The role of transvaginal ultrasound in screening for ovarian cancer

Authors
Campbell S,1 Gentry-Maharaj A2

Affiliations
1 Create Health Fertility, London, EC2V 6ET; 2 Gynaecological Cancer Research Centre, Department of Women’s Cancer, Institute for Women’s Health, University College London, London W1T 7DN;

Corresponding Author
Professor Stuart Campbell
Create Health Fertility
150 Cheapside
London EC2V 6ET
profscampbell@hotmail.com

Short title
Transvaginal ultrasound screening for ovarian cancer

Keywords
Ovarian cancer, screening, transvaginal ultrasound
Abstract/Precis (234 words)

Ovarian cancer is a low prevalence postmenopausal cancer with a high mortality rate and is the 5th most lethal cancer in women. The most common serous subtype with TP53 mutations spreads rapidly throughout the peritoneal cavity (stage III) when 5 year survival is 10%. If diagnosed while confined to the ovary (stage I) the survival exceeds 90%. This is the rationale for screening. Annual transvaginal ultrasound (TVU) scans used as a primary screening modality or as a second line test following primary screening with serum CA125 (multimodal) has been investigated in several trials. Only two large randomised controlled trials have provided mortality data. The US PLCO trial studied over 78,000 women (randomised to screening with either TVU or CA125, or control) over 6 years with 14 years follow up and found no mortality benefit from screening and increased morbidity in the screened arm. The UKCTOCS studied over 202,000 women randomised to TVU, multimodal or control in a 1:1:2 ratio over 7-11 years with 11 years follow up. CA125 was interpreted by the Risk of Ovarian Cancer algorithm which identifies a rise in the level rather than a fixed cut-off. There was a late reduction in mortality after 7 years in the screened arm (23% in the multimodal and 21% in the TVU) but the overall reduction was not significant. Further follow-up may reveal whether TVU has a primary or secondary role in ovarian cancer screening.
Introduction (3,702 words)

Ovarian cancer (OC) is principally a disease of postmenopausal women with only 9% occurring before the age of 50 years.\textsuperscript{1} In developed countries it is the second most common genital tract malignancy, with women having a 1-2% life-time risk of developing the disease.\textsuperscript{1} It is the most lethal gynaecological malignancy, with an overall 5-year survival of 45%.\textsuperscript{2} In 2017 in the United States it is projected that 22 440 women will develop OC and 14 080 women will die from the disease.\textsuperscript{3} In Europe, the corresponding figures are 65 600 and 42 700, respectively.\textsuperscript{1} 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 and most of these cases are associated with germline mutations in the \textit{BRCA1}, \textit{BRCA2} and \textit{MMR} genes.\textsuperscript{4,5} More recently moderate penetrance genes (\textit{RAD51C}, \textit{RAD51D} and \textit{BRIP1}) have been described which confer risk of OC of 5-12%.\textsuperscript{6} Between 80 and 85% of cancers are epithelial in origin (EOC), serous being the most common subtype, which usually presents at an advanced stage and has the poorest outcome.\textsuperscript{7} 60% of cases of OC present late when the cancer is disseminated throughout the peritoneal cavity (Stage III) as early symptoms are often vague such as abdominal distension and a feeling of bloatedness.\textsuperscript{8} The 5-year survival in these women is as low as 10% but exceeds 90% when the disease is diagnosed at Stage I (i.e. confined to one or both ovaries).\textsuperscript{9} This forms the rationale for OC screening programs, the premise being that early detection may affect long-term survival.

Pathogenesis

EOC presents as a heterogeneous group of tumours. Borderline ovarian tumours represent about 20% of EOC and have a survival rate of over 80% as characteristically they do not invade the basement membrane and are usually diagnosed at an early stage. Based on this, the focus of screening efforts is to detect invasive cancers. Over the past decade it has become apparent that on a morphologic and molecular genetic basis, invasive (iEOC) are divided into two subtypes: Type I are slow-growing cancers with good prognosis, such as low-grade serous (which according to the WHO 2014 criteria\textsuperscript{10} include serous borderline tumours with invasive implants), low-grade endometrioid, clear cell and mucinous carcinomas. They are usually detected at an early stage; however, they constitute only 25% of OCs and account for approximately 10% of OC deaths. Type II tumours are fast growing aggressive cancers and include high-grade serous and undifferentiated carcinomas representing 75% of all ovarian carcinomas and 90% of OC deaths. They are more difficult to detect at an early stage due to their rapid growth and dissemination and display TP53 mutations in over 80% of cases.\textsuperscript{11}
The traditional view of ovarian carcinogenesis has been that the various different tumours 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 previous ovulations. This theory goes some way to explain why previous use of oral contraceptives reduces the risk of developing EOC. However, recent studies on the origin of OC have identified a precursor in-situ lesion called serous intraepithelial tubal carcinoma (STIC) in the distal part of the Fallopian tube which morphologically and molecularly, resembles high-grade ovarian serous carcinoma. Rather than developing de novo from the ovary, as previously proposed, the majority of Type II cancers appear to arise from a STIC in the fimbriated end of the Fallopian tube before spreading to the ovary and are thus of Mullerian not mesothelial origin. Based on these new concepts of ovarian carcinogenesis the ACOG have issued a Committee Opinion recommending prophylactic salpingectomy for women having hysterectomy with ovarian conservation and also as an alternative to tubal ligation as “an opportunity to prevent ovarian cancer in their patients”. No data on the effectiveness of this advice is yet available.

In this review the term ovarian cancer will include cancers of the fallopian tube. In view of the WHO 2014 classification, it must be highlighted that ultrasound as a modality was not designed/has the ability to detect the rarer primary peritoneal cancers therefore will have limited impact on this entity.

**Screening for Ovarian cancer**

As 90% of OCs occur in postmenopausal women, the concept of universal screening of this population to detect cancer at an early treatable stage seems logical. A number of recent studies have evaluated screening for familial OC in pre and postmenopausal women with a lifetime risk greater than 10%. Annual screening was shown not to be effective but multimodal screening using the Risk of Ovarian Cancer (ROC) algorithm at 4 monthly intervals shows a significant stage shift in EOC. Nevertheless prophylactic salpingo-oophorectomy is recommended in these women when childbearing is completed. This review will therefore confine itself to universal screening of asymptomatic postmenopausal women.

The prevalence of OC 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 I disease to make screening worthwhile, but, most importantly, the specificity must be very high
 (>99.6%) to give a positive predictive value of at least 10%, i.e. a maximum of 10 operations per cancer detected.

The aim of all screening efforts to date has been detection of the disease at an early stage. With the novel insights into the existence of STIC lesion, future strategies may be targeted towards detecting the disease at a pre-malignant stage, akin to cervical screening. However, no screening strategies are currently able to detect STIC lesions. Circulating tumour (ctDNA) has been the most promising to date and may provide a more specific test in the future that would be able to detect Type II cancers and STIC lesions.

Demonstrating a stage shift towards detection of iEOC at Stage I/II 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. Early detection through screening (with either the multimodal or ultrasound strategy) may not necessarily translate into a mortality benefit as this lead time bias is impacted by many factors, including treatment delivered. If the cancers detected at an earlier stage have surgery with a curative intent, then the opportunity to impact on mortality is much greater. 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. This review will focus on such studies.

Universal annual ultrasound screening is a huge logistical exercise, although the difficulty of the task has sometimes been exaggerated. After all, in most countries at least 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 but, as EOC is not a cancer of the socially deprived, this may not be a significant bias. Nevertheless to minimise bias ideally recruitment should be done on the basis of inviting women to participate through local population registers. Scans in recent trials were usually performed by highly trained sonographers who record details of ovarian size and morphology and also Doppler indices if required.

**Screening tests for ovarian cancer**

There are two currently used modalities for OC screening, transvaginal ultrasound (TVU) and serum CA125 performed on an annual basis. Transabdominal (TA) scans were performed in the original
ultrasound screening trial in 1989 but this technique is now only used rarely due to the superior resolution of TVU. The aim of TVU is to visualise both ovaries in both longitudinal and transverse planes and calculate the volume by the prolate ellipsoid formula (length × width × depth × 0.523) (Figure 1). Any sudden increase in volume or change in morphology (Figure 2) is scored according to protocol and a decision to operate made by the appropriate clinician. The normal postmenopausal ovary is a small structure lateral to the uterine fundus and close to the pelvic sidewall with a range in volume of about 1-2 mL and cannot always be identified because of shadowing by bowel, fibroids or other factors even by skilled practitioners of TVU. In the Kentucky screening study at least one ovary was not seen in 16% of cases. In most studies, failure to visualise the ovary is regarded as a negative screen. In the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), in order to minimize the chance of missing an abnormal ovary, sonographers were requested to demonstrate a 2-cm length of a clearly defined iliac vein in the pelvic side wall if the ovary was not visualised. A similar protocol was used in the PLCO Screening Trial and a minimum time of 5 min was spent to identify each ovary. Another problem with ultrasound screening is the high incidence of benign cysts which when complex are frequently difficult to differentiate from malignant tumours. In the PLCO Trial 21% of women over the age of 50 had an ovarian cyst, of which 5.5% were complex. In UKCTOCS adnexal masses were present in 10% of women at the first scan of which 17% were complex with solid components and 1% were malignant. As a result a large number of studies have been carried out notably from the IOTA group with some success to generate ultrasound morphological and blood flow indices to help identify which adnexal tumours are malignant (Figure 2). A final problem is that women found to be screen positive require surgical removal of tubes and ovaries rather than a simple biopsy as in cervical and breast cancer screening programmes.

Cancer antigen 125 (CA125) was first described by Bast et al in the early 1980s. Serum levels at a fixed level of 35kU/L and above has a high overall sensitivity for EOC of 80% in postmenopausal women but this falls to 50% for Stage I disease. False positive rates are high as the antigen is also raised in benign conditions such as endometriosis. In order to improve sensitivity, Skates et al introduced a more sophisticated approach by rejecting a fixed cut-off CA125 level and analysing serial values that are available in the screening context. They demonstrated that elevated CA 125 levels in women without OC 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 OC based on the rise in CA125 and allows women to be triaged into low, intermediate or high-risk categories. Thus the absolute value is less important than the rise; for example a rise in value from 8 to 16 kU/L (i.e. values which would usually be regarded as normal) over a period of 3
months could put a woman in ‘at high-risk’ category. The ROC algorithm was evaluated in a screening trial\textsuperscript{28} with TVU employed as a secondary test to visualise the ovaries in screen positive cases in order to improve specificity (multimodal screening). This randomised trial of 13,582 postmenopausal women demonstrated a specificity of 99.8\% and a PPV of 19\% for primary invasive EOC. These results, in conjunction with an indication of survival benefit with screening,\textsuperscript{29} prompted the UKCTOCS multicenter trial which is discussed below.

**Recent studies**

There are now four large recently completed trials on OC screening by means of TVU and/or CA125 that have published data in the last decade:

1. The *University of Kentucky ovarian cancer screening trial*\textsuperscript{22} 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. 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 EOC, the sensitivity decreased to 76.3\%. Most (82\%) of the primary OCs were early stage (Stage I/II). Serum CA 125 levels were increased (>35 kU/L) at the time of detection in 13 of 15 (87\%) patients who had Stage III EOC but in only three of 15 (20\%) patients who had Stage I or II disease. At a mean follow up of 5.8 years, the women in the trial had a significantly longer 5-year survival (74.8±6.6\%) compared to the women from the same institution, treated by the same surgical and chemotherapeutic protocols, who were not screened (53.7±2.3\%).\textsuperscript{30}

2. The *Japanese Shizuoka Cohort Study of Ovarian Cancer Screening*\textsuperscript{31} 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 CA125 using a cut-off of 35 kU/L. 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. The screening strategy achieved sensitivity for ovarian cancer of 77.1\% and a specificity of 99.9\%. The proportion of Stage I OC 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).*\textsuperscript{23} This is a randomised controlled trial of 78,216 women aged 55-74 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 CA125, using a fixed cut-off of 35 kU/L, and transvaginal ultrasonography for 4 years, followed by CA 125 alone for a further 2 years. On ultrasound an ovary with a volume greater than 10 ml or containing a cyst of any size with any solid area or papillary projection was considered a positive screen. 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,32 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 OC. A total of 72% of the screen-detected cancers were late stage (Stage III/IV). Recently, mortality data have been reported and these are discussed below.

4. The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS).18 33 In this trial, 202,638 postmenopausal women aged 50-74 years were randomised to either control or annual screening with ultrasound or a multimodal strategy in a 2:1:1 fashion for 7-11 years. In the multimodal group, CA125 was interpreted using the ROC algorithm to triage the women into low, intermediate and elevated risk. Those at intermediate risk had a repeat CA125 in 12 weeks, whereas those with elevated risk were referred for a transvaginal scan and repeat CA125 in 6 weeks. In the prevalence screen, 91% were classified as low risk by the ROC algorithm and returned to annual screening. Only 9% of women required a repeat CA125 test and an ultrasound scan and 0.2% had surgery. However, the performance characteristics from the prevalence screen of the two screening strategies33 suggest that for detection of primary OC the multimodal strategy had superior sensitivity (89.4%), specificity (99.8%) and PPV (23%) to ultrasound screening alone (sensitivity 84.9%, specificity 98.2%, PPV 5%). When restricted to the detection of primary invasive EOC, the sensitivity of multimodal screening was maintained at 89%, whereas the sensitivity of the ultrasound-based strategy decreased to 75.0%. 50% were detected at an early stage in both arms. At the end of the follow up there was a significant stage shift to earlier diagnosis (I/II) in the multimodal arm (36.1% versus 23.9% in control) which was absent from the ultrasound arm (22.4%).34 However although the sensitivity in the multimodal arm was determined by the ROC algorithm, ultrasound was essential to reduce false positive diagnoses; ROC algorithm has a PPV of 15% without ultrasound but 23% with ultrasound as a second line test. Despite the ultrasound strategy performing much better at detecting borderline tumours, the proportion of Type II cancers in the multimodal or ultrasound arm was similar (74.9% vs 77.9%). Mortality data was published in 2016 and is discussed below.

Effect of TV ultrasound and CA125 screening on mortality
As described above the only proof of a mortality effect of screening is through randomised controlled trials with sufficient numbers to provide statistical significance. Data from only two trials fit this criterion.

In the PLCO trial at a median follow-up of 12.4 years a total of 212 women had a screen-detected cancer in the intervention arm and 176 were identified in the control arm. There was no significant effect on mortality from screening; 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 CA125 using an absolute cut-off and TVU did not reduce mortality from the disease. Moreover, the excess morbidity of carrying out surgery in women with false-positive results was 15%. An extended 14.7 year follow up of PLCO indicated no mortality benefit from screening for OC with CA125 and TVU. The PLCO study results influenced the FDA which made the following FDA safety Communication in September 2016 “...there are currently no screening tests for ovarian cancer that are sensitive enough to reliably screen for ovarian cancer without a high number of inaccurate results”. For physicians: Do not recommend or use tests that claim to screen for ovarian cancer in the general population of women.

In UKCTOCS at a median follow-up of 11 years OC was diagnosed in 1282 women: 338 in the multimodal, 314 in the ultrasound, and 630 in the no screening group. Of these women, 148 women in the multimodal group, 154 in the ultrasound group, and 347 in the no screening group had died of OC. The primary analysis using a Cox proportional hazards model gave a non-significant mortality reduction over years 0-14 of 15% with multimodal and 11% with ultrasound. However there was a late effect in mortality with a mean reduction in years 7-14 of 23% in the multimodal and 21% in the ultrasound group. The authors of the paper stated that a weakness of the study was that in the original statistical design a test such as the weighted log-rank test (used in the PLCO trial) which would show this late effect was not employed. A post hoc analysis with this test showed a significant reduction in OC deaths in the multimodal arm.

The PLCO and UKCTOCS trials came up with very different mortality results and it is worthwhile exploring the reasons for this. Concerns have been raised about the PLCO trial design; a screening failure could be recorded as long as 10 years after the end of the trial (40.6% of women had OC diagnosed after screening ended), CA125 was interpreted using an absolute cut-off, positive screen was defined as either CA125 or TVS positive, and management of screen positives was at the discretion of the treating clinician. The UKCTOCS study design avoided all of these flaws; mortality was
established within 3 years of the end of the study, there was central control of all aspects of the trial by the Trial Management team and the ROC algorithm was employed to determine a CA125 rise. Also unlike the PLCO trial a comparison could be made between the performance characteristics of the two screening modalities.

Conclusion

Both ultrasound and serial CA125 are well accepted by patients with compliance in both arms of the UKCTOCS trial of 98.9%. Alternative screening tests both for imaging the ovaries and as liquid markers have been researched but there is little evidence that alternative screening tests are likely to appear in the near future.\(^\text{38}\) The principal weakness of ultrasound screening is the high false positive rate which in most instances involves the detection of a benign adnexal tumour. In a screening study small adnexal tumour may be cancerous and yet may not have the typical characteristics described for advanced tumours so they require detailed follow up until they are proven to be innocent. With CA125 screening false positives are not usually associated with adnexal masses so TVU is necessary to improve specificity and PPV. Although the Kentucky ultrasound-based study described an improved survival in women who had ultrasound screening, the lack of randomisation is likely to have introduced bias. Furthermore survival rates rather than mortality were reported which are subject to lead-time bias. The only large randomised study with an ultrasound arm was UKCTOCS. This showed that ultrasound has inferior performance characteristics to multimodal screening with lower sensitivity for invasive disease and a higher false positive rate and an inferior stage shift. It is therefore surprising that the mean mortality reduction in years 7-14 after cessation of screening in the ultrasound arm was 21% which is not dissimilar to the reduction in the multimodal arm of 23%.\(^\text{34}\) One has to be cautious about over interpretation of the data as the confidence limits are large in these later follow up years but one can conjecture that there may be an unexpected mortality reducing effect operating in the ultrasound arm irrespective of performance characteristics. Data from UKCTOCS shows that ultrasound has a much higher sensitivity for borderline tumours (91% vs. 55%) than CA125.\(^\text{34}\) It is possible that the surgical removal of these tumours together with complex benign cysts may have reduced the prevalence of invasive EOC in the ultrasound arm. The UKCTOCS follow up is being extended for a further 4 years; analysis of this extended mortality data will hopefully provide further clarification of the role of ultrasound both as a primary screening test or secondary to CA125 in the reduction of mortality from what is one of the greatest health problems in postmenopausal women. Whether ultrasound will have a role as a primary screening test or secondary as in the multimodal screening arm of UKCTOCS it will remain an essential tool in screening for ovarian cancer.
References


Figures

**Figure 1:** Transvaginal ultrasound scan showing normal right ovary with on-screen digital calipers. The volume is 2.2 ml.

**Figure 2:** Transvaginal ultrasound scan showing enlarged cystic right ovary (25 ml) with vascular solid elements 1 year after a normal scan. Low-volume type II serous carcinoma; the patient survived.