mass could not be imaged completely transvaginally, an abdominal scan was performed. The ultrasound findings were compared to the histology of the tumor following its complete removal. The ultrasound classification was unilocular, unilocular solid, multilocular, multilocular solid or solid cyst. Papillae were graded according to their number as 0 (none), 1 (one to five) or 2 (more than five). The sensitivity for malignancy was 82% and the specificity 92% in their study cohort. Subsequently, several prediction algorithms with weighted gray-scale scoring systems were developed[24–27]. Ferrazzi *et al.*[27] introduced the concept of removing dermoid cysts (which could easily be recognized subjectively) from the algorithm and giving a corpus luteum a score equivalent to that of an anechoic cyst to reduce the false-positive rate. Two non-gray-scale modifications were also introduced into the scoring of adnexal masses in an attempt to improve performance. In 1989, Bourne *et al.*[28] introduced the concept of using visualization of color flow and spectral Doppler (pulsatility index, PI) of tumor vessels as a second-stage test to improve the specificity of predictions. In 1990, Jacobs *et al.*[29] introduced CA 125 into the prediction model, describing the Risk of Malignancy Index (RMI), which is the product of an ultrasound morphology score, CA 125 level and menopausal status. Perhaps because scans were performed transabdominally, the ultrasound score was simple, with a score of 1 for multilocular cyst and 3 for cysts with solid areas, metastasis, ascites and bilaterality. Menopausal status scored 1 for pre- and 3 for postmenopausal status. The score was developed on 143 adnexal masses. Using an RMI cut-off level of 200, the sensitivity for ovarian cancer was 85% and the specificity was 97%. The index was validated prospectively on a further 124 masses[30]. This index had the advantage of simplicity and became very popular, being recommended by the Royal College of Obstetricians and Gynaecologists (RCOG)[31] for triaging postmenopausal cysts. They designated RMI < 25 as low risk (risk of cancer < 3%), RMI = 25–250 as moderate risk (risk of cancer 20%) and RMI > 250 as high risk (risk of cancer 75%). Although modifications of the RMI have been made[32,33], the original Jacobs's score is the one most commonly employed. In the mid 1990s, logistic regression modelling was used to determine the optimal ultrasound variables for the prediction of malignancy. Tailor *et al.*[34] included 10 gray-scale and Doppler variables and only age, papillary projection score and time-averaged maximum velocity contributed significantly to the prediction of malignancy. Both training sets and test sets were constructed and, at a cut-off of 25%, the sensitivity and specificity were 93% and 90%, respectively. Timmerman *et al.*[35], using a different logistic regression model, found that a visual estimate of color score, papillary projection score and postmenopausal score contributed best to the prediction of malignancy; at a cut-off of 25% the corresponding figures for sensitivity and specificity were 96% and 87%. In an external validation of these two models in a Swedish population, Valentin *et al.*[36] found that the logistic regression models were comparable but that the sensitivities and specificities were lower than in the original studies. Valentin *et al.* conjectured that this could have been due to differences in the tumor populations. What was revealing was that subjective assessment of the masses by an expert, by means

of gray-scale pattern recognition, was superior to both models. In the Geomini *et al.*[5] analysis of 83 models that had been subjected to external validation, the RMI had the highest accuracy and, because of its simplicity was their recommended prediction algorithm. The main criticism of the models tested by Geomini *et al.* was the small number of tumors in each study; 109 studies were analyzed and 21 750 tumors evaluated, giving an average of just under 200 tumors per study. Furthermore, there was frequently a wide difference in the spectrum of tumors assessed between the original algorithm-generating studies and the validation studies carried out in different centers, and definitions of the morphological features were frequently inconsistent.

International Ovarian Tumor Analysis (IOTA) group studies

To overcome the problems outlined above, Dirk Timmerman from Leuven set up the IOTA group of experts with the aim of studying over 1000 patients with persistent adnexal masses and of providing rigorous and systematic statistical analysis of the results. The international group consisted of nine centers, with the majority of the tumors coming from Leuven and the center of Lil Valentin in Malmo. The first paper[37] established the definitions and qualitative and quantitative end-points to describe both morphological and Doppler features of adnexal masses; in other words, what to record, how to define and where to measure. In 2005[38], the group published two logistic regression models, M1 and M2 (later renamed LR1 and LR2), which were based on 1066 patients with persistent tumors, all of which were surgically removed within 12 weeks of the ultrasound examination. Only one mass, i.e. the most complex, was included for each patient. The examinations were performed by an expert and over 40 clinical and ultrasound variables (morphological and Doppler) were included in the analysis. CA 125 was not an essential requirement and was performed in about half the cases. The mean age of the cohort was 47 years, with 60% of women being premenopausal and 40% postmenopausal. Model LR1 included the full set of 12 variables that were independently predictive following stepwise multivariate regression analysis: (1) personal history of ovarian cancer; (2) current hormonal therapy; (3) patient age (years); (4) maximum diameter of lesion (mm); (5) pain during the examination; (6) ascites; (7) blood flow within a solid papillary projection; (8) a purely solid tumor; (9) maximum diameter of solid component (in mm, but with no increase > 50 mm); (10) irregular internal cyst walls; (11) acoustic shadows; and (12) color score (1, 2, 3 or 4). The simpler model, LR2, employed six variables: 3, 6, 7, 9, 10 and 11 from above. In Phase 1 of the study, the area under the receiver–operating characteristics (ROC) curve (AUC) for LR1 on the test set was 0.94, giving a sensitivity of 93% and a specificity of 76%. This compared unfavorably with the sensitivity of 96% and specificity of 90% for pattern recognition by an expert but compared favorably with the corresponding values of 86% and 80% by a less experienced operator which had been established previously[39]. The performance of LR1 was significantly better than that of LR2 (AUC, 0.92) and the RMI (AUC, 0.87). In Phase 2 of the IOTA series of publications, both LR1 and LR2 models underwent extensive validation, both temporal (941 patients from seven original centers) and external (997 patients from 12 new centers)[40]. Comparisons were made by AUC, and those established in the original study for both models were maintained. A further significant innovation in the IOTA analysis was the introduction of the simple ultrasound-based rules[41], which were recently validated prospectively on 1938 patients with an adnexal mass[42]. Adnexal masses were triaged by morphological characteristics and Doppler into five features that conclusively indicated malignancy (M features) and five that indicated benignity (B features). When there was a mixture of features resulting in an inconclusive result, subjective assessment was undertaken. The M and B rules gave similar results to pattern recognition in the 77% of masses with a certain diagnosis. In the 23% of masses that were inconclusive, only pattern recognition by an expert achieved good results. The LR1 and LR2 models performed poorly in the inconclusive group, with very low specificities. Interestingly, simple rules had a higher number of inconclusive diagnoses among the postmenopausal women (28%). Although overall the RMI was inferior, it performed much better in the postmenopausal patients, with better specificities than LR1 and LR2, and may have achieved similar sensitivities if a cut-off of 250 had been chosen.

The IOTA project has brought many benefits to the subject of classification of adnexal masses, not least of which is its bringing to a stop the algorithm mania that has dominated gynecological scanning since Granberg *et al.*'s seminal paper. Essentially, the IOTA group are generating logistical models that we can have on our computer to help us differentiate between benign and malignant tumors. The two logistic models LR1 and LR2 have undergone rigorous validation within the IOTA centers and their performance has held up well. In this issue of the Journal, Nunes *et al.*[1] report on an external validation of the LR2 model in a non-IOTA center, carried out by a Level-II (i.e. well trained but non-expert) sonographer. The accuracy in terms of AUC was similar to that of the original report; the sensitivity was higher (97%) but the specificity was lower (69%), so, although the sonographer was instructed not to evaluate the tumors subjectively, there may have been an element of a spectrum bias effect due to the high incidence (53%) of malignancy in the cohort. As one of the purposes of the IOTA algorithms is to enhance the performance of the non-expert sonographer, this is reassuring, as high sensitivity is essential in a diagnostic test for cancer. There has been a recent tendency[43] to test only the LR2 model, which is surprising when LR1 consistently outperforms LR2 in triaging persistent tumors before surgery. If the model is on our computer, it would take only a few

minutes to record the additional variables, such as the overall color score, and would encourage the examiner to take time to examine the adnexal mass in detail.

The aim of these models is to place less reliance on subjective assessment by an expert, as it is argued that such an individual is not always around in a community hospital setting. The human brain is uniquely adapted to perform pattern recognition and so it is no surprise that subjective assessment by an expert has consistently been the most successful method of discriminating between benign and malignant ovarian tumors. Nevertheless, both models in the validation studies came close to matching pattern recognition and in the foreseeable future it may be mandatory to put variables into a computer program and get a risk score when assessing pelvic masses. Validation of the IOTA models has been extensive[40] and in the most recent study[44] comparisons are made between 11 IOTA models and 12 non-IOTA models in the 12 new centers previously described[40]. The centers, however, were recruited by the IOTA group and the variables collected were for the IOTA models, necessitating statistical devices such as multiple imputation to address the problem of missing data in the non-IOTA models. This may partly explain why, when using the diagnostic odds ratio for comparison, the IOTA models were unrealistically three times better than the non-IOTA algorithms. Nevertheless, on the basis of the available evidence, we should assume that LR1 and LR2 are the best current ultrasound models for preoperative triage.

The IOTA team have been scrupulous in stating that their algorithms are to be used to triage women for surgery once a decision to operate has been made. This is because all the current algorithms, including theirs, are based on women with ovarian masses who were scheduled for surgery. In the IOTA studies, an expert had already made the subjective decision, based on a prior scan, to schedule the patient for surgery. We have no information as to what percentage of women with masses was triaged for expectant management or indeed as to the outcomes of this group of women. To some extent, this could limit the applicability of the algorithms. No-one disputes that it is important to have an appropriate surgeon to operate on malignant tumors[45]. However, if all we are doing with the logistic regression models is triaging women with pelvic masses to the most appropriate surgeon, it might be more effective if each hospital ensures that it has at least one experienced expert in transvaginal ultrasound to scan all complex adnexal tumors scheduled to be removed by surgery.

What we require are models to triage women either to surgery or to conservative management and in this regard high sensitivity is the default situation. The consequences of triaging a woman who has a malignant mass to conservative management could seriously affect her chances of survival. The LR1 and LR2 models were not created on a sequential unselected group of women with persistent pelvic masses so they may function less well, for example, in screening programs for ovarian cancer, where it is very important to reduce the number of unnecessary procedures in an aging group of women. Sixty per cent of the 1066 women used to develop the two models were premenopausal (mean age, 47 years) and tumors such as endometriomas, which accounted for 32% of benign masses[46], are much less frequent in the postmenopausal woman. Valentin *et al.*[36] correctly identified this spectrum bias effect when they stated: 'Naturally a mathematical model yields better results if it is tested in a tumor population very similar to that in which it was created rather than a dissimilar one'.

The paper by Di Legge *et al.*[3] from the IOTA group in this issue of the Journal highlights another potential problem in that small persistent adnexal masses < 4 cm in maximum diameter had both a lower rate of invasive malignancy and a lower detection rate by the IOTA models even though 71% of these lesions were Stage 3. The authors suggest that the models may be suboptimal for diagnosing small-volume malignancies. This is especially relevant when screening postmenopausal women, in whom the aim is to detect the cancer at an early stage.

In their recent validation study[44], the IOTA group found that their models were superior to other models in identifying Stage 1 cancers. Clearly, it is important to triage all such cancers, including borderline tumors, to a specialist oncology team which is best placed to decide whether conservative or radical surgery is required. Indeed, all borderline and most low-grade Stage-1a and Stage-1b epithelial ovarian cancers can be treated by simple oophorectomy without adjunctive chemotherapy when, for example, fertility sparing surgery is required[47]. Apart from a study performed on the original database[48], little information has been provided in the IOTA studies on the histology of the malignant tumors or regarding whether they were Type I or II. This early study[48] showed that the incidence of borderline tumors in the original database was high compared to that found in screening studies[49] and that only one third of early-stage invasive cancers were of the serous type. Indeed both the borderline and Stage 1 cancers were of larger size than the more advanced invasive cancers implying that the removal of most of the early stage cancers detected by ultrasound, while important for the clinical management of the patient, may not provide a significant mortality benefit.

The simple rules triage model by the IOTA group is an attempt to develop a system that could be used to select women for conservative management, but so far studies have been performed on a group already selected for surgery, so it is in a 'double triage' phase. An external validation of the simple rules by Hartman *et al.*[2] in this issue was carried out on 110 masses in women scheduled for surgery. The sensitivity and specificity in the inconclusive group did not match that of the IOTA study, but was improved significantly by the addition of CA 125 at a 37.5 U/mL fixed cut-off value.

The role of CA 125 in the discrimination of adnexal tumors has become a contentious issue. The IOTA group calculated that a fixed CA 125 value of 30 U/mL did not

improve their model[50]. In a recent retrospective analysis[43] they found that the LR2 model was superior to the RMI in categorizing both benign and malignant tumors, and recommended that the RCOG abandon their espousal of the RMI for classification of postmenopausal adnexal masses. While the RMI has possibly reached its 'sell-by date', I do not think it should be abandoned just yet. The performance of the RMI appears to improve on a postmenopausal population[43] and has the advantage of having the actual CA 125 level rather than a fixed level in the equation. Some of the advantages of the IOTA model in this study may be due in part to the inclusion of low-grade malignancies and borderline tumors in the high-risk group, and the effect of missing CA 125 values is difficult to assess. It is also now apparent that a rise in serial CA 125 measurements is better than a fixed cut-off in identifying ovarian cancer[51] and this should be considered when the nature of a persistent mass is in doubt. We now need logistic regression models to be constructed for adnexal tumors in an unselected postmenopausal population of women with and without the addition of serial CA 125 measurements to determine the best means of achieving maximum sensitivity in the diagnosis of ovarian cancer. In the meantime, however, while advocating the use of the IOTA models, I believe that CA 125 measurements should also be performed when a persistent complex adnexal mass is detected in the postmenopausal woman, especially if a decision has to be made as to whether to operate or continue surveillance.

Screening for ovarian cancer

The premise that early detection may affect long-term survival forms the rationale for ovarian cancer screening programs. As 85% of ovarian cancers occur in postmenopausal women, the concept of screening this population to detect cancer at an early treatable stage seems logical. Although a number of studies have evaluated screening for familial ovarian cancer, the sensitivity and effectiveness of screening in the younger high-risk population has yet to be established[52] and this Editorial will only consider studies on the general population. Demonstrating a stage shift towards detection of EOC at Stage 1/2 is not sufficient to imply an improvement in rates of mortality from this condition. Lead and length time biases can give a false impression that screening is working. Therefore, to prove a mortality benefit in terms of 5-year survival that is attributable to screening, an unscreened control group is required and numbers have to be sufficiently large to show an effect. Also, national cancer mortality registries have to be up to the task of recording the cause of all deaths accurately so that ascertainment is rigorous. Ultrasound screening is a huge logistical exercise, although the difficulty of the task has sometimes been exaggerated. After all, in most countries two routine screening scans are performed on all pregnant women and this does not appear to be problematic. The ovarian screening studies that have been or are being performed recruit volunteers, which means the population is biased towards the motivated and middle class[53], but, as EOC is not a cancer of the socially deprived, this may not be a significant bias. It is probable that the Japanese Shizuoka Cohort Study[54] and the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)[55] come closest to the ideal in recruitment, in that all women were contacted through their local population registers and offered a chance to participate. Scans in recent trials were usually performed by sonographers who record details of ovarian size and morphology and also Doppler indices if required. Ovaries were visualized in both longitudinal and transverse planes and the volume calculated by the prolate ellipsoid formula (length $\times$ width $\times$ depth $\times$ 0.5233). Any abnormal morphology was scored according to protocol and a decision to operate made by the appropriate clinician. In most studies this was left to the surgeon's clinical judgment, but in the UKCTOCS study there was a management plan that specified indications for surgery.

The normal postmenopausal ovary has a range in volume of about 1–2 mL and cannot always be identified because of shadowing by bowel, fibroids or other factors. In the Kentucky screening study[56] at least one ovary was not seen in 16% of cases. In most studies, failure to visualize the ovary is regarded as a negative screen. In the UKCTOCS study, in order to minimize the chance of missing an abnormal ovary, sonographers were requested to demonstrate a 3-cm length of a clearly defined iliac vein in the pelvic side wall if the ovary was not visualized. A similar protocol was used in the PLCO Screening Trial[57] and a minimum time of 5 min was spent to identify each ovary.

The prevalence of ovarian cancer in the postmenopausal population is 1 in 2500, which makes population screening a challenge. A high sensitivity of $> 75\%$ is required with a significant shift towards Stage 1 disease to make screening worthwhile, but, most importantly, the specificity must be very high ($> 99.6\%$) to give a positive predictive value of 10%, i.e. a maximum of 10 operations per cancer detected. These figures should be borne in mind when assessing the efficacy of a screening program.

Historical perspective

The first ovarian cancer screening study[8] was carried out over a 5-year period in the 1980s on over 5000 volunteers using transabdominal scanning which required the woman to fill her bladder. Despite this, compliance was good. All five primary cancers were detected, most of which were borderline. In addition, four metastatic cancers were found. The study demonstrated that screening by ultrasound was feasible. The first prospective ovarian cancer screening study using CA 125[58] involved 5550 Swedish women aged 50 years and older. Of the 175 women with high CA 125 levels, six were found to have ovarian cancer: two each in Stages 1A, 2B and 3C. Three ovarian cancers were missed. This study showed that screening with CA 125 was feasible, but subsequent studies showed that CA 125 at a fixed cut-off of 35 U/mL

had a low sensitivity for Stage 1 disease[59]. In order to improve sensitivity, Skates *et al.*[51] introduced a more sophisticated approach by rejecting a fixed cut-off CA 125 level and taking into account the serial values that are available in the screening context. They demonstrated that elevated CA 125 levels in women without ovarian cancer had a flat or static profile or decreased with time, whereas levels associated with malignancy tended to rise. This led to the development of the ROC algorithm which estimates a woman's risk of ovarian cancer based on the rise in CA 125 and allows women to be triaged into low-, intermediate- or high-risk categories. It is important to realize that, for example, a rise in value from 8 to 16 U/mL (i.e. a value which would usually be regarded as normal) over a period of 3 months could put a woman in the high-risk category. Jacobs *et al.*[60] then introduced the ROC algorithm into a randomized controlled screening study using transvaginal ultrasound to visualize the ovaries of women in the high-risk group in order to improve specificity. This was called multimodal screening. The trial of 22 000 postmenopausal women showed a significantly increased median survival in women who developed ovarian cancer in the screened group compared with the control group. These results prompted the UKCTOCS multicenter trial which is discussed below.

Recent studies

There are now four large ongoing or recently completed trials on ovarian cancer screening by means of transvaginal scanning and CA 125 that have published data in the last decade:

1. *The University of Kentucky ovarian cancer screening trial*[56] is a single-arm (i.e. uncontrolled) annual ultrasound screening study of 25 327 volunteers over a period of 9 years, in which 120 569 scans (mean, 4.8 per participant) were performed. An ovarian volume > 20 mL (premenopause) or > 10 mL (postmenopause) or any cystic ovarian tumor with a solid or papillary projection into its lumen was considered abnormal. The mean age of the cohort was 55 years. The reported sensitivity for primary EOC was 81%, with 9.3 operations carried out per case detected. When restricted to primary invasive ovarian cancer, the sensitivity decreased to 76.3%. Most (82%) of the primary ovarian cancers were early stage (Stage 1/2). Serum CA 125 levels were increased (> 35 U/mL) at the time of detection in 13 of 15 (87%) patients who had Stage 3 EOC but in only three of 15 (20%) patients who had Stage 1 or 2 disease. At a mean follow up of 5.8 years, the women in the trial had a significantly longer 5-year survival ($74.8 \pm 6.6\%$) compared to the women from the same institution, treated by the same surgical and chemotherapeutic protocols, who were not screened ($53.7 \pm 2.3\%$)[61].
2. *The Japanese Shizuoka Cohort Study of Ovarian Cancer Screening*[54] is a randomized controlled trial of 82 487 low-risk postmenopausal women from 212 hospitals in 35 townships carried out over a 15-year period. Women with a median age of 58 years were screened by annual transvaginal ultrasound exam and CA 125 using a cut-off of 35 U/mL. The mean number of screens per woman was 5.4; the uptake of screening fell from 82% to 56% from the second to the fifth screen. Abnormal ovarian morphology was classified as simple cyst (single, thin walled, anechoic cyst with no septa or papillary projections) or complex cyst (abnormal ovarian morphology other than simple cyst). The screening strategy achieved a sensitivity for malignancy of 77.1% and a specificity of 99.9%. The proportion of Stage-1 ovarian cancer was higher in the screened group (63%) than in the control group (38%) but the difference was not statistically significant. The effect on mortality has not yet been reported.
3. *The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)*[57]. This is a randomized controlled trial of 78 216 women aged 55 to 74 years assigned to undergo either annual screening or usual care at 10 screening centers across the United States between November 1993 and July 2001. Women were screened by serum CA 125, using a cut-off of 35 U/mL, and transvaginal ultrasonography for 3 years, followed by CA 125 alone for a further 2 years. The following transvaginal ultrasound results were classified as abnormal: (1) ovarian volume greater than 10 mL; (2) cyst volume greater than 10 mL; (3) any solid area or papillary projection extending into the cavity of a cystic ovarian tumor of any size; and (4) any mixed (solid and cystic) component within a cystic ovarian tumor. Evaluation and management of positive screening tests was at the discretion of the participant's clinician. Women were followed up for a median of 12.4 years. During four rounds of incidence screening[62], 89 invasive ovarian or peritoneal cancers were diagnosed, of which 60 were detected by screening (sensitivity of 68.2%), with 13 surgeries carried out per case of ovarian cancer. A total of 72% of the screen-detected cancers were late stage (Stage 3/4). Recently, mortality data have been reported[57]. A total of 212 women had a screen-detected cancer in the intervention arm and 176 were identified in the control arm. The screening and control arms included 118 and 100 deaths, respectively, with a mortality rate ratio of 1.18. These data showed that simultaneous screening with CA 125 using an absolute cut-off and transvaginal scanning did not reduce mortality from the disease. Moreover, the excess morbidity of carrying out surgery in women with false-positive results was 5.1%.
4. *The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)*[63]. In this trial, 202 638 postmenopausal women aged 50–74 years were randomized to either control or annual screening with ultrasound or a multimodal strategy in a 2 : 1 : 1 fashion. In the multimodal group, CA 125 was interpreted using the ROC algorithm to triage the

women into low, intermediate and elevated risk. Those at intermediate risk had a repeat CA 125 in 12 weeks, whereas those with elevated risk were referred for a transvaginal scan and repeat CA 125 in 6 weeks. In the prevalence screen[63], 91% were classified as low risk by the ROC algorithm and returned to annual screening. Only 9% of women required a repeat CA 125 test and an ultrasound scan and 0.2% had surgery. The study commenced in 2001 and final results are expected in 2015. However, the results from the prevalence screen in the multimodal arm suggest that the multimodal strategy has superior sensitivity (89.4%) and specificity (99.8%) to ultrasound screening alone (sensitivity, 84.9%; specificity, 98.2%) for primary ovarian cancer. When restricted to the detection of primary invasive ovarian cancers, the sensitivity of multimodal screening was maintained at 89%, whereas the sensitivity of the ultrasound-based strategy decreased to 75.0%[63].

The aim of early detection of ovarian cancer is to save lives. There is no evidence at the moment that serial scanning of postmenopausal women will reduce mortality from ovarian cancer. The PLCO study[57] found no difference in mortality between screened and control groups and in fact found a significant increase in morbidity in the screened group, mainly from unnecessary surgery due to false-positive diagnoses by ultrasound. The absence of a stage shift in the screen arm towards earlier diagnosis is surprising for there is evidence of such a shift in the other screening studies. In the Kentucky study, 63% of cancers were at Stage 1 and in the Japanese and UKCTOCS studies the figures for Stage 1/2 cancers were 85% and 50%, respectively. If absence of benefit on mortality rate from the PLCO study is confirmed by the Japanese and UKCTOCS studies then a reason has to be found as to why the early detection of ovarian cancers does not save lives. A possible explanation is length time bias, which means that the screening test tends to detect slow-growing tumors more often than the more lethal fast-growing serous epithelial cancers. The paper by Sharma *et al.*[7] from the UKCTOCS group in this issue of the Journal suggests that this may be true. This study investigated the cysts detected in the ultrasound arm of the UKCTOCS study and found that detection of the more indolent Type-I tumors occurred twice as often as did the Type-II tumors whereas if ultrasound was equally effective across the spectrum of malignancies this ratio would be almost exactly reversed. Also, nearly all screening studies show that the detection of borderline tumors is disproportionately high in relation to other EOCs; for example, in the UKCTOCS study in the prevalence screen[63] there were 20 borderline cancers detected in the ultrasound arm but only eight in the multimodal group. The detection of these tumors is clearly advantageous for the individual woman but will have no effect in terms of reducing ovarian cancer mortality rates and therefore would not justify a screening program. This may be rather disappointing for ultrasound enthusiasts but it may not be all bad news. There are some weaknesses in the PLCO study[64]. One of these is that there was a fixed cut-off of 35 U/mL for CA 125, which is generally accepted as being too high a threshold for early ovarian cancer detection. In the UKCTOCS study, serial CA 125 measurements were assessed by the ROC algorithm and the results in the multimodal arm for the prevalence screen were strikingly good, with 89% sensitivity and 99.8% specificity for invasive EOC (i.e. one cancer found for 2.9 operations). This still gives some hope that, at least in the multimodal arm, screening may be a life-saver and that a rise in serial CA 125 levels below the standard 35 U/mL cut-off may at least detect a significant number of Type-II tumors. We will find this out when the mortality data are available from the screened and control groups. In the meantime, it seems likely that ultrasound as the primary screening test will not be effective in reducing mortality but will still retain its pivotal role in reducing the false-positive rate in the multimodal arm.

Future developments

At the moment, whether made subjectively by an expert or by means of an equation based on morphological criteria, the preoperative diagnosis of malignancy in persistent adnexal masses by transvaginal ultrasound is an inexact science. At best, about 15% of benign cysts and 10% of malignant masses will be miscategorized. Because subjective assessment by an expert using pattern recognition appears to be the best method of distinguishing between benign and malignant tumors, attempts have been made to use computational systems based on machine-learning techniques, such as artificial neural networks and support vector machines, but these do not appear to confer additional discriminating power[65]. There is also no evidence that any other imaging modality will perform any better than does ultrasound for this purpose[66]. It seems extraordinary that CA 125, which was first described over 30 years ago[67] remains pre-eminent as a biological marker for ovarian malignancy. With recent progress in genomics and proteomics[68] it appears likely that better markers will appear, but although a number of studies have shown improved performance with a panel of ovarian cancer biomarkers over CA 125 alone[69], they appear to add little to CA 125 when used in prediagnostic samples[49]. Human epididymis protein 4 (HE4) is the only new serum marker to be adopted into clinical practice and was recently shown to be effective as a second-line screen in women with an adnexal mass and elevated CA 125[70]. Cell-free DNA has also shown promise as a prognostic indicator in ovarian malignancy[71], but there is no evidence yet that it can be used as a diagnostic or screening test. So, while the search for better markers for malignancy goes on, it is important that we optimize the use of transvaginal ultrasound and CA 125 to detect Type-II ovarian cancers in asymptomatic postmenopausal women and make an accurate diagnosis of malignancy when an adnexal mass is detected.

REFERENCES

1. Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. Prospective evaluation of the IOTA logistic regression model LR2 for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2012; **40**: 355–359.
2. Hartman CA, Juliato CRT, Sarian LO, Toledo MC, Jales RM, Morais SS, Pitta DD, Marussi EF, Derchain S. Ultrasound criteria and CA 125 as predictive variables of ovarian cancer in women with adnexal tumors. *Ultrasound Obstet Gynecol* 2012; **40**: 360–366.
3. Di Legge A, Testa AC, Ameye L, Van Calster B, Lissoni AA, Leone FPG, Savelli L, Franchi D, Czekierdowski A, Trio D, Van Holsbeke C, Ferrazzi E, Scambia G, Timmerman D, Valentin L. Lesion size affects diagnostic performance of IOTA logistic regression models, IOTA simple rules and risk of malignancy index in discriminating between benign and malignant adnexal masses. *Ultrasound Obstet Gynecol* 2012; **40**: 345–354.
4. Granberg S, Norstrom A, Wikland M. Tumors in the lower pelvis as imaged by vaginal sonography. *Gynecol Oncol* 1990; **37**: 224–229.
5. Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BWJ. The accuracy of risk scores in predicting ovarian malignancy – a systematic review. *Obstet Gynecol* 2009; **113**: 384–394.
6. Weber G, Merz E, Bahlmann F, Leber AM. A new sonomorphologic scoring-system (Mainz score) for the assessment of ovarian tumors using transvaginal ultrasonography. Part II: a comparison between the scoring-system and the assessment by an experienced sonographer in postmenopausal women. *Ultraschall Med* 1999; **20**: 2–8.
7. Sharma A, Apostolidou S, Burnell M, Campbell S, Habib M, Gentry-Maharaj A, Amso N, Seif MW, Fletcher G, Singh N, Benjamin E, Brunell C, Turner G, Rangar R, Godfrey K, Oram D, Herod J, Williamson K, Jenkins H, Mould T, Woolas R, Murdoch J, Dobbs S, Leeson S, Cruickshank D, Fourkala E-O, Ryan A, Parmar M, Jacobs I, Menon U. Risk of epithelial ovarian cancer in asymptomatic women with ultrasound-detected ovarian masses: a prospective cohort study within the UK collaborative trial of ovarian cancer screening (UKCTOCS). *Ultrasound Obstet Gynecol* 2012; **40**: 338–344.
8. Campbell S, Bhan V, Royston P, Whitehead M, Collins W. Transabdominal ultrasound screening for early ovarian cancer. *Br Med J* 1989; **299**: 1363–1367.
9. Cancer Research UK. CancerStats: ovarian cancer molecular biology and genetics. http://info.cancerresearchuk.org/cancerstats/types/ovary/molecularbiologyandgenetics/[Accessed 10 July 2012].
10. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin* 2009; **59**: 225–249.
11. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer* 2010; **46**: 765–781.
12. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol* 2007; **25**: 1329–1333.
13. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomäki P, Mecklin JP, Järvinen HJ. Cancer risk in mutation carriers of DNA-mismatch-repair genes. *Int J Cancer* 1999; **81**: 214–218.
14. Seidman JD, Horkayne-Szakaly I, Haiba M, Boice CR, Kurman RJ, Ronnett BM. The histologic type and stage distribution of ovarian carcinomas of surface epithelial origin. *Int J Gynecol Pathol* 2004; **23**: 41–44.
15. Bankhead CR, Kehoe ST, Austokera J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. *BJOG* 2005; **112**: 857–865.
16. Cancer Research UK. CancerStats: ovarian cancer survival statistics. http://info.cancerresearchuk.org/cancerstats/types/ovary/survival/#stage [Accessed: 10 July 2012].
17. Kurman RJ, Shih I-M. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer – shifting the paradigm. *Hum Pathol* 2011; **42**: 918–931.
18. Scully RE. Pathology of ovarian cancer precursors. *J Cell Biochem Suppl* 1995; **23**: 208–218.
19. Piek JM, van Diest PJ, Zweemer RP, Jansen JW, Poort-Keesom RJ, Menko FH, Gille JJ, Jongsma AP, Pals G, Kenemans P, Verheijen RH. Dysplastic changes in prophylactically removed Fallopian tubes of women predisposed to developing ovarian cancer. *J Pathol* 2001; **195**: 451–456.
20. Piek JM, Verheijen RH, Kenemans P, Massuger LF, Bulten H, van Diest PJ. BRCA1/2-related ovarian cancers are of tubal origin: a hypothesis. *Gynecol Oncol* 2003; **90**: 491.
21. Hartge P, Hayes R, Reding D, Sherman ME, Prorok P, Schiffman M, Buys S. Complex ovarian cysts in postmenopausal women are not associated with ovarian cancer risk factors: preliminary data from the prostate, lung, colon, and ovarian cancer screening trial. *Am J Obstet Gynecol* 2000; **183**: 1232–1237.
22. Valentin L, Akrawi D. The natural history of adnexal cysts incidentally detected at transvaginal ultrasound examination in postmenopausal women. *Ultrasound Obstet Gynecol* 2002; **20**: 174–180.
23. Meire HB, Farrant P, Guha T. Distinction of benign from malignant ovarian cysts by ultrasound. *Br J Obstet Gynaecol* 1978; **85**: 893–899.
24. Sassone AM, Timor-Tritsch IE, Artner A, Westhoff C, Warren WB. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. *Obstet Gynecol* 1991; **78**: 70–76.
25. DePriest PD, Varner E, Powell J, Fried A, Puls L, Higgins R, Shenson D, Kryscio R, Hunter JE, Andrews SJ, van Nagell Jr JR. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. *Gynecol Oncol* 1994; **55**: 174–178.
26. Lerner JP, Timor-Tritsch IE, Federman A, Abramovich G. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. *Am J Obstet Gynecol* 1994; **170**: 81–85.
27. Ferrazzi E, Zanetta G, Dordoni D, Berlanda N, Mezzopane R, Lissoni AA. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study. *Ultrasound Obstet Gynecol* 1997; **10**: 192–197.
28. Bourne T, Campbell S, Steer CV, Whitehead MI, Collins WP. Transvaginal colour flow imaging: A possible new screening technique for ovarian cancer. *BMJ* 1989; **299**: 1367–1370.
29. 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**: 922–929.
30. Davies AP, Jacobs I, Woolas R, Fish A, Oram D. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. *Br J Obstet Gynaecol* 1993; **100**: 927–931.
31. Rufford BD, Jacobs IJ. Green-top Guideline No. 34. Ovarian cysts in postmenopausal women. Royal College of Obstetricians and Gynaecologists: London, UK, 2003.
32. Tingulstad S, Hagen B, Skjeldestad FE, Onsrud M, Kiserud T, Halvorsen T. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. *Br J Obstet Gynaecol* 1996; **103**: 826–831.
33. Tingulstad S, Hagen B, Skjeldestad FE, Halvorsen T, Nustad K, Onsrud M. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. *Obstet Gynecol* 1999; **93**: 448–452.
34. Tailor A, Jurkovic D, Bourne TH, Collins WP, Campbell S. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. *Br J Obstet Gynaecol* 1999; **106**: 21–30.
35. Timmerman D, Bourne T, Tailor A, Collins WP, Verrelst H, Vandenberghe K, Vergote I. A comparison of methods for the pre-operative discrimination between benign and malignant

adnexal masses: The development of a new logistic regression model. *Am J Obstet Gynecol* 1999; **181**: 57–65.
36. Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses. A prospective cross-validation. *Ultrasound Obstet Gynecol* 2001; **18**: 357–365.
37. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I; International Ovarian Tumor Analysis (IOTA) Group. 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**: 500–505.
38. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML. International Ovarian Tumor Analysis Group. 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**: 8794–8801.
39. Timmerman D, Schwarzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses using ultrasonography: An analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999; **13**: 11–16.
40. Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D,Lissoni AA, Van Holsbeke C, Fruscio R, Czekierdowski A, Jurkovic D, Savelli L, Vergote I, Bourne T, Van Huffel S, Valentin L. Ovarian cancer prediction in adnexal masses using ultrasound based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010; **36**: 226–234.
41. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, Paladini D, Van Calster B, Vergote I, Van Huffel S, Valentin L. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008; **31**: 681–690.
42. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, Van Holsbeke C, Savelli L, Fruscio R, Lissoni AA, Testa AC, Veldman J, Vergote I, Van Huffel S, Bourne T, Valentin L. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010; **341**: c6839. DOI: 10.1136/bmj.c6839.
43. Van Calster B, Timmerman D, Valentin L, McIndoe A, Ghaem-Maghami S, Testa AC, Vergote I, Bourne T. Triaging women with ovarian masses for surgery: observational diagnostic study to compare RCOG guidelines with an International Ovarian Tumour Analysis (IOTA) group protocol. *BJOG* 2012; **119**: 662–671.
44. Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, Fruscio R, Lissoni AA, Czekierdowski A, Savelli L, Van Huffel S, Valentin L, Timmerman D. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. *Clin Cancer Res* 2012; **18**: 815–825.
45. Woo YL, Kyrgiou M, Bryant A, Everett T, Dickinson HO. Centralisation of care may prolong survival in women with ovarian cancer, and possibly more generally, gynaecological cancer. *Cochrane Database Syst Rev* 2012; Issue 3. CD007945.
46. Van Calster B, Timmerman D, Bourne T, Testa A, Van Holsbeke C, Domali E, Jurkovic D, Neven P, Van Huffel S, Valentin L. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007; **99**: 1706–1714.
47. Park JY, Kim DY, Suh DS, Kim JH, Kim YM, Kim YT, Nam JH. Outcomes of fertility sparing surgery for invasive epithelial cancer. *Gynecol Oncol* 2008; **110**: 345–353.
48. Valentin L, Ameye L, Testa A, Lécuru F, Bernard JP, Paladini D, Van Huffel S, Timmerman D. Ultrasound characteristics of different types of adnexal malignancies. *Gynecol Oncol* 2006; **102**: 41–48.
49. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, Urban N, Mor G, Bast RC Jr, Moore LE, Lokshin AE, McIntosh MW, Skates SJ, Vitonis A, Zhang Z, Ward DC, Symanowski JT, Lomakin A, Fung ET, Sluss PM, Scholler N, Lu KH, Marrangoni AM, Patriotis C, Srivastava S, Buys SS, Berg CD; PLCO Project Team. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. *Cancer Prev Res (Phila)* 2011; **4**: 375–383.
50. Timmerman D, Van Claster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, Van Holsbeke C, Van Huffel S, Vergote I, Bourne T. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. *J Clin Oncol* 2007; **25**: 4194–4200.
51. Skates SJ, Menon U, MacDonald N, Rosenthal AN, Oram DH, Knapp RC, Jacobs IJ. 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–210s.
52. Manchanda R, Rosenthal A, Burnell M, Fraser L, Mackay J, Skates S. Change in stage distribution observed with annual screening for ovarian cancer in BRCA carriers. *J Med Genet* 2009; **46**: 423–424.
53. Burnell M, Gentry-Maharaj A, Ryan A, Apostolidou S, Habib M, Kalsi J, Skates S, Parmar M, Seif MW, Amso NN, Godfrey K, Oram D, Herod J, Williamson K, Jenkins H, Mould T, Woolas R, Murdoch J, Dobbs S, Leeson S, Cruickshank D, Campbell S, Fallowfield L, Jacobs I, Menon U. Impact on mortality and cancer incidence rates of using random invitation from population registers for recruitment to trials. *Trials* 2011; **12**: 61.
54. Kobayashi H, Yamada Y, Sado T, Sakata M, Yoshida S, Kawaguchi R, Kanayama S, Shigetomi H, Haruta S, Tsuji Y, Ueda S, Kitanaka T. A randomized study of screening for ovarian cancer: a multicenter study in Japan. *Int J Gynecol Cancer* 2008; **18**: 414–420.
55. Menon U, Skates SJ, Lewis S, Rosenthal AN, Rufford B, Sibley K, Macdonald N, Dawnay A, Jeyarajah A, Bast RC Jr, Oram D, Jacobs IJ. Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. *J Clin Oncol* 2005; **23**: 7919–7926.
56. van Nagell Jr. JR, Depriest PD, Ueland FR, DeSimone CP, Cooper AL, McDonald JM, Pavlik EJ, Kryscio RJ. Ovarian cancer screening with annual transvaginal sonography: findings of 25,000 women screened. *Cancer* 2007; **109**: 1887–1896.
57. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, Reding DJ, Greenlee RT, Yokochi LA, Kessel B, Crawford ED, Church TR, Andriole GL, Weissfeld JL, Fouad MN, Chia D, O'Brien B, Ragard LR, Clapp JD, Rathmell JM, Riley TL, Hartge P, Pinsky PF, Zhu CS, Izmirlian G, Kramer BS, Miller AB, Xu JL, Prorok PC, Gohagan JK, Berg CD; PLCO Project Team. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* 2011; **305**: 2295–2303.
58. Einhorn N, Sjovall K, Knapp RC, Hall P, Scully RE, Bast RC Jr, Zurawski VR Jr. Prospective evaluation of serum CA 125 levels for early detection of ovarian cancer. *Obstet Gynecol* 1992; **80**: 14–18.
59. Jacobs I, Bast RC Jr. The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 1989; **4**: 1–12.
60. Jacobs IJ, Skates SJ, MacDonald N, Menon U, Rosenthal AN, Davies AP, Woolas R, Jeyarajah AR, Sibley K, Lowe DG, Oram DH. Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 1999; **353**: 1207–1210.
61. van Nagell JR Jr, Miller RW, DeSimone CP, Ueland FR, Podzielinski I, Goodrich ST, Elder JW, Huang B, Kryscio RJ, Pavlik EJ. Long-term survival of women with epithelial ovarian cancer detected by ultrasonographic screening. *Obstet Gynecol* 2011; **118**: 1212–1221.
62. Partridge E, Kreimer AR, Greenlee RT, Williams C, Xu JL, Church TR, Kessel B, Johnson CC, Weissfeld JL, Isaacs C,

Andriole GL, Ogden S, Ragard LR, Buys SS; PLCO Project Team. Results from four rounds of ovarian cancer screening in a randomized trial. *Obstet Gynecol* 2009; **113**: 775–782.
63. Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, Lewis S, Davies S, Philpott S, Lopes A, Godfrey K, Oram D, Herod J, Williamson K, Seif MW, Scott I, Mould T, Woolas R, Murdoch J, Dobbs S, Amso NN, Leeson S, Cruickshank D, McGuire A, Campbell S, Fallowfield L, Singh N, Dawnay A, Skates SJ, Parmar M, Jacobs I. 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). *Lancet Oncol* 2009; **10**: 327–340.
64. Menon U, Gentry Maharaj A, Jacobs I. Ovarian cancer screening and mortality. *JAMA* 2011; **306**: 1544–1545.
65. Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, Valentin L, Timmerman D. Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumor analysis study. *Clin Cancer Res* 2009; **15**: 684–691.
66. Fenchel S, Grab D, Nuessle K, Kotzerke J, Rieber A, Kreienberg R, Brambs HJ, Reske SN. Asymptomatic adnexal masses: correlation of FDG PET and histopathologic findings. *Radiology* 2002; **223**: 780–788.
67. Bast RC Jr, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest* 1981; **68**: 1331–1337.
68. Zhang B, Cai FF, Zhong XY. An overview of biomarkers for the ovarian cancer diagnosis. *Eur J Obstet Gynecol Reprod Biol* 2011; **158**: 119–123.
69. Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, Gajewski W, Kurman R, Bast RC Jr, Skates SJ. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009; **112**: 40–46.
70. Urban N, Thorpe JD, Bergan LA, Forrest RM, Kampani AV, Scholler N, O'Briant KC, Anderson GL, Cramer DW, Berg CD, McIntosh MW, Hartge P, Drescher CW. Potential role of HE4 in multimodal screening for epithelial ovarian cancer. *J Natl Cancer Inst* 2011; **103**: 1630–1634.
71. Kamat AA, Baldwin M, Urbauer D, Dang D, Han LY, Godwin A, Karlan BY, Simpson JL, Gershenson DM, Coleman RL, Bischoff FZ, Sood AK. Plasma cell-free DNA in ovarian cancer: an independent prognostic biomarker. *Cancer* 2010; **116**: 1918–1925.

