Screening for ovarian cancer: a systematic review

R Bell
M Petticrew
S Luengo
TA Sheldon

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:
HTA Despatch   Email: orders@hta.ac.uk
c/o Direct Mail Works Ltd   Tel: 02392 492 000
4 Oakwood Business Centre   Fax: 02392 478 555
Downley, HAVANT PO9 2NP, UK   Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Screening for ovarian cancer: a systematic review

R Bell[1]
M Petticrew[1]
S Luengo[2]
TA Sheldon[1]

[1] NHS Centre for Reviews and Dissemination, University of York, UK
[2] Health Services Research Unit, Instituto de Salud Carlos III, Madrid, Spain

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Population Screening Panel.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In England, policy options in this area are to be considered by the National Screening Committee, chaired by the Chief Medical Officer, who will take into account the views expressed here, further available evidence and other relevant considerations.

Contents

List of abbreviations

Bart's	St Bartholomew's Hospital, London
BSO	bilateral salpingo-oophorectomy[*]
CA 125	cancer antigen 125
CDI	colour Doppler® imaging[*]
ERTOCS	European Randomised Trial of Ovarian Cancer Screening
FIGO	International Federation of Gynecology and Obstetrics
FNA/B	fine needle aspiration cytology/biopsy[*]
GHQ	General Health Questionnaire
HAD	Hospital Anxiety and Depression Scale
NIH	National Institutes of Health (USA)
PLCO	Prostate, Lung, Colon, Ovary trial
PPV	positive predictive value
RCT	randomised controlled trial
STAI	State–Trait Anxiety Inventory
TAH	total abdominal hysterectomy[*]
TAS	transabdominal sonography[*]
TVS	transvaginal sonography[*]
UKCCCR	UK Coordinating Committee for Cancer Research

[*] Used only in tables and figures

Executive summary

Background

Ovarian cancer is the seventh most common cancer in women. The overall 5-year survival rate is only 30%. For women whose disease at diagnosis is localised to the ovaries, survival is about 75% at 5 years, but only a quarter of cases in the UK are currently diagnosed at such an early stage. This has led to interest in screening methods that might result in earlier diagnosis and reduce both mortality and morbidity.

Screening methods include ultrasound scanning and the measurement of the tumour marker cancer antigen 125 (CA 125) in serum. When used for screening, CA 125 measurement is followed by ultrasound scanning in women with abnormal CA 125 levels ('CA 125-based screening'). Women with persistently abnormal findings are referred for diagnostic abdominal surgery for removal of ovarian tissue.

Objectives

- To evaluate the performance of current screening tests for ovarian cancer.
- To assess the adverse effects of screening, including morbidity associated with surgical intervention and psychological morbidity associated with false-positive diagnosis.
- To report on the stage of development of newer methods of screening.
- To investigate the potential cost-effectiveness of screening in different risk groups.

Methods

The review was carried out using structured guidelines for systematic reviews. These are described in detail in the full report.

Results

The effectiveness of screening
Although three large RCTs are in progress, no RCTs of screening for ovarian cancer have been completed. In the absence of evidence of effectiveness, it would be premature to establish any kind of screening programme.

Screening test performance
The evidence suggests that both CA 125-based screening and ultrasound screening can detect a higher proportion of ovarian cancers at Stage I than that currently observed in the UK. About 50% (95% CI; 23–77) are diagnosed at Stage I in CA 125-based screening studies, and about 75% (95% CI; 35–97) in ultrasound screening studies. These data should be interpreted cautiously, however, as they are based on small numbers of cancers detected in diverse studies carried out mainly on self-selected women.

From the limited data available, annual screening with ultrasound appears to have a sensitivity or detection rate close to 100%. The reported sensitivity of annual CA 125-based screening is about 80%. The precision of these estimates is low, however, as they are based on small numbers of cancers.

The false-positive result rate is about 1.2–2.5% for women screened by ultrasound scanning and 0.1–0.6%. for CA 125-based screening.

About 0.5–1% of women will suffer a significant complication due to surgery and most of those who do not have ovarian cancer will have a benign gynaecological condition. There is a risk that detection of benign and borderline tumours may become a target of ovarian screening, even though they would not have been associated with any morbidity during a patient's lifetime.

Intervals for ultrasound scanning of between 1 year and 3 years are under investigation in the RCTs. CA 125-based screening has been carried out annually. The effect of different screening intervals on the detection rate and false-positive rate has not been formally investigated.

About 3–12% of screened women are recalled for further testing and assessment, resulting in potential distress and anxiety to otherwise healthy women.

The potential impact of screening for ovarian cancer
The low positive predictive value of ovarian screening (3% for surgery and 0.6% for initial recall for annual ultrasound screening; 15% for

surgery and 1% for initial recall for annual CA 125-based screening) is due mainly to the relatively low prevalence of ovarian cancer, which limits the potential cost-effectiveness of general population screening.

Evidence suggests that ultrasound screening is more sensitive than CA 125-based screening but that the latter may result in fewer false-positives and, hence, a higher positive predictive value. However, a less sensitive test must to be repeated more frequently to achieve the same overall detection rate of ovarian cancers, which may reduce the apparent advantages of CA 125-based screening. The most efficient screening method and interval is unknown, but modelling studies suggest that annual CA 125-based screening may provide lower overall benefits but be more cost-effective at detecting early stage cancers than annual ultrasound screening.

It is suggested that the addition of colour Doppler® imaging (CDI) to ultrasound screening may reduce the false-positive rate but reported results are mixed.

Screening a higher-risk population

A family history of ovarian cancer is one of the strongest risk factors for developing the disease and some UK centres currently offer screening to women with a strong family history. Until RCTs have been completed, there is no evidence as to whether, or by how much, screening women at higher risk reduces mortality.

For some women with an extensive family history of ovarian and/or certain other cancers, the increased risk is associated with an inherited genetic mutation. Carriers of some specific mutations may have a lifetime risk of developing ovarian cancer as high as 50–60%. The identification of some of these mutations raises the possibility of testing individuals in these families to determine whether they are carriers, potentially enabling more accurate assessment of risk.

Conclusions

Implications for policy
- Further evidence is required before a decision can be made about the potential benefits, harms

and costs of screening for ovarian cancer. While awaiting the results of the current trials, demand for screening is likely to increase, and a strong national lead will be required.
- The relatively low prevalence of ovarian cancer means that the positive predictive value of screening tests is low. Since the consequence of a false-positive result is a surgical procedure, consideration of the overall impact of ovarian cancer screening is important. The low prevalence also limits the potential cost-effectiveness of population screening.
- Screening women who are at risk because of a strong family history may be more cost-effective but this has not been established. No RCTs are planned in this group, but a screening study has been established. This will provide some evaluation using intermediate outcomes of screening but may also increase demand for screening services.

Implications for research
- In a few years, RCTs should provide an estimate of the impact of screening on mortality. Assessment of the adverse effects of screening and the relative cost-effectiveness of different screening strategies would enhance information from the trials.
- New or modified screening tests should be compared with those being evaluated in current trials. Test developments which require further evaluation include: the marginal impact of adding CDI to ultrasound screening; the use of CA 125 levels in multivariate algorithms to determine thresholds for ultrasound and surgical intervention, and the marginal value of adding CA 125 measurement to ultrasound screening. The screening modalities will require continuous re-evaluation in line with technical developments.
- Research efforts should be directed towards evaluating both the clinical and cost-effectiveness of screening strategies for patients at high risk. This includes: investigation of any differences in the natural history; performance of screening tests compared with the strategies used in RCTs; investigation of age-specific risks of developing ovarian cancer, and psychological impact and value of risk assessment.
- Research is also needed into the impact of genetic testing on health outcomes and the level of demand for such services.

Chapter 1

Background

The size of the problem

Ovarian cancer is the seventh most common site for cancer in women worldwide and is most common in western industrialised countries.[1] In 1994, in England and Wales, there were 3859 deaths caused by ovarian cancer and, in 1989, 5100 new registrations of the disease.[2,3]

Pathology

Ovarian cancer is not a single disease but represents a group of cancers arising from a variety of different cell types. Histological classification is complex but the majority of primary malignant tumours, about 90%, are of epithelial origin.[4,5] One distinct subset of tumours has pathological features intermediate between benign and invasive malignant disease; these are termed borderline tumours (also referred to as 'low malignant potential' tumours) and have a much better prognosis than invasive tumours.[5] Non-epithelial ovarian cancer includes germ cell tumours and sex cord stromal tumours. Germ cell tumours (which comprise around 3% of all ovarian cancers) arise, on average, at an earlier age and have a better prognosis than epithelial ovarian cancers.

The focus of this review is on screening for invasive epithelial ovarian cancers, although some of the information discussed, particularly that derived from routine data sources, will relate to all primary ovarian cancers.

Symptoms and treatment

Ovarian cancer tends to give rise to vague or non-specific symptoms such as abdominal discomfort, swelling caused by tumour mass or ascites, menstrual irregularities or gastrointestinal symptoms. The tumour spreads from the ovaries locally and also by peritoneal seeding, which can lead to widespread disseminated intra-abdominal disease. This intra-abdominal spread can occur when the ovarian tumour mass is small, which, together with the insidious nature of the symptoms, means that the disease is frequently widespread at diagnosis. This has led to interest in the potential of screening for the disease, in the hope that identifying it before clinical presentation may increase the likelihood that treatment is effective.

The extent of spread of the disease at diagnosis is classified into four stages as shown in the box below. Establishing the stage of disease accurately requires extensive surgical exploration of the pelvis and abdomen.

FIGO 1986 staging system for ovarian cancer[6]	
Stage	**Definition**
I	**Growth limited to the ovaries**
IA	Limited to one ovary; no ascites; no tumour on external surfaces, capsule intact
IB	Limited to both ovaries; no ascites; no tumour on external surfaces, capsule intact
IC	Tumour either Stage IA or IB but with tumour on the surface of one or both ovaries, or with capsule ruptured, or with ascites containing malignant cells, or with positive peritoneal washings
II	**Growth involving one or both ovaries with pelvic extension**
IIA	Extension or metastases to the uterus or tubes
IIB	Extension to other pelvic tissues
IIC	Tumour either Stage IIA or IIB but with tumour on the surface of one or both ovaries, or with capsule ruptured, or with ascites containing malignant cells, or with positive peritoneal washings
III	**Tumour involving one or both ovaries with peritoneal implants outside the pelvis or positive retroperitoneal or inguinal nodes; superficial liver metastases; tumour limited to the true pelvis but with histologically verified malignant extension to small bowel or omentum**
IIIA	Tumour grossly limited to the true pelvis with negative nodes but with histologically confirmed microscopic seeding of abdominal peritoneal surfaces
IIIB	Tumour of one or both ovaries; histologically confirmed implants of abdominal peritoneal surfaces, none exceeding 2 cm in diameter, nodes negative
IIIC	Abdominal implants greater than 2 cm in diameter or positive retroperitoneal or inguinal nodes
IV	**Growth involving one or both ovaries with distant metastases; if pleural effusion is present, there must be positive cytological test results to allot a case to Stage IV; parenchymal liver metastases equals Stage IV**

Treatment of the disease consists of surgical removal of as much of the tumour as possible, followed by adjuvant therapy if indicated. The usual treatment for Stage I low-grade disease is surgery alone, while more advanced disease may be treated with surgery followed by adjuvant chemotherapy.[6]

Incidence and mortality

The incidence of ovarian cancer can be estimated from the number of cancer registrations. The overall registration rate for ovarian cancer in England and Wales in 1989 was 19.7 per 100,000 women.[3] The incidence of ovarian cancer is strongly related to age, and increases markedly over the age of 40 years; 94% of registrations occur in women over the age of 40 years and 48% in women aged between 50 and 69 years. The registration rate in this age group is around 44 cases per 100,000 women per year *(Figure 1)*.

The mortality rate for ovarian cancer in 1994 was 14.7 per 100,000 women.[2] The mortality rate rises with age and, for women between the ages of 50 and 69 years, the mortality rate was 30 per 100,000 in 1994. Approximately 1.3% of all deaths in women are caused by ovarian cancer, the relative impact on mortality being greatest between the ages of 40 and 59 years, when about 5% of all deaths in women are attributed to ovarian cancer. When compared with breast cancer, the most common cancer in women, ovarian cancer results in one-fifth as many cases per year and less than one-third of the number of deaths.[2,3]

The overall mortality rate from ovarian cancer in Great Britain has been stable for the past 20–30 years.[7,8] However, within this period, there has been a slight decrease in the mortality rate in women under the age of 55 years. An analysis of survival trends in Scotland shows that survival in younger age groups has improved over the past few decades,[9] suggesting that the reduction in mortality may partly reflect improved treatment effectiveness.[8]

The number of deaths caused by ovarian cancer in the future will be affected by the ageing of the population, the effectiveness of treatment and changes in the prevalence of factors such as family size, oral contraceptive use and oophorectomy. The effect of trends in these factors may take some time to become apparent, and the overall impact on the mortality rate and the overall number of deaths is difficult to predict.
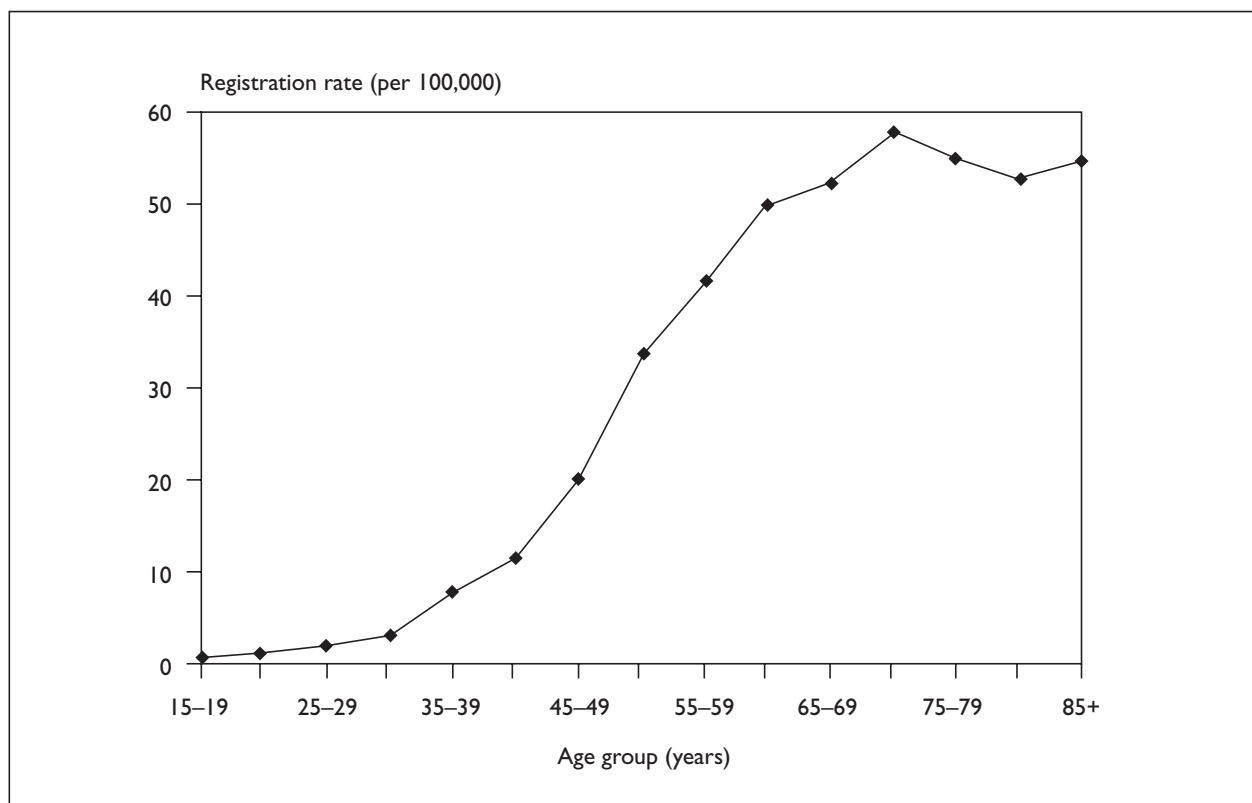
FIGURE 1 *Registration rates for ovarian cancer in England and Wales[3]*

Survival

Overall, the 5-year survival rate for ovarian cancer in Great Britain is about 30%, and there has been minimal improvement in this figure over the past 20–30 years.[9,10]

Survival by stage from information from a number of population-based cancer registries in the UK is shown in *Table 1,* with an indication of current survival and proportions diagnosed at each stage within these populations. There is some variation in the reported survival rates and the proportion diagnosed at each stage. These variations may be caused by a number of factors:

- variation in the completeness of staging data
- variation in the population age structure, since younger patients present with earlier disease on average and experience better stage-specific survival[11]
- variation in practice of performing staging laparotomies and classifying the results[12]

- real differences between populations in the proportion presenting early and in the effectiveness of treatment.

The data show a consistent and strong relationship between stage at diagnosis and 5-year relative survival. The percentage of patients diagnosed at Stage I, when the tumour is localised to the ovaries, varies between 22% and 28%, with the 5-year survival rate for these patients varying between 72% and 81%. Survival rates for the majority of cancers which present at Stages II–IV are much poorer. A similar picture is apparent in published data from international registries.[11,14,15]

This suggests that there may be scope for outcomes to be improved by increasing the proportion of cancers diagnosed early. However, it is possible that the observed survival advantage for early ovarian cancer reflects differences inherent in the tumour biology rather than the effectiveness of treatment. Clinically detected early cancers may be slower growing and have less propensity to become

TABLE 1 *Survival by stage and proportion diagnosed at each stage for selected UK regions*

Registry (age range)	Number of cases	Years diagnosed	Stage at diagnosis (%)		5-year survival by stage (%)	
Thames (15–74 years)	4570	1986–90	I	28	I	72
			II	21	II	36
			III	3	III	22
			IV	39	IV	13
			n.k.	9	n.k.	48
East Anglia (all ages)	654	1989–91	I	22	I	72
			II	8	II	30
			III	39	III	11
			IV	13	IV	2
			n.k.	18	n.k.	14
West Midlands[13] (all ages)	1603	1985–87	I	22	I	81
			II	4	II	35
			III	33	III	10
			IV	5	IV	7
			n.k.	35	n.k.	–
Scotland (35–64 years)	835	1987, 1992–94	I	28	I	77
			II	10	II	49
			III	44	III	17
			IV	15	IV	8
			n.k.	3	n.k.	4
Scotland (all ages)	1829	1987, 1992–94	I	23	I	74
			II	9	II	41
			III	45	III	13
			IV	16	IV	3
			n.k.	9	n.k.	2

n.k., not known

widely disseminated than those diagnosed at a more advanced stage. Screen-detected early cancers may not have the same favourable characteristics and may not therefore demonstrate the same survival advantage observed for early cancers in an unscreened population.

In Great Britain, approximately 75% of women with Stage I disease currently survive for 5 years. If screen-detected early cancers show similar survival rates, a significant proportion of women with ovarian cancer would not be 'cured', even if screening could detect all cancers while they are localised. However, a higher standard of staging and treatment might be achieved in a screening programme, leading to survival rates exceeding those currently observed in an unscreened population. Clinical trials restricted to women with accurately staged low-grade Stage IA and IB disease have demonstrated long-term survival rates in excess of 90%.[16] However, this may not be possible in the case of screen-detected Stage I disease.

Risk factors and aetiology

A wide range of risk factors has been postulated for ovarian cancer;[17,18] the most reliable information relates to reproductive factors, oral contraceptive use and family history *(Table 2)*.

An analysis of pooled data from 12 case–control studies undertaken in the USA has investigated risk factors for invasive epithelial ovarian cancer.[19] These data showed a protective effect for pregnancy, with the risk reducing for each additional term pregnancy. Pregnancies ending in miscarriage or termination were also protective. Ovarian cancer risk also reduced with increasing duration of breast feeding.

TABLE 2 Major risk factors for epithelial ovarian cancer

Risk factor	Relative risk/odds ratio (95% confidence interval)
None	1.0
Oral contraceptive use[19]	0.66 (0.55–0.78)
Any term pregnancy[19]	0.47 (0.4–0.56)
One first- or second-degree relative with ovarian cancer[22]	3.1 (2.2–4.4)
Two or three relatives with ovarian cancer[22]	4.6 (1.1–18.4)

Use of oral contraceptives has consistently been reported to reduce the risk of ovarian cancer.[19–21] The risk appears to reduce with increasing duration of oral contraceptive use.[19] There is uncertainty over whether newer oral contraceptive formulations confer the same degree of protection as older, higher dose formulations.[19,20]

Women with a family history of ovarian cancer are at increased risk of ovarian cancer.[22] This is discussed further below.

The contribution of each of these risk factors to ovarian cancer incidence depends both on the strength of the association and the prevalence of the risk factor in any given population. An analysis of data from the USA suggests that the most important risk factor on a population basis is the use of oral contraceptives; over half of all ovarian cancers in the USA might be prevented if all women used oral contraceptives for at least 4 years.[23]

Based on these observations, it has been hypothesised that the suppression of ovulation, whether by pregnancy, breast feeding or the oral contraceptive pill, confers protection from ovarian cancer. This was first proposed by Fathalla,[24] who hypothesised that 'incessant ovulation', and the subsequent trauma and healing of the ovarian epithelium, predisposed to malignant change. An alternative hypothesis, that the high levels of circulating gonadotrophins associated with ovulation were responsible for inducing malignant change, has also been proposed.[25] The available epidemiological evidence is not wholly consistent with either of these hypotheses.[26]

Genetics of ovarian cancer

A family history of ovarian cancer in a first- or second-degree relative is one of the strongest risk factors for epithelial ovarian cancer. However, only about 7% of women with ovarian cancer report a family history of ovarian cancer disease.[27] Of these, the majority will have only one affected relative; however, a small group of women, perhaps 1% of all those with ovarian cancer, will report a more extensive family history of ovarian and certain other cancers.[28]

Data from case–control studies suggest that the risk of ovarian cancer for a woman with one first-degree relative with ovarian cancer is about three times the average risk.[22,27,29,30] Cohort studies, which are less susceptible to errors such as recall bias, of the incidence and mortality of ovarian cancer in relatives of women with ovarian cancer indicate

a slightly lower risk of about twice the risk compared with women with no family history.[31,32]

Data on the risk of developing ovarian cancer for women with more than one affected close relative are much more sparse; however, the risk is estimated at about 10–15% of the risk of developing the disease by age 70 years.[28,31] This is about ten times the risk in a woman with no family history.

In perhaps half of these families, the pattern of cancers suggests the presence of a dominantly inherited gene conferring susceptibility to ovarian cancer and cancers at other sites. Three distinct clinical patterns of hereditary ovarian cancer are recognised:[28]

- ovarian cancer with breast cancer
- ovarian cancer with colorectal, endometrial, stomach and, possibly, pancreatic cancer (the 'Lynch II' syndrome)
- site-specific ovarian cancer syndrome.

Identification of the genetic mutations responsible for these syndromes is a complex and rapidly evolving field. Several predisposing genetic loci have been identified through genetic linkage studies. One of these, BRCA1, has been cloned and appears to act as a tumour suppressor. Mutations in this gene are thought to account for the majority of breast-ovarian cancer families and also for many apparently site-specific ovarian cancer syndromes. It is estimated that carriers of the BRCA1 gene may have a risk of up to 60% of developing ovarian cancer by the age of 70 years.[33] A second gene associated with breast-ovarian cancer, BRCA2, has also been identified, as have a number of genetic loci which may account for some of the Lynch II families.[33]

The identification of these genes raises the possibility that eventually the ability to test individuals in these families may be possible in order to establish whether or not they have inherited the gene and, thus, to assess more accurately their risk of developing ovarian or other cancers. However, this is a complex and resource intensive process, and currently risk assessment is based mainly on a detailed family history.

Chapter 2

Screening for ovarian cancer

Principles of screening

The aim of screening is to reduce mortality and morbidity from ovarian cancer by detecting it at an earlier stage when treatment may be more effective. Any potential beneficial effect of screening is indirect and dependent on a causal chain of events. A screening test must be performed which indicates an increased probability of the disease; this must be followed-up by further assessments to confirm the diagnosis and lead to earlier treatment, which must then result in improved survival.[34] Screening also has harmful effects, related to any risks of the screening and diagnostic process, and to the extent to which women without the disease have abnormal test results leading to unnecessary further investigations. In particular, screening may lead to diagnosis of tumours of borderline malignancy which may not have been clinically detected during the woman's lifetime, and this may result in over-treatment. There is also the possibility that detection and treatment of such borderline malignancies becomes regarded as a goal in itself. In screening, because a healthy population is tested in order to detect the small percentage of individuals who have pre-clinical disease, any harms resulting from screening may be experienced by a much larger number of people than would potentially benefit from it. Deciding whether screening is worthwhile involves assessing the balance between benefits and harms.

This balance of benefits and harms is related to the ability of a screening test to distinguish between women who have ovarian cancer and those who do not. This can be expressed as the sensitivity and specificity of the test. When the test is undertaken, four outcomes are possible *(Table 3):*

(i)   the test correctly identifies women with the disease (true-positive, a)
(ii) the test is positive when in fact the woman is healthy (false-positive, b)
(iii) the test is negative when in fact the woman has cancer (false-negative, c)
(iv) the test is negative and the woman does not have the disease (true-negative, d).

The sensitivity of the test expresses its ability to correctly identify women with the disease and is

TABLE 3 Calculating the performance of a screening test

	Disease: present	absent	
Test: positive	a	b	a + b
negative	c	d	c + d
	a + c	b + d	a + b + c + d

Sensitivity (proportion of those with the disease testing positive) = a/(a + c)

Specificity (proportion of those without the disease testing negative = d/(b + d)

calculated as the proportion of those with the disease who are detected by screening. The specificity expresses the test's ability to correctly identify healthy women, calculated as the proportion of those without the disease who screen negative.

A screening test which discriminates well between diseased and healthy women has a high sensitivity and specificity. The two parameters are, however, interdependent and vary according to the threshold used to define a positive result. A low threshold, resulting in high sensitivity, will categorise more women without the disease as positive. A higher threshold will reduce the number of these false-positives, thus increasing the specificity of the test but at the expense of missing more women with the disease and therefore resulting in lower sensitivity.

Only women who have ovarian cancer which is detected by screening have the potential for their outcome to be improved. The magnitude of any potential benefits of screening depends on the extent to which treatment is more effective in these women, and on the sensitivity of the screening test to identify ovarian cancer. In a programme in which screening is at regular intervals, the number of cancers detected by screening will further depend on the interval between each screening round and the length of any preclinical phase of ovarian cancer. A rapidly developing cancer has less chance of being detected by screening at any given screening interval.

Screening is offered to otherwise healthy women who have not sought medical help, as they do in

clinically presenting disease; hence, there is a particular duty to minimise any harm done. The major sources of potential harm are any adverse effects of the screening tests and the risks of unnecessary investigations in women with false-positive results. The number of women affected in this way depends on the specificity of the screening test – the more specific the test, the lower the proportion of false-positives.

The possible harms of screening are presented in the box below. These can be described in terms of the psychological adverse effects and the risks of morbidity or mortality associated with diagnosis. In the case of ovarian cancer screening, some women will be recalled for further assessment, most of whom will not have cancer but who may experience a period of anxiety before being told that this is the case. Among this group there will be a small number in whom cancer will arise subsequently (false-negatives) and they may experience resentment and disillusionment. A number of women initially screened positive will need to undergo invasive investigations, with the associated risks of surgery, but will be found not to have cancer. A further group of women will have their cancer detected but the prognosis will be unchanged despite earlier treatment; the harm for this group will be the extra time for which they have had to live with a cancer diagnosis.

Assessing the value of a screening test involves balancing the harms and benefits experienced by different people. If there is no improvement in outcome for women with ovarian cancer detected by screening (true-positives), then screening is clearly ineffective. If a beneficial effect is demonstrated, however, this must be weighed against the magnitude of harmful effects and the number of women experiencing these effects. Finally, if the benefits are judged to outweigh the harms, then the resources needed to produce these benefits must be considered, since greater benefits might

result if these resources were used in some other way.

The major determinants of resource use of a screening programme are the equipment, staff and training needed to set up and maintain the programme. The total direct costs will depend on the overall numbers and costs of the screening tests, follow-up tests and diagnostic tests, and these will be influenced by the screening frequency and the number of women invited for screening. This is not an exhaustive list; other potential costs include the establishment of national standard setting bodies, legal costs, the costs of holding official enquiries when standards are not met and the cost of research into new methods. A full economic analysis should consider costs and benefits to the user as well as to the health service, and should compare a variety of screening options with the option of no screening, as illustrated in the box (see right).

Evaluating screening

The potential benefits and harms of screening can only be reliably estimated in a randomised controlled trial (RCT). This allows a direct measurement of the effect of screening on the length and quality of life, by comparing screened and unscreened populations who are otherwise similar.

Prospective screening studies, which resemble the intervention arm of an RCT but have no control group for comparison, can be used to measure outcomes which only occur in screened populations. These include the false-positive rates and the outcomes experienced as a result of these false-positives. With adequate follow-up, such studies can also estimate the sensitivity of screening at varying screening intervals. The costs of screening can also be estimated. However, these studies cannot be

Some factors influencing the benefits and harms of screening at a population level	
Factors influencing benefits	**Factors influencing harms**
- Effectiveness of treatment for early disease compared with advanced disease - Ability of test to detect early disease (test sensitivity) - Screening interval and duration of preclinical phase of ovarian cancer - Prevalence of ovarian cancer in screened population - Size of target population - Uptake of screening	- Adverse effects of screening tests - Proportion of screened women recalled for further assessment - Anxiety/distress experienced by these women - Proportion of screened women undergoing diagnostic interventions (false-positives – test specificity) - Morbidity/mortality experienced by these women - Size of screened population - Uptake of screening

Examples of direct health service costs of screening and no screening options	
Screened population	**No screening**
• Initial screening test (all women) • Repeated test/secondary test (subgroup of women – possible multiple recalls) • Diagnostic surgery • Treatment of screen-detected cancers • Diagnosis and treatment of cancers presenting clinically • Administration of follow-up and fail-safe procedures • Quality assurance and audit teams • Initial and continuing training • Costs of long-term follow-up of those with screen-detected abnormality	• Diagnosis and treatment of cancers presenting clinically • Possible treatment of benign conditions which may be averted by screening

used to evaluate potential benefits of screening, because survival may appear to be improved in screened women simply because a cancer has been detected earlier (lead-time bias). Furthermore, screening may preferentially identify slower growing tumours which have an inherently better prognosis (length bias).

In the absence of direct evidence on the benefits of screening, indirect evidence may be used to estimate possible benefits in a model of screening. This involves using data on the sensitivity and specificity of the screening tests, the effect of screening on stage at diagnosis, the effect of earlier treatment, and so on. The many assumptions usually made in constructing such models may lead to inaccuracies and their effects should be investigated using sensitivity analyses.

The decision as to whether screening is worthwhile depends on the overall benefits and harms of screening and the resources required. There are various ways of obtaining this information – by calculating, for example, the costs per life saved, the costs per life-year gained, or the costs per quality adjusted life-year. These methods subtract the negative effects on health from the positive effects on health to obtain an overall summary of the health outcomes 'produced' by screening. It may also be helpful to consider positive and negative effects separately, to enable an assessment of the distribution of benefits and harms. Screening may result in large benefits for a small number of people, with larger numbers affected by smaller negative effects.

There may also be a range of opinions among women and health professionals regarding the balance of benefits and risks for which they would consider screening worthwhile; formulating a

policy about screening needs to take account of these different perspectives.

Screening methods for ovarian cancer

Ultrasonography

Ultrasonography uses imaging of the ovaries to detect changes in size and shape which may indicate abnormality. Ultrasound scanning may be performed transabdominally or transvaginally. The size of ovaries measured by the two techniques is similar but more detail of the ovarian morphology can be obtained using the transvaginal route; hence, this has become the preferred method.[35,36] The transvaginal route also removes the need for women to have a full bladder on scanning, which may increase acceptability of the procedure. Ultrasonography is carried out by trained technicians, radiographers or physicians and, on average, each examination takes about 15 minutes.[37] Because ultrasound scanners are expensive and bulky, examination takes place at a central facility where the scanner is installed.

Changes in size and shape of ovaries can be transient or reflect normal physiological events, particularly in premenopausal women. After the menopause the ovaries are smaller and tend to reduce in size with age.[36] Criteria for defining an abnormally enlarged ovary therefore vary with menopausal status and age.

Persistently enlarged or abnormal-looking ovaries can occur as a result of benign or malignant tumours or tumour-like conditions. There are no universally-accepted criteria for distinguishing between benign and malignant conditions on the basis of ultrasound findings, although many

authors have described systems for classifying morphological abnormalities,[38,39] while others have attempted to derive numerical scoring systems which would provide a more objective way of identifying ovarian malignancies.[40] Common parameters included in such classification systems include the size of the ovary, the number of locules in cystic masses and the uniformity of echogenicity of solid masses. Some types of abnormal morphology, such as papillary projections into a cyst, are considered highly suspicious,[39] while many simple cysts either resolve or remain stable over long periods.[41]

A more recent technique which may be of use in distinguishing between benign and malignant ovarian abnormalities is the use of colour Doppler® imaging. This is used in conjunction with grey-scale ultrasonography, and enables visualisation of ovarian blood vessels and characterisation of the pattern of blood flow. Malignant tumours induce the formation of new blood vessels; these appear disorganised and have reduced smooth muscle in their walls, which leads to reduced resistance to blood flow and high flow velocity. A variety of methods may be used to calculate the velocity of blood flow but there is wide variation in the extent to which these methods have been found useful for discriminating benign from malignant masses.[42]

Cancer antigen 125

Cancer antigen 125 (CA 125) is a glycoprotein produced by some ovarian cancers.[43] Levels of CA 125 in serum can be measured by means of a blood test and laboratory assay of the serum. The test can therefore be undertaken at any suitable location and by any personnel trained in venepuncture.

Elevated levels of CA 125 have been reported in 61–96% of all clinically diagnosed epithelial ovarian cancers and in 29–75% of cancers diagnosed at Stage I.[44] Elevated levels have also been reported in other malignancies, for example, in endometrial and pancreatic cancer, and in a variety of benign gynaecological conditions, such as endometriosis, uterine leiomyoma (fibroids) and pelvic inflammatory disease.[45] Levels of CA 125 in healthy women vary with menopausal status and past history of hysterectomy.[46]

Studies of CA 125 levels in stored blood samples from population-based serum banks and observational cohort studies indicate that raised levels of CA 125 can occur many years before the clinical diagnosis of ovarian cancer; furthermore, over 95% of women who do not develop ovarian cancer do not have elevated levels of CA 125.[47–51] Serial measurements of levels of CA 125 suggest that women with ovarian cancer demonstrate rising levels, while elevated levels associated with other conditions may remain stable over time.[52]

When used for screening for ovarian cancer, CA 125 measurement is used in conjunction with an ultrasound scan. Ultrasound may be performed at the same time as blood is taken for CA 125, or women with elevated or rising CA 125 levels may be recalled for ultrasound scanning.

The screening process

Screening for ovarian cancer involves a number of stages. The initial test, either ultrasound or CA 125, is performed on all women. The findings of this initial screening test then determine whether the woman is recalled for further assessment, which may consist of a number of further stages. The initial test may be repeated one or more times in order to establish whether abnormalities have resolved, and a secondary test may be performed, such as ultrasound screening with colour Doppler imaging or ultrasound screening following initial CA 125 measurement. Women who have persistent abnormal findings at the end of this process are then referred for a definitive diagnosis to be made. For ovarian cancer, this involves an invasive surgical procedure, usually an open or laparoscopic oophorectomy. This enables ovarian tissue to be removed and examined histologically to confirm whether or not a malignant tumour is present. Women diagnosed with ovarian cancer will then require further surgery for accurate staging of the disease and removal of the tumour mass. Treatment may also involve chemotherapy or other adjuvant therapy.

Chapter 3

Methods

Objectives

The purpose of this review is to provide the NHS Health Technology Assessment programme with an overview of the results of research evaluating screening for ovarian cancer.

The specific objectives of the review are:

- to evaluate the performance of the current screening tests for ovarian cancer
- to assess the adverse effects of screening, including morbidity associated with surgical intervention and the psychological morbidity associated with false-positive diagnosis
- to report on the stage of development of newer methods of screening
- to investigate the potential cost-effectiveness of screening in different risk groups.

In addition, the review identifies those issues which require further research and the degree to which research in progress is likely to address these issues.

Sources

The review was undertaken using structured guidelines for systematic reviews.[53] A comprehensive search for studies and reviews evaluating screening tests was conducted to address the main objective of the review – an assessment of the performance of screening tests. Supplementary searches were performed specifically to address additional objectives.

Evaluation of screening test performance

A sensitive search strategy for studies evaluating screening tests was used (see Appendix 1). The following databases were searched: Current Contents, Medline (1966–May 1997), computerised Embase (1982–May 1997), the nursing database CINAHL (1982–97), the Cochrane Register of Controlled Clinical Trials (Issue 3, 1996) and Cancerlit (1966–May 1997). Researchers and experts in the field, and consultants to the review were also contacted with a view to identifying any unpublished studies. In addition, the bibliographies of literature reviews in the area were used as sources of relevant studies. Conference proceedings were identified through Cancerlit.

Assessing the adverse effects of screening

This involved a search of Medline (1982–97), Embase (1982–97), CINAHL and Psychlit (1974–97). Two search strategies were used: one to identify case series of surgical procedures similar to the diagnostic procedures used in screening for ovarian cancer and the other to identify research on the psychological aspects of screening (see Appendix 1).

Investigating screening methods under development

Formal systematic review procedures were not considered appropriate to address this issue. Relevant information was identified from the three sources listed below.

1. The authors of all studies identified in the main review, together with other known researchers in the field, were contacted to obtain information on new developments.
2. A major international workshop (Ovarian Cancer Screening International Meeting, Royal College of Obstetricians and Gynaecologists, 21–22 April 1997) was attended.
3. Data were extracted from any relevant abstracts identified during the main search for studies evaluating screening tests.

On the basis of the above, a view of likely future developments in this field was obtained.

Assessment of the cost-effectiveness of screening for ovarian cancer

Searches for economic evaluation studies were conducted in Econlit, Medline, and the NHS Economic Evaluation Database (see Appendix 1). In addition, studies reporting cost information were identified from the main search.

Inclusion criteria

Evaluation of the performance of screening tests

Studies which prospectively evaluated a test or a combination of tests to detect ovarian cancer in asymptomatic women were eligible for inclusion. Three criteria were used to define studies eligible for inclusion.

1. The women included in the study should be asymptomatic (i.e. not presenting clinically with symptoms suggestive of ovarian cancer).
2. The test should be performed before the diagnosis is known.
3. Women testing positive should be followed-up with diagnostic surgery to establish whether they have ovarian cancer.

Only prospective screening studies were included, so that estimates of sensitivity and specificity would be directly applicable to the use of the test in a screening situation. RCTs of screening were also eligible for inclusion under these criteria.

Many studies have evaluated the performance of these tests in detecting cancer in women already scheduled for surgical investigation. However, because these studies include women with clinically apparent ovarian abnormalities, they are likely to overestimate sensitivity compared with the use of the test in asymptomatic women. This type of study was therefore excluded from the review.

Three reviewers independently assessed the retrieved abstracts and titles for relevance, and the full versions of selected papers were independently assessed for inclusion by two reviewers. Multiple publications of single studies were included only once, with relevant data extracted from several separate papers where necessary. Studies in any language were considered for inclusion.

Assessing the adverse effects of screening

Studies eligible for inclusion were those which reported information on the surgical complications of the procedures used in diagnosing ovarian cancer, such as open or laparoscopic oophorectomy, and studies in which the psychological outcomes of screening for ovarian cancer were reported. These were identified both from the main search for studies evaluating screening tests and from the specific searches outlined above.

As the literature on psychological adverse effects was known to be very limited, any study reporting psychological effects of ovarian cancer screening in the general population or in women at high risk was included. With respect to adverse effects of surgery, only studies with more than 50 patients were included (e.g. case series reporting complications associated with oophorectomy or large comparative studies). Studies of oophorectomy carried out at the same time as hysterectomy and

studies of the long-term effects (e.g. osteoporosis and depression) of oophorectomy were excluded.

Assessment of the cost-effectiveness of screening for ovarian cancer

Any study reporting cost data for ovarian cancer screening was eligible for inclusion. This included economic evaluations, cost-effectiveness studies (including cost-minimisation and cost-consequences analyses), cost–benefit analyses and costing studies. In addition, any cost data reported in studies of the performance of screening were recorded.

Data extraction and assessment of study validity

Evaluation of the performance of screening tests

Data were extracted from studies meeting the inclusion criteria by one reviewer using a standard data extraction form and checked by a second reviewer. Authors were contacted for additional data, if appropriate. Information was extracted relating to the study population, all relevant details of the screening protocol, methods of follow-up and the outcomes of screening in terms of the number of women recalled, the number screened positive and the number with ovarian cancer. The data extraction form is presented in Appendix 2.

These data were then used to calculate summary statistics for each study – the prevalence of cancer detected in the screened population; the sensitivity, specificity and the probability of having ovarian cancer at diagnostic intervention (i.e. the positive predictive value (PPV)); and the false-positive and recall rates.

Information was also recorded relating to the methodological quality of each study, based on criteria recommended by the Cochrane Methods Working Group on systematic reviews of screening and diagnostic tests.[54] Information relating to the following methodological issues was recorded.

- The method and completeness of follow-up of women screened negative, which affects the reliability of estimates of false-negatives and, hence, test sensitivity.
- The clarity of cut-off points and explicitness of the description of the protocol. This affects the generalisability of the study and may also influence the reliability of the estimates of the outcomes of screening (numbers of true- and false-positives, etc.).

- The completeness of result reporting including drop-out rates at each stage of screening. This affects the reliability of the estimates of the outcomes of screening, particularly the false-positive rates if a significant proportion of women have not completed the screening process at the time of reporting.
- The description of the study population with respect to major risk factors. This may affect the generalisability of the results to other populations.

These quality criteria were not used to obtain an overall quality score, because they affect the validity of different aspects of the study (for example, the quality of the follow-up of women screened negative primarily influences the estimate of test sensitivity). Instead, these factors were considered separately in assessing the validity of each study in relation to the different outcomes investigated.

Chapter 4
Results from published studies

Studies identified

A total of 25 separate prospective studies of ovarian cancer screening in apparently healthy women were identified which fulfilled the inclusion criteria. A list of studies which were assessed and judged not to meet the review's inclusion criteria is presented in Appendix 3, together with the reasons for their exclusion. The most frequent reasons for exclusion were that the women screened were not asymptomatic (7 studies) or that the study was not a prospective investigation of test performance, with definitive diagnosis in those testing positive (20 studies). A total of 11 articles were excluded because they duplicated publication of data relating to the same women.

The 25 studies included were all prospective screening studies, in which women were screened for ovarian cancer but no comparisons were made with unscreened women. Details of the study designs and results are summarised in Appendix 4. The search also identified three on-going RCTs of screening for which no results have so far been published.

Appraising the information available from prospective screening studies

Uncontrolled screening studies cannot provide reliable evidence concerning the effect of ovarian cancer screening on health outcomes such as mortality and quality of life. However, outcomes which only occur in screened women, such as the risks of screening, can be measured, together with information about the performance of the screening tests in discriminating between women with and without cancer.

Measurement of the sensitivity and specificity requires comparison of the screening test against the best available reference standard for the diagnosis of ovarian cancer – the histological examination of ovarian tissue removed from the woman. However, this is itself subject to inter- and intra-observer variability in interpretation. Also, since the histological examination involves an invasive procedure in a screening study performed on apparently healthy volunteers, this information can only be obtained for those testing positive. Only the numbers of true- and false-positives can therefore be directly measured with reference to histological diagnosis. The numbers of true- and false-negatives cannot be directly observed at the time of screening but can be estimated by following-up women who have screened negative and measuring the subsequent occurrence of ovarian cancer. This is illustrated in *Figure 2.* The number of false-negatives therefore increases with increasing duration of follow-up, and the sensitivity estimates obtained measure the ability of screening to detect cancers which would otherwise become clinically apparent within a defined period. It is not possible to distinguish whether these false-negatives were present at the time of screening or developed some time after screening.

The accuracy of the estimates of sensitivity and specificity obtained from these studies depends on the completeness and accuracy of the data reported. The information in the boxes with broken lines in *Figure 2* indicate the points at which incomplete follow-up data can introduce inaccuracies into the estimates of sensitivity and specificity. If a large number of women do not complete the screening process, either because they have chosen not to attend or because, at the time of reporting the study, they are still undergoing further tests, then the estimates of true- and false-positive rates will be inaccurate. Similarly, the accuracy of estimates of the number of false-negatives will be affected by the quality of follow-up of women who screened negative.

Also illustrated in *Figure 2* are the various stages of the screening process: the application of an initial screening test does not lead directly to diagnostic surgery but is preceded by an intermediate stage at which women with an initial positive or equivocal test result are recalled for further assessment. This further assessment may involve a number of repeated tests, or the use of different tests, before a final decision on whether to refer the woman for surgery can be made. Women undergoing further assessment may experience considerable anxiety as they await the results of further tests; hence, the number of women affected in this way needs to be considered when assessing the impact of screening. Repeated tests will also add to the costs of the screening process.
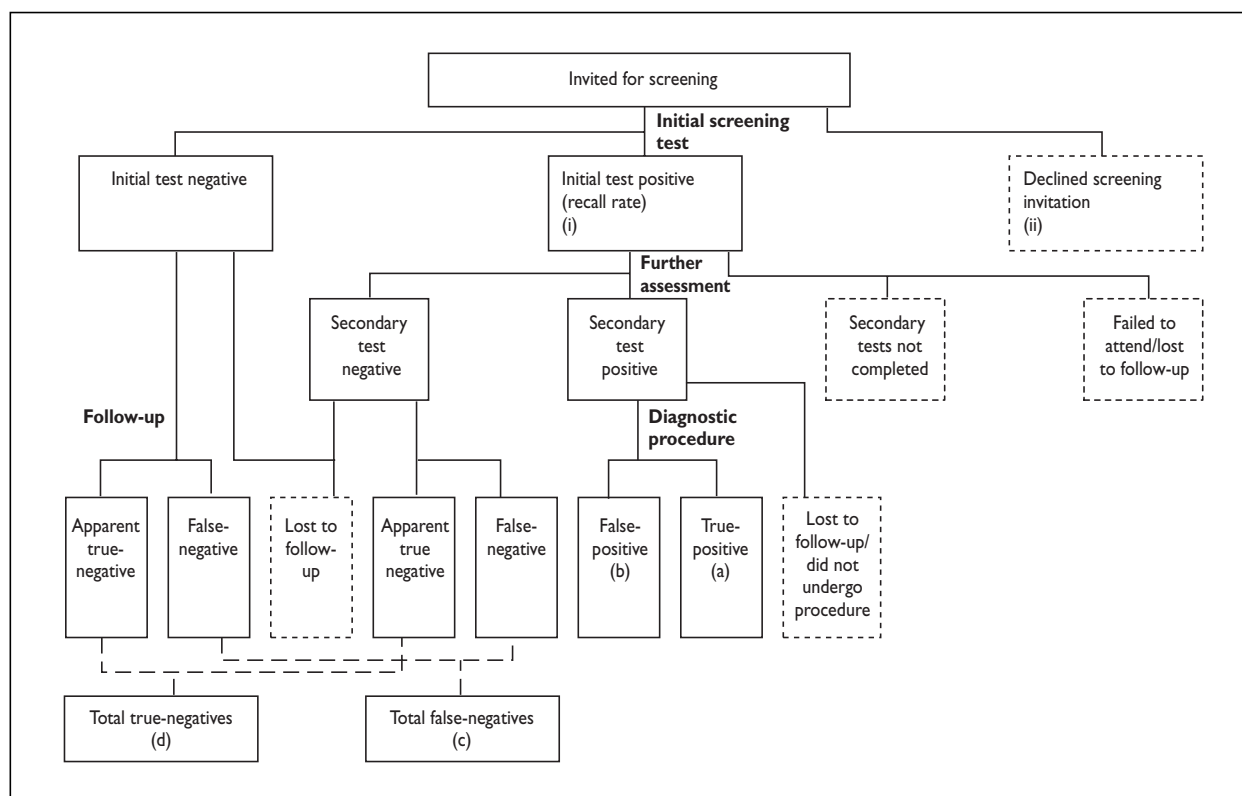
FIGURE 2 *Stages in the screening process*

Study populations and sample size

Of the 25 studies identified, 16 screened women who were at average risk for their age of developing ovarian cancer; hence, their results may be relevant to general population screening. A variety of recruitment methods were used in these studies. In two studies, a random sample from a population register was invited,[55,56] while in two further studies women already attending other screening programmes were invited to participate.[36,57] These studies reported uptake rates for screening of between 50% and 74%. In the remaining studies self-selected volunteers, who responded to publicity about the study, were the main source of recruitment or the studies did not record their recruitment methods. Such studies do not allow calculation of an uptake rate or assessment of the extent of selection bias and, if their subjects differ from the general population in their risk of ovarian cancer, the results may be misleading.

The 16 general population studies all stipulated a lower age limit for eligibility, which ranged from 18 years to 55 years, but most frequently was either 45 or 50 years of age. Several studies restricted entry to postmenopausal women and some had other exclusion criteria, such as previous hysterectomy (see Appendix 4).

Of the nine further studies recruiting women at higher-than-average risk of developing ovarian cancer, seven investigated screening in women with a family history of ovarian or certain other cancers. The precise inclusion criteria varied for each study. On average, these studies included younger women than the general population studies. All of those on women with a family history were performed on volunteers responding to publicity or referred by their doctors.

The size of the studies varied between 435 and 22,000 women for general population samples and between 137 and 1601 for studies on high-risk populations. The median size of the studies undertaken in a general population was 2572 women. If the incidence of ovarian cancer in this type of population is about 1 in 2500 per year (comparable to that in England and Wales for women over the age of 40 years), then a study of this size would expect to detect, on average, perhaps one or two cancers on initial screening.

Screening methods

The studies identified evaluated a wide variety of screening tests and combinations of tests. The numbers of studies using each method of screening are shown in *Table 4*.

TABLE 4 *Number of studies using each screening method*

Screening method		Number of studies	
Initial screening test	Follow-up test*	General population	High-risk population
TAS/TVS	–	6	1
TVS with CDI	–	2	1
TVS	CDI	1	1
TVS	FNA/B	1	–
TVS	CA 125 or other markers	2	–
CA 125	TVS/TAS	4	–
TVS with CA 125	–	–	6

Test performed only on women with positive result on initial test for further assessment; women positive on follow-up testing referred for diagnostic surgery.

TAS, transabdominal sonography; TVS, transvaginal sonography; CDI, colour Doppler® imaging; FNA/B, fine needle aspiration cytology/biopsy.

In addition, four studies also included a comparison with pelvic examination.[58–61]

Studies using the same modality of screening test did not necessarily use the same criteria for defining positive results. Most of the studies using grey-scale ultrasound (transabdominal or trans-vaginal sonography) as an initial test used a combination of ovarian volume and morphological criteria to define abnormal results, although there were differences in the detailed definitions of abnormalities. For colour Doppler imaging, the criteria used to define positive results were much less consistent, with a variety of parameters and cut-off points used, including pulsatility index, resistance index and peak systolic flow. Studies using the CA 125 test, followed by ultrasound scanning in those with elevated levels ('CA 125-based screening'), used either 30 U/ml or 35 U/ml as the cut-off for abnormal results. Not all the studies specified the definitions of abnormal results or the full screening protocol used (see Appendix 4).

The studies demonstrated considerable hetero-geneity, both in the study populations and in the screening tests used. Furthermore, many studies did not permit measurement of test sensitivity (see below). It was therefore considered inappropriate

to attempt to calculate pooled measures of the performance of the various screening tests or to construct summary receiver operating curves. Instead, a qualitative summary of the findings is presented, including a discussion of the validity of the individual studies.

Sensitivity of screening tests

The sensitivity of a screening test is that pro-portion of women with cancer (true-positives and false-negatives) who are correctly identified by the test. As discussed earlier, in this type of study the number of cases missed can only be estimated by follow-up of those women who screened negative to see if they subsequently developed clinical ovarian cancer.

Of the 25 screening studies identified, only six reported a method for follow-up of women who screened negative that was reasonably complete (better than 85% response) at least 1 year after screening.[35,56,60,62–64] The details and results of this follow-up, together with the estimated sensitivity of screening, are presented in *Table 5.* Information is also presented for three further studies in which cancers arising in screen-negative women were reported or it was stated that no such cancers had arisen, but no description was given of the method or completeness of follow-up.[36,59,65]

Pooled estimates of sensitivity are not given because it is inappropriate to consider sensitivity without reference to the specificity achieved, since the two parameters are interdependent.[54] Furthermore, in each study either a different screening method was used or a different population was recruited. The individual sensitivity estimates were also very imprecise. For these reasons, a summary receiver operating curve was not constructed to summarise test sensitivity and specificity.

No studies using ultrasound as an initial test reported any ovarian cancers arising clinically within 1 year of a negative screen. Two such studies followed-up women for more than 1 year, and both reported cancers arising clinically 24 months after screening.[35,56] One of these reported follow-up information at 4 years, which indicated a sensitivity of 60%.[35] One study reported the results of three screening rounds at approximately 18-month intervals, with no interval cancers found.[62]

The small size of the studies and the lack of long-term follow-up limits the conclusions that can be drawn regarding test sensitivity. However, it appears

TABLE 5 Sensitivity of screening tests at 1-year follow-up

Study	Sample	Test	Number of cancers detected at screening	Method of follow-up	Number of cancers arising in women screened negative	Sensitivity after 1-year follow-up (Exact 95% CI)
Campbell et al.[62]	5479 Aged > 45 years or with family history.	TAS	5 (after 3 screening rounds)	89% of women contacted at 1 year.	None	100% (48–100)
Vuento et al.[56]	1364 Aged 56–61 years, eligible for mammography screening.	TVS + CDI	1	Finnish cancer registry.	None at 1 year; one at 2.5 years.	100% (3–100)
Bourne[35]	1601 Family history, mean age, 47 years (range, 17–79 years).	TVS then CDI	6	100% of women contacted between 6 and 16 months following screening.	None at 1 year; four at 4 years.	100% (54–100)
Schincaglia et al.[64]	3541 Postmenopausal, aged 50–69 years.	TVS then FNA/B	2	Cancer registry and annual questionnaire – 100% complete.	None at 1 year.	100% (16–100)
van Nagell et al.[65*]	8500 Aged 50+ years and postmenopausal, or 25+ years with family history.	TVS	8	Not stated.	One at 1 year (discovered at surgery).	88% (47–100)
Parkes et al.[36*]	2953 Aged 50–64 years.	TVS then CDI	1	Not stated.	None at 1 year; one at 19 months.	100% (3–100)
Jacobs et al.[60]	1010 Postmenopausal, aged 45+ years (mean, 54 years).	CA 125 then ultrasonography	1	Postal questionnaire, 100% response.	None at 1 year.	100%
Jacobs et al.[63]	22,000 Postmenopausal, aged 45+ years (median, 56 years).	CA 125 then ultrasonography	11	Postal questionnaire, 99% response at 1 year, 57% at 2 years.	Three at 1 year; eight at 2 years.	73% (39–94)
Adonakis et al.[59*]	2000 Aged 45+ years (mean, 58 years).	CA 125 then ultrasonography	2	Not stated.	None at 1 year.	100% (16–100)

* *Studies with poor details of follow-up.*

TAS, transabdominal sonography; TVS, transvaginal sonography; CDI, colour Doppler imaging; FNA/B, fine needle aspiration cytology or biopsy.

that significant numbers of interval cancers arise in the first year following CA 125 screening, while for ultrasound, interval cancers have appeared at about 18 months after screening. The suggestion that ultrasound screening alone is more sensitive than CA 125 followed by an ultrasound scan is supported by one study in which the performances of these two screening methods were compared in the same cohort of women.[66] This is much more valid than comparisons between different studies, because it eliminates differences caused by the individual characteristics of the women – in effect, each woman acts as her own control. In this analysis, three of the six cancers (one of three invasive tumours and two of three borderline tumours) would have been missed using a CA 125

cut-off of 30 U/ml or 35 U/ml as an initial screen, suggesting that, at least in this population of women with a family history of ovarian cancer, ultrasound is considerably more sensitive.

Because of the limited follow-up information, and the fact that only one study has reported results for successive screening rounds, little can be inferred about appropriate screening intervals or the natural history of ovarian cancer. A more sensitive screening test is likely to require less frequent screens to detect the same proportion of cancers. The rate of appearance of clinical ovarian cancer following screening in these studies suggests that screening intervals between 1 year and, perhaps, 3 years merit further investigation.

Stage at diagnosis of screen-detected cancer

Measures of test sensitivity do not necessarily indicate the likelihood or extent to which screening will detect cancers earlier. A more significant observation would be an increase in the proportion of cancers detected at an early stage. Although this would not in itself demonstrate that screening is effective, evidence to the contrary would suggest that screening is unlikely to improve outcomes.

The stage distribution of the cancers detected in the general population screening studies is shown in *Table 6.* One study has been excluded because stage at diagnosis was not reported.[67] The individual studies are relatively small and only a few cancers were detected in each; hence, the confidence intervals for the proportion diagnosed at Stage I are extremely wide. A more precise estimate can be obtained by calculating the average proportion diagnosed at Stage I; this is shown in *Table 6* separately for ultrasound-based screening and for CA 125-based screening. In studies using CA 125 followed by ultrasound, 50% (95% CI, 23–77) of cancers detected were at Stage I and in studies using ultrasound as an initial screening test, 61% (95% CI, 38–80). In comparison, data from cancer registries indicates that, in the largely unscreened UK population, the proportion of

TABLE 6 *Stage at diagnosis and prevalence of screen-detected cancer in general population studies*

Study	Number screened	Number of cancers (of which borderline tumours)	Prevalence of screen-detected cancer (95% CI) per 100 000	Percentage diagnosed at Stage I (95% CI)
Ultrasound-based screening				
Goswamy *et al.*[68]	1084	1	92	100
Millo *et al.*[57]	500	0	0	–
Campbell *et al.*[62] (first screen)	5479	2 (1)	36	100
Demidov *et al.*[69]	11,996	11	91	36
van Nagell *et al.*[65]	8500	8	94	75
Tabor *et al.*[55]	435	0	0	–
Kurjak & Predanic[70]	5013	4	80	100
Vuento *et al.*[56]	1364	1 (1)	73	100
Parkes *et al.*[36]	2953	1	34	100
Schincaglia *et al.*[64]	3541	2	56	0
Holbert[71]	478	1	210	100
All ultrasound studies which appear to report only prevalence screen *	**32,843**	**23 (2)**	**70 (41–98)**	**61 (38–80)**
All ultrasound studies where it is clear that only the prevalence screen is reported **	**15,834**	**8 (2)**	**51 (16–90)**	**75 (35–97)**
CA 125 followed by ultrasound				
Jacobs *et al.*[60]	1010	1	99	100
Jacobs *et al.*[63]	22,000	11	50	36
Grover *et al.*[61]	2550	0	0	–
Adonakis *et al.*[59]	2000	2 (1)	100	100
All CA 125 studies	**27,560**	**14 (1)**	**51 (24–78)**	**50 (23–77)**

* *Excludes van Nagell* et al.; ** *excludes van Nagell* et al., *Demidov* et al., *Kurjak & Predanic.*

cancers diagnosed at Stage I is about 22–28%. This suggests that the evidence is consistent with some improvement in stage at diagnosis for screen-detected cancers, which may be greater for ultrasound-based screening.

A number of issues should be considered in assessing the significance of these findings. The first time that a population is screened, the stage distribution will reflect the prevalence of advanced and early cancers in that population and, thus, the proportion of early cancers may be lower than in subsequent screening rounds. Therefore, in an attempt to ensure comparability, only studies reporting initial screening rounds have been included in the above calculation, which might underestimate the potential impact of repeated rounds of screening on stage at diagnosis. Not all studies explicitly stated that they were only reporting prevalence screening, and it is therefore possible that some results relate to several screening rounds reported together. Furthermore, an accurate estimate of the stage shift resulting from screening should include all the cancers arising in the screened population, including those 'missed' by screening. This was not possible because of the limited follow-up information available.

Many of these studies gave few details of their recruitment methods, making it difficult to assess whether the women were truly representative of the general population. If those women presenting for screening who had signs that were suspicious of cancer were excluded from the reported results of screening, or if women with more advanced cancers were less likely to volunteer for the studies, the proportion of screen-detected cancers diagnosed at Stage I may be misleadingly high.

A further potential source of error is the classification of borderline tumours. These are those with features intermediate between benign tumours and frankly invasive cancers; they have a good prognosis, are more likely to be detected at an early stage and are not thought to be precursors of more aggressive cancers. The classification of borderline tumours varies; hence, the screening studies may not have reported them in a comparable way. However, the proportion of screen-detected cancers reported to be borderline tumours was less than 10%, suggesting that the over-diagnosis of such tumours may not be a significant problem. Excluding such tumours slightly reduces the proportion of screen-detected Stage I cancers.

The proportion of tumours diagnosed at Stage I in studies of women at high risk are shown in *Table 7*. In these studies a much larger proportion were reported to be borderline tumours and the proportion of invasive tumours diagnosed at Stage I was only 25%. It is difficult to assess to what extent

TABLE 7 *Stage at diagnosis and prevalence of screen-detected cancer in studies screening women at high risk*

Study	Number screened	Risk group	Number of cancers (of which borderline tumours)	Percentage diagnosed at Stage I (95% CI)
Ultrasound screening				
Andolf *et al.*[58]	805	Outpatient department attenders	3 (2)	67
Bourne *et al.*[35]	1601	Family history	6 (3)	83
Weiner *et al.*[72]	600	History of breast cancer	3	33
Ultrasound with CA 125				
Karlan *et al.*[73]	597	Family history	1 (1)	100
Muto *et al.*[74]	384	Family history	0	–
Schwartz *et al.*[75]	247	Family history	0	–
Dorum *et al.*[76]	180	Family history	7 (3)	43
Belinson *et al.*[77]	137	Family history	1	0
Average for all studies	**4551**		**21 (9)**	**57 (34–78)**
Average for all studies on women with family history	**3146**		**15 (7)**	**60 (32–84)**
Average for all studies on women with family history excluding borderline tumours	**3146**		**8**	**25 (3–65)**

this may be a real difference in women at high risk or if it simply reflects the way tumours were classified in these particular studies. There is evidence for the latter explanation: one study with a high proportion of borderline tumours[35] was undertaken at the same institution, and with some of the same investigators, as one of the studies in a general population which also reported a high proportion of borderline tumours.[62]

Prevalence of screen-detected cancer

Any benefits of screening result from its ability to detect cancer before it would be clinically diagnosed and when it may be more amenable to treatment. The average time between the detection of cancer at screening and the time at which it would have been detected clinically in the absence of screening is known as the 'lead time'. It is determined by both the sensitivity of the test and the rate of growth of the cancer. The longer the lead time provided by a screening test, the greater its potential to influence the outcome in those screened positive.

The prevalence of screen-detected cancer can be used to estimate the lead time.[78] If the number of cases detected at screening is compared with the number which would be expected to present clinically per year, then the length of time it would take to clinically detect the number of cases detected at screening can be used to estimate the lead time. The prevalence of screen-detected cancer is about 50 per 100,000 *(Table 6)*. If it is assumed that the screened women would have had an average annual incidence of clinically detected cancer of 40 per 100 000 (the incidence in England and Wales in women over 40 years of age[3]), then screening detects about 1.25 years' worth of cancer cases. If there is no length bias (i.e. no tendency for screening to preferentially detect slow-growing cancers), then this is about double the average lead time. This suggests that ovarian cancer screening may result in a lead time of only 7–8 months. If ovarian cancer has a short natural history, however, this may be sufficient to produce a clinically significant improvement in outcome, although it would also imply that the screening interval would need to be relatively short.

False-positive results

Ideally, when assessing the performance of a screening test which has been evaluated in several studies, the sensitivity and specificity obtained in each study should be considered together. However, because so few studies permit a reliable estimate of sensitivity, and the studies use differing tests, thresholds for positive results and study populations, such an analysis would be of little value.

The practical significance of the specificity of a test is its relationship with the false-positive rate. This is defined as the proportion of all women without the disease (true-negatives plus false-positives) who are wrongly classified as positive on testing (false-positives). Because most of the studies do not have adequate follow-up information, the proportion of true-negatives is not known. However, because ovarian cancer is relatively rare in the general population, with an incidence of around 1 in 2500 for the age groups screened, the proportion of women without the disease is close to the total number of women screened. The false-positive rate can therefore be approximated by the proportion of screened women who are false-positives, that is, who undergo diagnostic surgery but prove not to have primary ovarian cancer. This is summarised for studies undertaken in women at average risk in *Table 8* and for women at high risk in *Table 9.*

False-positive rates varied considerably between studies. There are a number of possible explanations for this:

* random variation
* the screening method used
* the threshold used to define a positive result
* the characteristics of the study population, in particular, the menopausal status of the women
* the completeness of reporting and follow-up of women screened positive.

Some of these factors are shown in *Tables 8* and *9.*

Studies which used grey-scale ultrasound alone as a screening method have reported higher false-positive rates than either those using ultrasound with colour Doppler imaging or those using CA 125 followed by ultrasound. However, because of the differences outlined above, comparisons of screening methods between studies should be interpreted cautiously. For example, two of the studies using colour Doppler imaging reported significant numbers of women who were still undergoing follow-up and had not been definitively classified as screen-positive or -negative; this could give a misleadingly low false-positive rate.[41,79]

Further evidence that the addition of colour Doppler imaging to ultrasound screening may

TABLE 8 *False-positive rates reported in general population studies*

Study	Number screened	Population age (years) and menopausal status	False-positive rate (%) of all women screened (95% CI)
Ultrasound screening			
Demidov et al.[69]	11,996	18+	2.1 (1.8–2.4)[†]
Campbell et al.[62] (screen 1)	5479	45–78	2.5 (2.1–2.9)
Campbell et al.[62] (screen 2)	4914	45–78	1.8 (1.4–2.2)
Campbell et al.[62] (screen 3)	4201	45–78	1.2 (0.8–1.6)
Tabor et al.[55]	435	46–65	2.1 (0.9–3.9)
Millo et al.[57]	500	45+ or postmenopausal	1.2 (0.5–2.6)[†]
Goswamy et al.[68]	1084	39–78 postmenopausal	1.3 (0.7-2.1)[‡]
de Priest et al.[80]	3220	33–90 (mean, 60) postmenopausal	1.3 (0.9–1.7)
Ultrasound with CDI			
Kurjak et al.[41]	5013	40–71 (mean, 45)	0.7 (0.4–0.9)[†‡]
Vuento et al.[56]	1364	56–61 (mean, 59)	0.3 (0.1–0.8)[†‡]
Ultrasound followed by CDI			
Parkes et al.[36]	2953	50–64	0.5 (0.3-0.8)
Ultrasound followed by other secondary tests			
Sato et al.[67]	15,282	30+	0.3 (0.2–0.4)[†]
Schincaglia et al.[64]	3541	50–69	0.5 (0.3–0.8)
Holbert[71]	478	30–89 postmenopausal	1.9 (0.9–3.6)
CA 125 followed by ultrasound			
Grover et al.[61]	2550	40+ (median, 51)	0.3 (01–0.6)
Adonakis et al.[59]	2000	45+ (mean, 58)	0.6 (0.3–1.0)
Jacobs et al.[60]	1010	45+ (mean, 54) postmenopausal	0.2 (0.02–0.7)
Jacobs et al.[63]	22,000	45+ (median, 56) postmenopausal	0.1 (0.09–0.2)

[†] Criteria for positive screening result not fully reported.
[‡] Incomplete follow-up: significant numbers of women awaiting further assessment or significant numbers of screen-positive women did not undergo diagnostic intervention.
CDI, colour Doppler imaging.

TABLE 9 *False-positive rates reported in studies on women at high risk*

Study	Population age (years) and menopausal status (if given)	False-positive rate (%) of all women screened (95% CI)
Ultrasound screening		
Andolf et al.[58]	40–70	4.5 (3.2–6.1)[†‡]
Bourne et al.[35]	17–79 (mean, 47)	4.9 (3.6–6.4)
Ultrasound with CDI		
Weiner et al.[72]	20-69	2.5 (1.4–4.1)[‡]
Ultrasound followed by CDI		
Bourne et al.[35]	17–79 (mean, 47)	1.0 (0.4–2.2)
Ultrasound with CA 125		
Akulenko et al.[81]	18+	1.3 (0.7–2.2)[†]
Karlan et al.[73]	35+	1.5 (0.7–2.8)[†‡]
Muto et al.[74]	25+	3.9 (2.2–6.4)
Schwartz et al.[75]	30+ (median, 42.5)	0.4 (0.0–2.4)[†‡]
Dorum et al.[76]	18+ (mean, 43)	8.9 (5.2–14.0)
Belinson et al.[77]	23+ (mean, 43)	0.7 (0.2-4.0)[†‡]

[†] Criteria for positive screening result not fully reported.
[‡] Incomplete follow-up: significant numbers of women awaiting further assessment or significant numbers of screen-positive women did not undergo diagnostic intervention.

increase specificity, however, comes from two studies which report separately the proportion of women positive on ultrasound alone compared with the proportion positive after repeat scanning with colour Doppler imaging.[35,36] In one study, the proportion of false-positives in the first phase of the study using trans-vaginal sonography alone was 4.9%, and this reduced to 1.0% after the intro-duction of colour Doppler imaging.[35] In the second study, the proportion reported positive by trans-vaginal sonography was 3% and, following repeat scanning with colour Doppler imaging, the pro-portion actually referred for surgery was 0.47% (a further 0.3% were referred for surgery outside the screening protocol).[36] These more direct com-parisons have greater validity than comparisons between separate studies.

Similarly, direct comparison between the CA 125 test and ultrasound found that the false-positive rate reduced with increasing cut-off points of CA 125 levels, from 1.1% at 20 U/ml to 0.47% at 35 U/ml, compared with 3.8% for ultrasound alone.[66] This strongly supports the suggestion from individual studies that CA 125 measurement is more specific than ultrasound (see page 17). In this study the increased specificity of the CA 125 test was associ-ated with a lower sensitivity than ultrasonography.

Table 9 indicates that studies on populations at high risk tended to have higher false-positive rates than studies using the same screening method on popu-lations at average risk. This may reflect the gener-ally younger age group, with more premenopausal women, in these studies, or it may reflect the use of a lower threshold for defining a positive result.

Recall rates

The discussion so far has considered as false-positives only those women referred for diagnostic testing at the end of the screening process but found not to have ovarian cancer. However, many more women test positive on the initial screen and are then recalled for repeat tests than are referred for diagnostic interventions following further assessment. These women will not receive the reassurance of a negative result after attending for screening and may experience distress and anxiety while waiting for their follow-up appointments.

Not all of the studies reported the number of women recalled for further tests. The recall rate for each study which reported it is given in *Table 10.* Studies in which grey-scale ultrasonography was used as the initial screening test reported recall

rates between 5% and 12%. The three studies in which ultrasound with colour Doppler imaging was used had recall rates between 8.5% and 17%, and the studies using CA 125 levels followed by ultrasound reported recall rates between 0.9% and 4%. This provides further evidence of the greater specificity of CA 125-based screening. Thus, the recall rate varied considerably across the studies, reflecting the different screening methods and thresholds used, and the characteristics of the women screened.

Positive predictive value of screening tests

If the number of true-positives and true-negatives are known, then the proportion of those under-going diagnostic tests who have cancer, the PPV, can be calculated. The PPV is determined by the test specificity and the prevalence of the disease in a given population. It gives an indication of the relative balance between the potential benefits and harms of screening, by measuring the probability that any individual who screens positive, and therefore undergoes diagnostic surgery, does in fact have cancer.

Table 10 gives the PPV reported in each of the screening studies. The largest study using ultra-sonography alone reported a PPV of 6.7%,[65] where-as the largest study using the CA 125 test achieved a PPV of 27% – just under three women, on average, undergoing unnecessary surgery for every cancer detected.[63] The PPVs reported in individual studies are of limited value, however, because the studies are small and the measures of prevalence of screen-detected cancers are imprecise.

Most women undergoing diagnostic surgery are found to have benign pelvic pathology, which may or may not have required treatment if screening had not been undertaken (see Appendix 4).

Pelvic examination as a screening test

Four studies compared pelvic examination with other screening tests. In three, pelvic examination was compared with the CA 125 test,[59–61] and in one pelvic examination was compared with trans-abdominal ultrasound.[58] Details of these compar-isons are given in Appendix 4. Pelvic examination failed to detect the three cancers detected by ultrasound. In the studies comparing pelvic examination with the CA 125 test, all three

TABLE 10 *Summary of results of prospective screening studies*

Study	Number screened [N]	Population	Sensitivity at 1 year (%)	Specificity (%)	Recall rate (%)	Proportion of false-positives in women screened (%) [b/N]	Proportion of screen-detected cancers (%) [a/N]	PPV (%) [a/(a+b)]
Ultrasound screening (trans-abdominal or trans-vaginal)								
Goswamy et al.[68]	1084	general (post-menopausal)			Not stated	1.29	0.09	6.7
Millo et al.[57]	500	general			5.6	1.20	0	–
Campbell et al.[62] (screen 1)	5479	general	100*	96.5	6.1	2.52	0.04	1.0
Campbell et al.[62] (screen 2)	4914	general	100*	98.2	7.0	1.81	0.06	3.3
Campbell et al.[62] (screen 3)	4201	general	100*	98.8	7.4	1.21	0	–
Demidov et al.[69]	11,996	general			Not stated	2.1	0.09	4.2
van Nagell et al.[82**]	1000	general			5.4	2.40	0	–
DePriest et al.[80**]	3220	general (post-menopausal)			Not stated	1.27	0.09	6.8
van Nagell et al.[65**]	8500	mixed	88	98.7	Not stated	1.33	0.09	6.6
Tabor et al.[55]	435	general			12.4	2.07	0	–
Andolf et al.[58]	805	high risk			10.3	4.47	0.37	7.7
Trans-vaginal ultrasound with CDI								
Kurjak et al.[41]	5013	general			8.5	0.68	0.08	10.5
Vuento et al.[56]	1364	general	100* 50 at 2 years	99.7	11.7	0.29	0.07	20.0
Weiner et al.[79]	600	high risk			16.7	2.50	0.50	16.7
Trans-vaginal ultrasound followed by CDI as a follow-up test (test method)								
Parkes et al.[36]	2953	general	100 50 at 19 months	99.5	Not stated	0.47	0.03	6.7
Bourne et al.[35] (TVS)	1000	high risk	100* 42 at 44 months	95.1	56.8	4.90	0.30	5.8
Bourne et al.[35] (TVS/CDI)	601	high risk	100*	99.0	56.8	1.00	0.50	33.3
Ultrasound followed by other follow-up tests (test method)								
Sato et al.[67] (tumour markers)	15,282	general			5.5	0.3	0.01	4.2
Schincaglia et al.[64] (FNA/B)	3541	general (post-menopausal)	100*	99.5	9.8	0.50	0.06	10.5
Holbert[71] (CA 125)	478	general (post-menopausal)			6.1	1.88	0.21	10.0
CA 125 with ultrasound as a follow-up test (cut-off point)								
Jacobs et al.[60] (30 U/ml)	1010	general (post-menopausal)	100*	99.8	3.1	0.20	0.10	33.3
Jacobs et al.[63] (30 U/ml)	22,000	general (post-menopausal)	79* 58 at 2 years	99.9	1.5	0.14	0.05	26.8
Grover et al.[61] (35 U/ml)	2550	general	–	99.7	4.0	0.30	0	–
Adonakis et al.[59] (35 U/ml)	2000	general	100	99.4	0.9	0.60	0.10	14.3
Bourne et al.[66] (20 U/ml)	1502	high risk	83	98.9	25.2	1.10	0.33	23.8
Bourne et al.[66] (25 U/ml)	1502	high risk	67	99.1	16.1	0.87	0.26	23.5
Bourne et al.[66] (30 U/ml)	1502	high risk	50	99.3	8.5	0.67	0.20	23.1
Bourne et al.[66] (35 U/ml)	1502	high risk	50	99.5	5.5	0.47	0.20	33.3

* *Adequate follow-up for 12 months of women screened negative; ** some overlap in study subjects.*

TABLE 10 contd *Summary of results of prospective screening studies*

Study	Number screened [N]	Population	Sensitivity at 1 year (%)	Speci-ficity (%)	Recall rate (%)	Proportion of false-positives in women screened (%) [b/N]	Proportion of screen-detected cancers (%) [a/N]	PPV (%) [a/(a+b)]
CA 125 with ultrasound								
Akulenko et al.[81]	1003	high risk			Not stated	1.30	0.10	7.1
Karlan et al.[73]	597	high risk			19.3	1.51	0.17	10.0
Muto et al.[74]	384	high risk			Not stated	3.90	0	–
Schwartz et al.[75]	247	high risk			Not stated	0.40	0	–
Dorum et al.[76]	180	high risk			Not stated	5.00	8.9	43.8
Belinson et al.[77]	137	high risk			Not stated	0.73	0.73	50.0

cancers detected had either an abnormal or ambiguous examination. The use of pelvic examination as a screening test would have resulted in more false-positives than the CA 125 test, however. These results, although limited, suggest that pelvic examination does not perform as well as either ultrasound or level of CA 125 as a screening method.

Adverse effects of screening

The decision whether or not to adopt a screening programme needs to take into account the potential harms of screening. As discussed previously (Chapter 2), these include possible over-diagnosis and over-treatment of women with borderline tumours and benign conditions which might not otherwise have caused any morbidity during the woman's lifetime. If operative and psychological morbidity are unacceptably high and the benefits in terms of increased life expectancy are low, then screening will be difficult to justify. The psychological and surgical adverse effects of screening for ovarian cancer are discussed below.

Psychological adverse effects of screening: examples from other screening programmes

It is known that in other cancer screening programmes, such as those for cervical and breast cancer, false-positive results result in a high level of anxiety.[83,84] This anxiety may be more than simply a transient side-effect; in the case of breast cancer screening, a false-positive mammogram may cause a significant increase in long-term psychological morbidity. Gram and colleagues, for example, reported that after 6 months the prevalence of anxiety was still higher in women with a false-positive mammogram result than in a reference group with a negative result.[85] The prevalence of anxiety remained twice as high at 18-months follow-up. Women who have received false-positive results

from mammography screening also report more negative experiences about the screening process, such as pain and discomfort.[86] Abnormal test results following Pap tests and mammograms appear to increase reporting of many other symptoms of psychological distress, such as fear, depression, sleep disturbance, sexual dysfunction and disruption of normal daily activities, although long-term increases in anxiety in false-positives are not consistently reported in all mammography studies.[87] It might be expected that, as a further consequence of false-positive diagnosis, such women would delay in seeking further health investigations because of disillusionment with screening. However, the opposite effect has also been reported in breast cancer screening, in that women with a false-positive diagnosis following mammography are subsequently more likely to practice breast self-examination than those initially screened negative, although this is due to increased anxiety.[88]

Finally, the effects of anxiety are not confined to the individual; anxiety is also likely to be provoked in family members and friends.

Psychological adverse effects of screening for ovarian cancer

Little is known about the reactions to screening of women undergoing screening for ovarian cancer. The background literature search on this subject in Medline, Embase and the psychological abstracts database, Psychlit, identified only a small number of studies that examined this issue *(Table 11)*.

True-negative diagnosis

In one study of the psychological impact of a true-negative diagnosis of ovarian cancer, a Swedish programme using abdominal ultrasound,[89] the reactions of a randomly selected sample of 319 women who had been screened negative (apparent true-negatives) were reported. These women with normal results had been 'at risk' on the grounds of family history, previous cancer or

TABLE 11 *Studies reporting psychological effects of screening for ovarian cancer*

Study, country	Sample	Design and data source	Main results
Pernet et al., 1992[90] UK	10 women with family history of ovarian cancer with false-positive results following surgery.	Qualitative study involving interviews.	Women broadly accepted surgery but great anxiety before biopsy results known.
Andolf et al., 1990[89] Sweden	319 women 'at risk' on grounds of family history or previous cancer or symptoms.	Questionnaire sent to random sample of women screened negative.	Anxiety low on receipt of invitation; majority of women satisfied with screening.
Wardle et al., 1993[91] UK	302 women at 'high risk'.	Prospective study comparing anxiety among those with false-positive and true-negative results.	Short-term anxiety associated with a false-positive result but no serious long-term (3 months) psychological effects.
Wardle et al., 1994[92] UK	31 self-referred women from general population. Positive on initial scan, subsequently shown to be false-positive by scan (n = 31) or surgery (n = 12).	Prospective study of women with false-positive scan, followed-up at 1 year.	No significant difference at 1 year between those originally scanned negative, and the 'false-positive' group.
Wardle et al., 1995[93] UK	358 women interested in ovarian family screening; 379 women who been screened 1 year previously; 186 controls.	Three groups of women were compared with respect to perceived cancer risk and worry about cancer.	Worry about cancer highest in those who had attended screening 1 year previously. Perception of risk not related to participation in screening.
Wolfe & Raju, 1994[94] UK	1833 women aged 45–74 years.	Questionnaire and information leaflet sent to women identified from FHSA lists to assess acceptability of ovarian/ endometrial screening.	Majority (76%) willing to be screened but those more worried less willing to attend.

Note: Papers by Wardle et al., *1993; 1994; 1995 are all part of the same study.*

symptoms, but were not necessarily previously aware of this increased risk. Anxiety was low on receipt of invitation to screening: a score of 3.5 (range, 0–100). The majority of women did not find ultrasound disagreeable (72%) and the majority were satisfied with the amount of information received (92%). Some 74% were reassured by the negative result and the remaining 26% felt "much as before". As regards future screening, 88% wished to be examined on a regular basis, 8% were undecided and 4%, most of who had received hysterectomy and salpingo-oophorectomy, were against future check-ups. Most women considered the examination worthwhile (82%), many saying they now felt reassured and grateful that they had been relieved of their worry about cancer. However, this was a highly selected sample – these women agreed both to participate in screening and subsequently to selection for the survey. A greater prevalence of negative attitudes are likely to be found among the non-responders (275/319 (86%) responded to the survey). It is also well-known that general questions about the acceptability of health-care do not usually elicit many negative responses. It has been also been proposed that one consequence of a true-negative result may be a 'certificate of health' effect, where those screened negative alter their lifestyle in such a way that they believe they can afford to take other risks (e.g. continue to smoke).[95] However, this has not been examined for ovarian cancer screening.

False-negative diagnosis

A preliminary search for literature on the impact of a false-negative diagnosis has recently been carried out by information staff from the NHS Centre for Reviews and Dissemination in order to develop a protocol for a review on the subject. No literature relating to this issue in ovarian cancer screening was found. Andolf and colleagues suggest that one result of a false-negative diagnosis may be for a patient to delay seeking medical advice in future and ignore symptoms; alternatively the screening programme may sensitise the patient to symptoms, making her more likely to seek professional help.[89]

Disillusion and resentment are also likely to be a more immediate result. However, no data are presented to elucidate these issues.

False-positive diagnosis

Evidence of the psychological impact of a false-positive diagnosis of ovarian cancer is limited to three studies. In the first of these, short-term follow-up results from a group of women at high risk because of family history, who were undergoing either transabdominal or transvaginal sonography, were reported.[91,96] A total of 302 women received a scan and participated in the study. Women with an initial positive result were asked to return for rescanning after 6 weeks and none were found to have cancer at surgery. Questionnaires were mailed after the first scan; 31% of women with a positive scan were worried about cancer, compared with 7% with negative scans. Longer-term follow-up data were also collected 3 months after surgery, or after a comparable period in the other screening groups and in a control group. The anxiety scores (General Health Questionnaire (GHQ) and Hospital Anxiety and Depression Scale (HAD)) of women with positive results at scanning but negative results at surgery had returned to baseline levels. This suggests that while screening may be associated with distress in the short term, it does not persist. However, it should be noted that this group of women were at increased risk of ovarian cancer and they may have had a long period to adjust to the possibility of a positive diagnosis at some point in their lives.

One study examined the long-term effects of false-positive results in a sample of women from the general population, that is, not selected to be at high risk because of their particular family history.[92] In this survey, 379 women had referred themselves for screening by ultrasound and 333 received a negative result after the first scan. A total of 46 women were referred for a second scan. The authors were therefore able to categorise women into two groups: 'scan false-positives' (those who received a negative result on the second or third scan, n = 31) and those who proceeded to surgery but were found to be disease-free (surgery false-positives, n = 12). A third group of women with negative results were also included. All three groups were aged 51–53 years (± 8 years) and had a mean of two to three relatives with cancer. There were no statistically significant differences between groups at 1-year follow-up in either GHQ or State–Trait Anxiety Inventory (STAI) scores. However, both sets of scores appear to have been markedly higher in the surgery false-positive group, and the sample size is likely to have been too small to have been able to detect a statistically significant

difference. The possibility cannot be excluded therefore that long-term anxiety is a consequence of a false-positive diagnosis but a larger study would obviously be required. One other notable finding from this study was that anxiety about ovarian cancer was considerable in these women: 29% overall reported themselves "very much worried" about ovarian cancer and 27% of women who had had surgery were more worried since receiving the result, compared with 10% of scan-positive and 3% of scan-negative women.

While the long-term effects (such as shock, distress and fear) of a positive diagnosis of cancer have been demonstrated in studies employing standard anxiety scales, such measures do not give an adequate depiction of the broader impact of a diagnosis on women's quality of life. Qualitative research may provide some of these details, and one small quali- tative study has reported women's reactions to a false-positive diagnosis in a screening programme of asymptomatic women with a family history of ovarian cancer.[90] Ten women aged 27–64 years were interviewed at 12–21 months after surgery. Psycho- metric measures were also employed although, as the sample was small, with no control group, these are not discussed here. The interview showed that six of the women had not previously worried about the health of their ovaries and, indeed, had been unaware that ovarian cancer runs in families before they saw the request for volunteers for the study. Despite their results, most women did not feel their participation had been pointless but were pleased to have taken part in screening and would recommend screening to other women.

Most of the women were very anxious at some time during the scanning procedure and anxiety levels were highest between the operation and the results of the biopsy becoming available. Four women were told little about their operation until much later and two women left hospital without a final biopsy result and spent weeks or months chasing- up the results. It is known that delays in notifying results are associated with distress in screening for cervical cancer.[93] In the case of ovarian cancer screening, if some women require repeated retest- ing before a negative result is notified, this will further increase the psychological costs of the programme. Many spouses apparently also found the experience stressful.

The perceived benefit of taking part can be sum- med up in a quote from one of the participants:

> "I feel a very lucky lady because knowing my family history I am positive that the cyst would

have developed into cancer. I have that firmly fixed in my mind."

Finally, one study was found which examined the attitudes of women to screening for ovarian cancer; 76% of women in inner city practices were reported to be willing to be screened using transvaginal sonography, although the response rate was low.[94]

Summary

For most women the psychological effects of a false-positive diagnosis will be short-lived and only a small minority may suffer long-term anxiety. It is, therefore, easy to assume that screening overall has few harmful psychological effects: most women with false-positive results appear to feel grateful that they have been screened and, in the absence of full information about the risks and benefits of screening, will probably interpret their eventual negative result as a benefit. However, these are asymptomatic and healthy women without ovarian cancer, and in whom ovarian cancer is unlikely to develop, and it is therefore doubtful whether the final negative result can be considered a real benefit to these women.

Heightened awareness of ovarian cancer in healthy women who have been screened negative also appears be another consequence of screening.

There is not enough information to determine whether women at high risk and women from the general population differ greatly in their psychological reactions to false-positive diagnoses. Greater anxiety may be provoked in those with a family history, as they may feel that cancer is very likely to be diagnosed at surgery. In one survey of 242 women with a first-degree relative with ovarian cancer, who were enrolled in the Yale Early Detection Programme, it was found that half of the women reported being increasingly anxious about their own health as they approached the age at which their relative had died.[97] These women may place a different value on having surgery than women in the general population who have been previously concerned about ovarian cancer.

Adverse effects associated with diagnostic surgery

There are few published data on the complications associated with oophorectomy. Case series which examine outcomes of laparoscopy tend to include only a small number of oophorectomies, while the larger case series tend to examine oophorectomy performed at the same time as hysterectomy and do not report separately on oophorectomy carried out as a single procedure. Moreover, studies which explicitly examine the risks and benefits of oophorectomy tend to examine the long-term risks associated with prophylactic oophorectomy, such as osteoporosis and cardiovascular disease, rather than the short-term complications associated with the procedure.

However, the literature does provide some limited data (see *Table 12).* Only those studies which include more than 50 women are considered, because smaller studies provide limited information about less frequent complications.

Leetanaporn and Tintara reported on operative morbidity associated with laparoscopic and open salpingo-oophorectomy for benign ovarian cysts in 82 women.[98] Both methods were considered to be safe and effective, although little other information on complications is presented. A small comparative study has also assessed laparoscopic oophorectomy in 65 women.[99] Rectus muscle bleeding and haematoma formation were the only complications in this series (3.1%); there were no postoperative complications and no blood transfusions were required.[99]

Possover and colleagues reported their experiences with laparoscopic removal of ovarian tumours in 94 postmenopausal women.[100] Ovarian tumours had been discovered during routine gynaecological and/or ultrasound examination. No intra- or post-operative complications were reported and no significant blood loss occurred. Yuen and Rogers reviewed details of a series of 52 women undergoing laparoscopy for ovarian masses.[101] A vaginal ultrasound examination was undertaken in all cases before surgery to exclude possible malignancy. Complications occurred in four patients (7.7%): one involved laparotomy because of failure to achieve haemostasis, and one involved injury to the inferior epigastric vessels. The two remaining patients experienced postoperative complications: one required catheterisation for urinary retention and one developed a pelvic haematoma which resolved spontaneously.

A prospective study by Papasakelariou and colleagues compared the outcome of oophorectomy by laparotomy and laparoscopy in 57 women.[102] Of these, 26 women (mean age, 45 years) underwent laparoscopy and 31 women (mean age, 48 years) underwent laparotomy. Half of the women in each surgical group had adnexal masses. No serious complications occurred in the laparoscopy group; none of the women were readmitted with complications. In the laparotomy group, two serious complications (6.5%) – one bowel injury and one bladder injury –

TABLE 12 Studies reporting complications of laparoscopic examination of ovarian masses

Study, country	Procedure and sample size	Design and data source	Main results
Leetanporn & Tintara, 1996[98] Japan	82 women undergoing either laparoscopic or open salpingo-oophorectomy for benign ovarian cysts.	Comparative study comparing laparoscopic cases with historical open salpingo-oophorectomy controls.	Laparoscopic salpingo-oophorectomy is a safe and effective alternative to open salpingo-oophorectomy.
Minelli, 1996[105] Italy	Laparoscopic removal of ovarian cysts in 920 women.	Retrospective analysis of authors own data.	13 converted to laparotomy; 5 severe intra- or postoperative complications (0.5%).
Papasakelariou et al., 1995[102] USA	Oophorectomy by laparoscopic or laparotomy in 57 women with pelvic pain, adnexal masses or endometriosis.	Prospective analysis of data collected during 1992. Data collected from logs in the operating room, and review of medical records.	No serious complications in laparoscopy group. No laparoscopies converted to laparotomies.
Canis et al., 1994[106] France	Laparoscopy for adnexal masses in 757 women aged 36 ± 13 years.	Retrospective review of data on all patients between 1980 and 1991.	8 complications (1.1%), three involving spillage of cyst contents.
Possover et al., 1994[100]	Laparoscopy in 94 post-menopausal women with ovarian tumours (mean age, 61 years).	Retrospective data from one medical centre between 1992 and 1993.	No peri- or postoperative complications and no significant blood loss.
Yuen & Rogers, 1994[107] Hong Kong	52 women undergoing laparo-scopy for ovarian masses (median age, 35 years). Present-ing complaints included pelvic pain (19 patients), infertility (4), abnormal uterine bleeding (7) and asymptomatic pelvic mass detected during routine check-up (23).	Review of data on 52 consecutive patients.	Post-operative pain reported to be minimal. Overall complication rate was 7.7% and included need for catheterisation, haematoma, epigastric vessel injury.
Reich et al., 1993[103] USA	Laparoscopic oophorectomy in 312 women undergoing unilateral or bilateral oophorectomy for symptoms including pain and/or adnexal mass.	Retrospective analysis of physician and hospital data from 1982 to 1990.	Intra-operative and/or postoperative complications reported in 12 (3.8%) women. Blood loss > 300 ml in two women (0.6%). Length of hospitalisation < 24 hours for 78% of women.
Daniell et al., 1990[99] USA	Laparoscopic oophorectomy by ligature, bipolar coagulation or stapling in 65 patients aged 16–57 years. Indications included pain, ovarian endometriosis, and recurrent benign ovarian cysts.	Comparative study of three methods of oophorectomy. Data collected retrospectively from hospital records from 1989 to 1991.	Rectus muscle bleeding and haematoma formation occurred in two patients (3.1%).
Mage et al., 1991[104] France	481 patients with suspected benign ovarian cyst undergoing diagnostic laparoscopy.	Retrospective analysis of hospital data (1981–88).	3/420 patients (0.7%) undergoing intraperitoneal or transparietal cystectomy developed a complication; no details given.

occurred. These were repaired intra-operatively with no long-term sequelae. In addition, one woman received a blood transfusion, and two had postoperative fever.

One case series was found which specifically reported information on 312 laparoscopic oophorectomies performed over an 8-year period on women, median age 39 years.[103] Intra-operative or postoperative complications occurred in 12 women (3.8%). These included bowel injuries to two women, one of whom also developed adult respiratory distress syndrome. Bleeding from the anterior abdominal wall at the secondary puncture site occurred in one woman. Postoperative complications included postoperative voiding difficulty requiring catheterisation, two urinary tract infections, one case of hydronephrosis requiring a stent, and one ileus which resolved spontaneously. Blood loss greater than 300 ml occurred in two women (0.6%).

Three other relatively large studies, in which the outcomes of laparoscopic investigation for ovarian abnormalities were reported, also suggested that the incidence of complications (such as haemorrhaging and postoperative inflammation) is in the range 0.5–1%. The first of these studies was a large case series reporting the outcomes of diagnostic laparoscopy for suspected benign ovarian cysts in 481 women between 1981 and 1988.[104] Laparotomy was undertaken in 61 women (mean age, 34 years), in whom anatomical conditions made laparoscopy difficult, or in whom malignancy was suspected. In the 420 women who underwent laparoscopy with either intraperitoneal cystectomy or transparietal cystectomy, there were three unspecified complications (0.7%). The second study reported on the laparoscopic management of ovarian cysts in 920 women.[105] A total of five severe complications occurred, either intra- or postoperatively (0.5%). These included one ovarian abscess leading to subsequent laparoscopic adnexiectomy, one inflammation of the abdominal wall, one case of uncontrollable intra-operative haemorrhaging, one case of postoperative haemorrhaging, and one case of postoperative acute abdomen. No other information is provided regarding the selection or characteristics of the cases reported in this series. Finally, Canis and colleagues studied the immediate and long-term consequences of laparoscopic diagnosis of adnexal cystic masses.[106] Long-term follow-up used data obtained either clinically or by mailed questionnaire. A total of 757 women aged about 36 years were investigated and eight complications were attributed to the investigative procedure (1%), three of these being spillage of cyst contents.

Prospective screening studies are another potential source of information on the adverse effects of surgery and the adverse psychological effects of screening. However, only one study has reported this information.[74] In 15 women undergoing surgery, one small bowel perforation requiring segmental resection is reported, with no other intra- or postoperative complications (6.7%).

Thus, these small case series examining outcomes of oophorectomy or of management of ovarian and related abnormalities tend to suggest that complications are rare, and that when they do occur they are minor. However, these studies are too small to provide a reliable estimate of the incidence of rare events. No large (i.e. > 1000 women) case series were found.

An estimate of the risks associated with laparoscopic management of suspected ovarian malignancy can, however, be made. On the basis of the larger case series, 0.5–1% of women may experience complications associated with laparoscopic oophorectomy. These include bleeding and postoperative infection but may also involve more serious complications such as bowel injury. However, this figure should be interpreted with caution as the techniques used in the studies cited here may differ from those likely to be performed as part of a screening programme. There are likely to be differences in age and case mix and, in particular, the actual procedure will be different, as most of the screening studies identified in this review were open procedures rather than laparotomies. Moreover, these studies do not provide estimates of the risk of mortality, nor do they accurately represent the true morbidity associated with diagnostic surgery, rather than just the rate of surgical complications.

Costs of screening

The search strategy identified one model of the relative cost-effectiveness of different screening strategies, which is discussed in Chapter 5. In addition, several articles were found in which data on charges in the USA are reported for various components of screening. Charges for transvaginal ultrasound scans vary between $150 and $275,[108,109] while costs for CA 125 testing vary between $45 and $61.[77,109] The charge for a laparoscopy is reported to be $3000, which indicates that a critical component of the total costs of screening will be the proportion of women referred for surgery.[108] One published report gave the actual costs, rather than the charges, incurred over time in establishing

and running an ultrasound-based screening programme.[37] The marginal cost for each ultrasound scan was only $25 once the programme was running at full capacity; this illustrates that the actual costs of screening may bear little resemblance to putative costs based on charges. In this study, the majority of costs incurred for each case of ovarian cancer detected resulted from the diagnostic procedures undertaken.

It seems reasonable to assume that the cost of screening a woman with ultrasound will be higher than the cost for CA 125 testing followed by ultrasound screening. Ultrasound screening requires investment in equipment and trained personnel, and requires centralised facilities for women to attend for screening. In contrast, collecting blood samples for CA 125 testing is a simple procedure which can be performed at any location, with samples transported for bulk analysis. However, the total cost of these screening options also depends on the numbers of women recalled for further assessments and for diagnostic procedures. A recent model, which compared annual transvaginal sonography with annual CA 125 testing followed by transvaginal sonography in those women with elevated or rising levels of CA 125 and used charges to estimate costs, found that the cost per life-year saved was lower for the CA 125 strategy.[108]

Chapter 5

Research in progress

RCTs of ovarian cancer screening

Three RCTs are currently in progress, investigating the effect of screening for ovarian cancer on mortality from the disease in women from the general population. Two of the trials are being coordinated in the UK: the trial at St Bartholomew's Hospital, London (Bart's), is recruiting women within the UK,[110] and the European Randomised Trial of Ovarian Cancer Screening (ERTOCS) is a multicentre trial open to participating centres throughout Europe.[111] The third trial is being carried out in the USA; the Prostate, Lung, Colon, Ovary (PLCO) trial coordinated by the National Institutes of Health (NIH) investigating screening for prostate, lung and colorectal cancer, as well as ovarian cancer.[112]

Protocols for each of these trials have been supplied by the investigators[110–112] and the key features are summarised in *Table 13.* These protocols are subject to on-going review and the details contained in the table are those most recently made available.

A trial of screening requires recruitment of healthy women who have not sought treatment. In this respect it is unlike a trial assessing treatment options, where the trial is ethically justified if there is uncertainty about the relative merits of different treatments. For a trial of preventive measures, there is a particular duty to ensure that the level of risk to which healthy volunteers are exposed is acceptable relative to the potential benefits, and that the volunteers are fully informed.

The following discussion appraises the design of these trials and assesses the information that they will provide if successfully completed.

Screening algorithms being evaluated
The Bart's trial is investigating the use of annual screening with CA 125 testing followed by an ultrasound scan as a screening strategy. The protocol differs from that used previously by these investigators,[63] in that decisions to recall for a repeat measurement of CA 125 level or scan will be based on the woman's risk of ovarian cancer, calculated on the basis of age and the level and rate of change of CA 125. In some circumstances, women will be referred on the basis of CA 125 measurements in

the presence of a normal scan. Women with equivocal results will be recalled for further testing, although to avoid repeated recalls a maximum of five recalls will be allowed for repeat investigations using ultrasound. A retrospective analysis of data from a previous screening study found that this algorithm resulted in a false-positive rate of 0.3%.[113] The cut-off point for recall for repeat testing is relatively low, at 15 U/ml, which means that a large proportion of women, perhaps more than 25%, will be recalled for repeat testing.

The ERTOCS trial uses transvaginal ultrasound as a screening test. The algorithm for determining a positive result is complex but essentially all ovarian abnormalities apart from small simple cysts will result in recall for further assessment. A maximum of three scans including the initial scan will be carried out before a definitive decision on referral is made.

A pilot study using a similar algorithm reported 3% of women with persistent abnormalities on ultrasound screening.[36] In this pilot phase of the study, colour Doppler imaging was used as a secondary test and a false-positive rate of 0.5% for surgical referral was reported; however, subsequently the use of colour Doppler imaging as part of the screening algorithm has been dropped because the reduction in false-positive rate was not maintained.

The PLCO trial is investigating annual CA 125 testing, pelvic examination and transvaginal ultrasound as screening tests. These tests are carried out independently and in a blinded fashion on each woman. The study protocol does not give specific definitions of abnormal results but states that women with abnormal or equivocal results will be referred back to their own doctor for follow-up.

This protocol is similar to the screening protocols which have been used in screening women at high risk. However, because of the looseness of the definitions of abnormal screening results, the lack of a protocol for further management, and the lack of repeat screening to rule out transient changes, the proportion of false-positives is likely to be higher than that reported in prospective screening studies.

TABLE 13 *On-going RCTs of screening*

Study	Bart's[110]	ERTOCS[111]	NIH PLCO[112]
Setting	UK	UK/Europe	USA
Start date	1995?	1995?	1993?
Screening protocol	CA 125 test followed by ultrasound in those testing positive – calculated on basis of age, level and rate of change of CA 125 level. Annual screenings for 6 years.	TVS at either 18 or 36 month intervals. Referred for repeat scan if ovarian volume ≥ 3 multiples of the median (MoM) (postmenopausal) or ≥ 4 MoMs (premenopausal) or if cyst present, unless simple unilocular cyst with regular outline, diameter < 50 mm.	TVS and CA 125 and pelvic examination. Positive results not strictly defined, but any positive/ suspicious result leads to referral to patients' own physician. Annual screening for 4 years.
Study population	Postmenopausal women aged over 50 years.	50–64 years.	Women aged between 60 and 74 years.
Recruitment	Volunteers: recruited via press, through occupational health departments and by invitation in participating general practices.	Women selected either from a population registry or invited when attending for breast cancer screening.	Volunteers: recruited through press.
Target number of subjects	60,000 in each arm.	30,000 in each intervention arm, 60,000 in control group.	37,000 in each arm.
Follow-up – method and length	Cancer registrations. Postal questionnaire at 4 and 7 years.	Cancer registrations/death notifications. Length of follow-up not given.	Annual postal questionnaire. Death registrations. Follow-up period: 10 years.
Estimated number of deaths from ovarian cancer in unscreened group	111 over 6 years.	Not given.	Not given.
Stated power	80% power to detect 30% reduction in mortality at 5% significance level.	78% power to detect one-third reduction in mortality at 5% significance level.	77% power to detect 30% reduction in mortality at 5% significance level.
Expected date of completion	7 years follow-up.	10-year follow-up?	10-year follow-up.
Uptake/acceptability assessed?	No	Yes	No
Evaluation of costs?	If funding permits.	Planned	No
Evaluation of harms?	Will measure surgical intervention rates in both arms. Evaluation of psycho-logical effects if funding permits.	Data on surgical complications to be recorded in those referred for diagnosis.	Stated that morbid events associated with screening or diagnosis will be recorded.
Comments	Preliminary results of stage distribution in screen-detected cancers available at 4 years.	Stage distribution in screened and control groups obtained by 4 years. Serum collected from screened women to enable retrospective analysis of tumour markers. CDI also undertaken at repeat screening for retrospective analysis.	Should allow comparison between the three methods. However, the looseness of definition of positive result means the false-positive rate is likely to be high.

In the populations being screened in these studies, the expected incidence of ovarian cancer is 40 per 100,000 per year and, based on published studies, it seems unlikely that the prevalence of cancer detected on screening will exceed twice this figure, or 0.08%. A false-positive rate of 1% will therefore result in about 12 diagnostic operations for each cancer detected. The Bart's study may achieve a higher PPV, and the ERTOCS and PLCO studies, a lower PPV. It is notable that one of the centres participating in the ERTOCS study stopped recruiting after 13,000 women were randomised, because of the unexpectedly large number of complications in women undergoing surgery (Tabor; personal communication, 1997).

The ERTOCS trial has two intervention arms, one screened at 18 months and one screened at 3 years. This allows investigation of the effect of screening interval on the benefits, harms and costs of screening.

Sample size and power

The two UK-based trials aim to recruit 120,000 women each to randomisation. This sample size is calculated to have 80% power to detect a 33% reduction in ovarian cancer mortality in the screened group, at a significance level of 5%. This is based on a total of about 110 ovarian cancer deaths in the 60,000 controls over 4 years.

The PLCO trial aims to recruit a total of 74,000 women to randomisation. It is calculated that this will give a 77% power to detect a 30% reduction in mortality after 10 years follow-up, at the 5% significance level. However, screening for ovarian cancer is already becoming widespread in the USA and contamination may therefore be a problem for this trial, reducing its power.

Population selection

Both the Bart's and PLCO trials are recruiting volunteers from a range of sources but principally from women who respond to publicity disseminated through a variety of channels. Neither trial is designed to allow a calculation of the uptake of screening in a randomly selected general population sample. Selection bias may lead to a healthier than average group of women participating in the trial, which might reduce the prevalence of cancers and, hence, the power of the study.

In the ERTOCS study, recruitment in the UK is by invitation to women attending breast screening, allowing uptake rates to be measured. These women may not be representative of a general population invited for screening, since they have already chosen to participate in a screening programme. Participating centres in Europe which sample women from a population register may give a more accurate estimate of the effect of selection bias and the uptake of screening.

Outcome measures: assessment of morbidity and costs, and of mortality

All three trials measure mortality from ovarian cancer as a final endpoint. The ERTOCS study relies on reporting of cancers via routine data systems such as cancer registries and death certificates for ascertainment in the control group, whereas the other trials also use a postal questionnaire as follow-up. There may therefore be differences in case ascertainment between trials.

In the Bart's trial, questionnaires are sent to both screened and control groups, enabling investigation of the effect of screening on the number of gynaecological procedures carried out. This may provide information relevant to the question of the natural history of screen-detected benign ovarian conditions.

All trials will record the nature of surgical procedures undertaken in the screened group and any complications arising, thus enabling quantification of the morbidity in false-positives arising as a result of screening. However, this will be restricted to description of the prevalence of complication rates, rather than a fuller measurement of morbidity and quality of life.

It is planned to incorporate an economic analysis of screening in the two UK-based trials. However, the Bart's study team has not yet developed a protocol for this aspect of the trial.

The Bart's study team also intend to incorporate investigation of the psychological impact of screening but, again, a protocol has not yet been developed. In the two remaining trials there are no plans to investigate this aspect of screening.

Scope for assessment of new screening strategies

All three trials will establish a serum bank with stored samples from screened women. This will facilitate the investigation of the value of newly-identified markers, or novel ways of combining or interpreting markers, in a large cohort of women with good follow-up data on the incidence of ovarian cancer. The ERTOCS and PLCO studies will have ultrasound data on all screened participants and will therefore be able to investigate the effect of using various serum markers either in comparison with, or combined with, ultrasound screening.

The ERTOCS trial is also recording information from colour Doppler imaging on women who have an abnormal ultrasound scan. This should enable a fuller assessment to be made of the value of colour Doppler imaging as a secondary test. The information recorded on the ultrasound findings will also allow correlation with the risk of malignancy. This could help in the development of more specific criteria for diagnostic intervention.

Conclusions

These three trials should all provide reliable information concerning the effect of screening on ovarian cancer mortality and incidence. They will also build-up valuable data sources for the development of improved screening methods.

The main weaknesses of these studies are the absence of an assessment of some of the 'softer' outcomes of screening, both in terms of the morbidity experienced by both true- and false-positives and the psychological impact of screening. This is particularly important in view of the large proportion of women who are likely to be recalled under the proposed protocols and the relatively poor information so far published on the effects of diagnostic surgery.

Another important issue is the measurement of the cost-effectiveness of screening. Ideally, the costs and consequences of screening should be measured in the trials in such a manner as to permit comparison of the cost-effectiveness of the very different screening strategies being evaluated. This would require increased collaboration between the research teams.

These studies all began in about 1994–95 and results relating to mortality effects may not be available until about 2003. However, intermediate outcomes of interest, such as the false-positive rate and the morbidity in false-positives, could be available much sooner than this if the investigators wished to publish them.

Studies on screening in women with a family history

There are no RCTs currently in progress in the high-risk population. However, if the natural history of ovarian cancer is similar in this group, the results of the RCTs carried out in the general population could be used to model the cost-effectiveness of screening in a group with a higher prevalence of the disease.

Screening is currently being offered as a service in the UK to some women with a strong family history of ovarian and other relevant cancers. There is a reluctance to establish RCTs in this group because of their high risk of developing ovarian cancer; the average lifetime risk of developing ovarian cancer for women with two affected close relatives is about 15%, which is about 1.5 times the average lifetime risk of developing breast cancer.

The UK Committee for Coordinating Cancer Research (UKCCCR) has recently established a prospective uncontrolled screening study for women with a family history of ovarian cancer.[114] The eligibility criteria for the study require a history of more than one affected close relative, such that this group has an average lifetime risk of developing ovarian cancer of at least 15%. This study has no comparison group and cannot, therefore, provide information about the effectiveness of screening. Like the prospective screening studies whose results have been published, it can only provide information about the performance of the screening test and the risks of screening.

The screening protocol proposed for use in this study includes annual CA 125 measurement and transvaginal ultrasonography. This protocol has been selected to maximise sensitivity at the expense of specificity and is not directly comparable with the protocols being evaluated in the UK-based RCTs. A high proportion of women may be recalled for repeat tests – an estimated 5% following ultrasonography and perhaps 15% for repeat venepuncture.[66] The main research objective of this study is to collect data to develop a model to determine an individual's risk of ovarian cancer. It is hoped that the use of this model in a screening algorithm might improve sensitivity and specificity.

The incidence of ovarian cancer will be measured through cancer registry data, giving information about the sensitivity of screening and the risk of developing ovarian cancer in this group. There is no provision in the protocol to collect data on operative morbidity of false-positives nor on the psychological impact of screening in this group of women. The study concentrates on attempting to devise an improved screening method, while paying less attention to the opportunities afforded by this study design to investigate the consequences of screening.

This study may have two unplanned consequences. First, it may improve the quality of screening already being offered to women at high risk in participating centres, by providing a clear

screening protocol and quality control measures. Second, the study seeks to actively recruit women from the general public by publicity. Recruiting to a non-randomised study in this way may give the erroneous impression that screening has been proven to be safe and effective. It will also increase awareness and anxiety about ovarian cancer among women with a family history, most of whom will not fulfil the criteria for entry into the study. This is likely to increase demand for screening among a far larger group of women than those at whom the study is aimed.

Unpublished studies

Authors of all prospective studies included in this review were contacted to establish whether they had further unpublished data relating to their screening studies. One group of investigators is preparing data for publication relating to

three annual incident screens in 11,000 women randomised to CA 125 screening (Jacobs; personal communication, 1997). These data will provide information relating to the outcome of repeated screening rounds based on CA 125 testing, and will also provide a comparison with the control group who received one initial screen.

There are further unpublished data relating to repeated screenings with transvaginal sonography. A cohort of women at high risk have undergone three screening rounds but published data relate only to the initial screen.[35] Because of lack of funding, the data relating to incident screenings are likely to remain unpublished (Bourne; personal communication, 1997). In the largest transvaginal sonography screening study in the USA, several screening rounds have also been undertaken in its participants. Unfortunately, the published data do not report the results of each round separately.[65]

Chapter 6

Discussion

Limitations of the published research evidence and the review methods

The published research evidence gives only limited information about the potential impact of ovarian cancer screening. Some of these limitations result from the study designs used and some from the quality of the conduct and reporting of the studies. Uncontrolled studies of screening cannot give reliable information about the effectiveness of screening, this requires randomised trials comparing mortality in screened and unscreened populations.[34] Thus the main role of such studies is in assessing whether screening test performance is adequate to justify the establishment of an RCT, and to help decide which screening methods should be investigated in such trials. Such studies can also be used to assess the adverse effects and costs of screening.

To investigate these issues reliably, prospective uncontrolled studies need to be designed robustly with a clear research aim. A study aiming to evaluate the performance of screening tests should define the screening protocol before the study commences and report data relating to all study subjects, including those lost to follow-up. If sensitivity is to be estimated, women must be followed-up after a negative test result and the study must be large enough to expect to detect a substantial number of ovarian cancers. Many published studies did not meet these standards.

The proportion of false-positives, and hence an estimate of test specificity, can be more readily estimated from this type of study. However, comparison of test specificities in different studies without also considering sensitivity gives limited information because the two parameters are interdependent. Specificity estimates can also be affected by the completeness of follow-up of women recalled for further tests; the proportion of false-positives may be underestimated if large numbers of women are still awaiting definitive results from screening or if some women fail to attend. Not all studies indicated the completeness of the information reported;

there was little detail of the actual procedures performed on women screened positive and on the outcomes or complications associated with such procedures.

Uncontrolled studies can also be useful in comparing the relative performance of different screening tests and algorithms, in order to define appropriate tests for use in a trial. The most valid way of making such a comparison is to perform each screening test on the same women, with observers blinded to the results. However, only one such comparison has been published – between CA 125 and ultrasound testing.[35] This means that there is little reliable evidence on which to base comparisons of different screening tests. Another area which has been under-reported is the effect of repeated screening rounds; data from each screening round should be reported separately to investigate the effect of repeated screening on detection rates and false-positive rates. So far, only one study has published such data.[62]

Most of the studies were carried out on volunteers and the method of recruitment was often not fully defined. This means that no estimates of uptake can be made, nor any assessment of the likely impact of selection bias. It is therefore difficult to assess the degree to which selection bias may have influenced the findings and to judge the relevance of the studies to screening being offered in an unselected population.

Many studies have investigated test performance in women undergoing surgery for suspected ovarian masses. However, such studies are not directly applicable to the screening situation because many of the women have clinically detectable ovarian abnormalities; hence, the estimates of sensitivity and specificity obtained cannot be directly applied to the detection of pre-clinical ovarian cancer in asymptomatic women. Studies of this type were not therefore systematically reviewed and fairly restrictive inclusion criteria were set to increase the validity of the findings of the review.

Obtaining evidence about the potential complications of surgery in false-positives necessitated the

use of a specific search strategy, because of the lack of information reported in the published screening studies. It was not intended to identify all potentially relevant studies but simply to identify case series of surgery potentially relevant to the diagnostic surgery undertaken as part of the screening process. The case mix and the nature of the surgical intervention may not be directly comparable to that observed in a screening study, and the resulting estimates of complication rates are clearly less reliable than if estimates taken directly from a screening study had been available.

Summary of research evidence

The available evidence suggests that screening with ultrasound, with or without colour Doppler imaging, or with CA 125 measurement followed by an ultrasound scan can detect ovarian cancer in asymptomatic women at an earlier stage than in an unscreened population (resulting in approximately 50–75% of cancers diagnosed at Stage I). Screening with ultrasound (trans-vaginally) appears to have higher sensitivity but lower specificity than CA 125 measurement followed by an ultrasound scan. Colour Doppler imaging added to grey-scale ultrasonography may increase specificity but the consequences for sensitivity are unknown.

Sensitivity of ultrasound screening, defined as the proportion of cancers arising in 1 year which are detected on screening, appears to be close to 100% but this is based on limited information. Sensitivity of the CA 125 test followed by an ultrasound scan at 1 year appears to be about 80% (95% CI, 49–95). The prevalence of screen-detected cancer and the proportion of cancers detected at Stage I are consistent with some improvement in stage at diagnosis compared with an unscreened population, but the precision of these estimates is low and the clinical significance, in terms of the potential impact on mortality from ovarian cancer, is unknown. There is little information on the impact of repeated screening or on the optimum interval between screenings.

Some women may have tumours of borderline malignancy diagnosed at screening which may not have been clinically detected in their lifetime, resulting in potentially unnecessary intervention and treatment. This may be more likely to occur with ultrasound screening. The available information suggests that the overall proportion of borderline tumours reported in screening

studies is consistent with that expected in the population as a whole;[5] however, these tumours have tended to be concentrated in particular studies, suggesting that there may be inconsistencies in the way these tumours are classified and reported. The extent of possible over-diagnosis can only be assessed in randomised trials which can compare the incidence of ovarian cancer in screened and unscreened populations.

The proportion of women screened who were false-positives ranges from about 0.1–0.6%, for screening with CA 125 followed by an ultrasound scan, to 1.2–2.5%, for grey-scale ultrasound screening alone. Most women undergoing diagnostic surgery who are found not to have cancer are found to have a benign ovarian or gynaecological condition. The potential benefits of surgical intervention in this situation are unknown. The risks of surgery include a small chance of death and, also, the risk of significant complications such as bowel or bladder damage, infection or excessive bleeding. The risks are difficult to quantify but may be about 0.5–1% for those undergoing diagnostic surgery. However, complication rates give only a limited picture of the adverse effects experienced by women who are screened positive but are found not to have cancer. In addition to these risks, a much larger proportion of women who do not have cancer (perhaps 3–12% of all screened women) will be recalled after the initial screening test for further assessment, and these women may experience distress and anxiety while awaiting the result of their tests.

A definitive answer to whether screening can improve the outcome for women with ovarian cancer requires an RCT of screening. Currently on-going RCTs, if successfully completed, will not report their results for at least 5 years. Such trials, as well as quantifying any benefits and harms of screening, could also provide information concerning the relative cost-effectiveness of different screening methods and screening intervals.

In the absence of direct evidence about screening from RCTs, models of the potential impact of screening can be constructed which may help in making judgements about screening. The reliability of these models depends on the accuracy of the assumptions made in place of empirical evidence. Nevertheless, they may be helpful in assessing the likely health gain which ovarian cancer screening might achieve and, therefore, in judging the relative priority of NHS support for further research in this area.

Modelling the impact of ovarian cancer screening

A number of authors have attempted to model the impact of ovarian cancer screening.[36,108,115–117] Details of the methods and findings of these modelling studies are given in Appendix 5.

The simplest model estimates the percentage reduction in ovarian cancer mortality 5 years after screening, by combining estimates of the proportion of cancers detected at Stage I in screened women with currently observed stage-specific survival rates.[36] The percentage mortality reduction predicted by this model over a range of different proportions of cancers diagnosed at Stage I are shown in *Table 14.*

TABLE 14 *Effect of proportion of ovarian cancers diagnosed at Stage I on estimated mortality reduction*[36]

Proportion of Stage I cancers in screened population	Expected reduction in mortality compared with unscreened population (%)[*]
34	0
50	15
80	43
100	61

[*] *Assumes 34% present at Stage I in the absence of screening, and 5-year survival of 75% at Stage I and 16% at all other stages.*

This highly simplified model demonstrates that a maximum reduction in mortality of 61% could theoretically be achieved, if all women with ovarian cancer are diagnosed at Stage I and the current 5-year survival rate applies. However, this theoretical reduction is extremely unlikely to be achievable in practice, as it assumes that no cases are missed and that the survival rates currently observed apply to screened Stage I cancers. It also assumes very frequent screening. If 75% of cancers are diagnosed at Stage I (based on published screening studies; see *Table 6*), this model predicts a mortality reduction that could range from 0% to 61% at 5 years but is most consistent with a reduction of 40%.

The authors calculate the cost per life saved by ovarian cancer screening to be about twice that for breast cancer screening. This is based on a number of fairly optimistic assumptions:

- 43% reduction in mortality in screened women, equivalent to 80% diagnosed at Stage I
- a screening interval of 3 years
- the same cost per screen as for breast cancer.

With these assumptions, for every 10,000 women screened there would be 1.6 extra 5-year survivors per year.

Similar methods using observational survival data from the USA suggest that screening might result in a 50% reduction in ovarian cancer mortality, equivalent to 85% of cancers in screened women diagnosed at Stage I.[117] For annual screening, this implies 1.7 additional 5-year survivors for every 6000 screening tests performed.

A decision-analysis model, based on a once-only screen and taking account of the adverse effects of laparotomy, calculated the average increase in life-expectancy in a screened population to be about three-quarters of a day for women screened at the age of 65 years.[115] These authors calculate that breast screening would achieve about twice as much gain as ovarian cancer screening. The simple stage shift model (which includes no allowance for adverse effects) assumed that each extra 5-year survivor would gain an extra 19.3 years of life.[36] This is equivalent to just over 1 day of life gained for each women screened, the same order of magnitude as that estimated by this decision-analysis model.

More complex models use computer simulation to reflect the dynamic nature of the growth of cancers and their likelihood of detection over time. One such model, which used clinicians' estimates to model the natural history of ovarian cancer, predicts an average of 3.4 years gained for every case of ovarian cancer (screen-detected and clinically detected) for CA 125 screening.[116] This is rather lower than that estimated by the simple stage shift model, which suggests about 7.7 years gained per ovarian cancer case.

Building on this model, using the same estimates of the natural history of ovarian cancer, Urban and colleagues have compared the relative cost-effectiveness of ultrasound (transvaginal sonography) screening and CA 125 screening followed by transvaginal sonography.[108] The model resulted in 66% of cancers diagnosed at Stage I for the ultrasound strategy and 51% diagnosed at Stage I for the CA 125/transvaginal sonography strategy, similar to the average proportions in published screening studies (see *Table 6*). The number of life-years saved according to this model was

equivalent to only about 1 year for each case of ovarian cancer in the population. This model estimates that, for annual screening, the CA 125 strategy saves nearly two-thirds of the number of life-years compared with the transvaginal sonography strategy – at about one-third of the cost. The CA 125 test retained its cost-effectiveness advantage across a wide range of sensitivity analyses. The authors did not consider the effect of less frequent ultrasound screening; this being investigated in one of the on-going RCTs.

The limitation of all these models is that they must rely on assumptions of treatment effectiveness based on currently observed survival rates at different disease stages, and may therefore over- or even underestimate the likely benefit. They cannot replace RCTs as a means of establishing whether or not earlier detection can improve outcome. Furthermore, the published models have, in general, been less useful for assessing the size and distribution of any adverse effects; these are either ignored, or added together with benefits to produce a figure for the 'net benefit', which does not allow consideration of the distribution of benefits and harms.

Potential benefits and harms

The potential balance between benefits and harms which may result from screening for ovarian cancer is considered here. In the absence of evidence concerning benefits, a level of benefit will be assumed that is consistent with the more encouraging results from published studies. If early detection and treatment is effective, benefits and harms could be experienced in terms of survival and quality of life. Potential benefits are only discussed here with respect to length of life or mortality.

The smallest effect on ovarian cancer mortality that the trials currently in progress can be confident of detecting is a 30% reduction. The maximum likely achievable reduction is 60%, equivalent to 100% of cancers being diagnosed at Stage I at current survival rates.

The absolute reduction in mortality rate corresponding to this range of relative mortality reduction is shown in *Table 15*, assuming that screening is offered to women between the ages of 50 and 64 years (who have a higher incidence of ovarian cancer than younger women but still have a reasonable life-expectancy) and that the mortality reduction occurs 5 years later. This shows that the number of extra 5-year survivors gained

TABLE 15 *Illustration of mortality reduction per year for a given percentage reduction in mortality*

	Percentage reduction in mortality from ovarian cancer			
	30	40	50	60
Absolute reduction in mortality in screened women (number of extra survivors at 5 years per 100,000 per year[*]).	11	15	18	22

* *Based on 1994 mortality rate in England and Wales in women aged between 55 and 69 years of 36.5 per 100,000.*

per 100,000 women screened per year is likely to be quite small, because the number of deaths due to ovarian cancer is relatively low (about one-third that from breast cancer in this age group). The total number of extra survivors depends on the period over which screening is offered and the duration of any effect on mortality. The number of life-years gained is difficult to estimate because the proportion of extra 5-year survivors who are 'cured' in the long term is unknown.

If we now assume that ovarian cancer screening results in a 40% reduction in mortality, this is equivalent to 15 extra survivors at 5 years per 100,000 women screened per year. This is equivalent, at current survival rates, to nearly 80% of cancers being diagnosed at Stage I and is consistent with the results reported from the more encouraging prospective screening studies (see *Table 6*). It is also more favourable than the stage shift resulting from the most sophisticated screening model[108] and might, therefore, be considered a relatively optimistic assumption.

Table 16 illustrates the outcomes which might result in a hypothetical cohort of 10,000 women who have an average annual incidence of ovarian cancer of 40 per 100,000 – the approximate incidence in the UK for women aged 50–64 years. The potential annual outcomes of screening in the 'steady state' are shown in the table. Two illustrative scenarios are given; the 'CA 125 scenario', which assumes annual screening with 3% of women recalled and 0.2% of women undergoing surgery for conditions other than ovarian cancer, and the 'transvaginal sonography scenario', which assumes bi-annual screening with 7% of women recalled and 1.3% of false-positives at diagnostic surgery. These assumptions are consistent with the results reported in the more favourable prospective screening studies, so

TABLE 16 *Annual outcomes of screening in a hypothetical cohort of 10,000 women assuming 40% mortality reduction and an annual incidence of 1 in 2500 for ovarian cancer*[*]

	CA 125 scenario	TVS scenario
Number of women participating in screening programme	10,000	10,000
Screening interval	Annual	Every 2 years
Number of screening tests carried out per year	10,000	5000
Number of women recalled for further assessment per year	300 (3% of screens)	350 (7% of screens)
Number of women undergoing surgery per year who do not have primary ovarian cancer	20 (0.2% of screens)	65 (1.3% of screens)
Maximum number of cancers detected on screening per year (if 100% sensitivity)	4	4
Number of additional 5-year survivors per year	1.5	1.5
Predictive value of recall (if 100% sensitivity)	1.3%	1.1%
Predictive value of diagnostic surgery (if 100% sensitivity)	17%	5.8%

[*] *See text for discussion of assumptions.*

may be optimistic. It is assumed that both of these strategies result in the same benefit, that is, a 40% reduction in ovarian cancer mortality.

This illustration raises a number of issues. First, the assumption that screening every 2 years with ultrasound would produce equivalent benefits to annual CA 125-based screening is likely to be inaccurate. It has been used to illustrate the greater sensitivity of transvaginal sonography compared with CA 125-based screening, which implies that a longer screening interval would detect the same proportion of cancers. The actual screening interval for transvaginal sonography which would be equivalent to annual CA 125-based screening is unknown. The shorter the screening interval, the greater the number of women with false-positive tests – and the greater the costs of screening. The optimum screening interval depends on the rate of growth of ovarian cancer and on the trade-off between costs, benefits and harms.

The number of women who might benefit from screening is shown to be small compared to the number of women who might suffer adverse effects. For every 1.5 extra survivors at 5 years, between 20 and 65 women who do not have ovarian cancer might undergo an operative procedure, and between 300 and 350 women would be told after initial screening that they required further tests, with the associated adverse psychological effects. However, it is possible that women undergoing surgery for benign conditions might derive benefit.

Researchers investigating screening for ovarian cancer have adopted an *ad hoc* benchmark to define an 'adequate' screening test; it has been

stated that a test which results in fewer than one in ten operations finding ovarian cancer would be unacceptable in clinical practice.[118] This is an arbitrary figure, however, which has no empirical basis. In the above illustration, the maximum predictive value for the CA 125 scenario was 17% and for the transvaginal sonography scenario under 6%. These figures are based on the optimistic assumption that all cancers are detected at screening (i.e. 100% sensitivity) and may therefore be overestimates. However, the only other determinants of these estimates are the incidence of ovarian cancer, the frequency of screening and the false-positive rate of the screening test. Thus, they fairly reliably indicate the maximum PPV likely for these tests in general population screening. Whether this is considered 'acceptable' involves a judgement about the likely benefits and risks; at present there is little information on which to base such judgements. Even if such information were available, individual women, clinicians and policy-makers might form different views about what would be an acceptable PPV in practice.

If this illustration is compared with the situation for breast cancer screening, which may reduce mortality by up to 40% in screened women (as opposed to a 25% reduction in the population invited for screening), this results in 38 deaths averted per 100,000 screened women per year.[119] This is more than twice as many deaths prevented compared with ovarian cancer screening (assuming a 40% reduction in mortality) because of the higher incidence of breast cancer. If the costs of screening are similar, this suggests that ovarian cancer screening is likely to be less cost-effective than breast cancer screening, even under quite

favourable assumptions. The same conclusion has been reached by other modelling studies.[36,115] Ovarian cancer screening might, however, need to be undertaken more frequently than the 3-yearly interval of the NHS breast screening programme, which would further reduce its relative cost-effectiveness. Moreover, as indicated in Chapter 4, many women will chose not to attend for repeated screening; in one study, even among a highly selected and motivated group of women at high risk because of their family history, 12% were against, or undecided about participation in further screening.[89]

It is relevant, therefore, to consider whether further research into screening for ovarian cancer should be considered a priority by the NHS, if it appears unlikely that general population screening could prove to be as cost-effective as screening for breast cancer, which has itself been considered a controversial use of NHS funds.[120] The impact that further developments in ovarian cancer screening might have on its potential cost-effectiveness are considered below.

Developments in ovarian cancer screening

The overall impact of ovarian cancer screening depends on the balance of potential benefits, harms and costs. This balance will be more favourable if benefits can be maximised, and harms and costs minimised.

Treatment effectiveness
Any benefits of screening depend on the relative effectiveness (and cost-effectiveness) of early compared with late treatment. Increasing the effectiveness of treatment of screen-detected cancers may improve the cost-effectiveness of screening. Some of this may involve optimising the treatment received by women in the screening programme; the estimates of potential mortality reductions above are based on the assumption that 5-year survival in early cancer is the same as that currently observed on a population basis, while 5-year survival rates in some trials in selected cases of early disease have exceeded 90%.[16] Conversely, however, any advances in treatment effectiveness for advanced cancer will reduce the potential for screening to improve outcomes.

Improvements to screening tests
Ultrasonography
Ultrasound screening appears to be quite sensitive in detecting ovarian abnormalities; however, many screen-detected abnormalities are not malignant but visualising them frequently leads to surgical intervention with a resulting high false-positive rate. The challenge is therefore to improve the specificity of this test whilst retaining or improving its sensitivity. Two methods under investigation are the role of colour Doppler imaging of ovarian blood flow and the use of morphological classifications to distinguish between benign and malignant lesions.

Despite the use of colour Doppler imaging in several prospective screening studies, its role in screening is still not clearly defined, partly because technical developments have meant continual changes in the cut-off points used to define abnormality. Several studies suggest that the use of colour Doppler imaging increases the specificity of screening but it is not known to what extent this may also reduce sensitivity.[35,36] It has been claimed that when used as an adjunct to grey-scale ultrasonography as an initial test, colour Doppler imaging can improve sensitivity by detecting small malignant lesions that are morphologically normal on the basis of abnormal areas of blood flow.[70] However, these results have not been replicated, and there has been considerable variation in the parameters used to define colour Doppler imaging abnormalities and in the results obtained.[42] The ERTOCS study was unable to maintain a useful reduction in false-positive rates by using colour Doppler imaging and is no longer using information from it as part of the screening protocol. Also, colour Doppler imaging equipment is expensive and its use would increase the costs of screening.

There is a need to characterise more reliably what colour Doppler imaging adds to screening by grey-scale ultrasound alone. The images being recorded as part of the ERTOCS trial will provide some answers but will not demonstrate its effect if used on all screened women. A reliable investigation of this question would require a blinded comparison on a cohort of women tested with and without colour Doppler imaging.

The specificity of ultrasound screening might also be improved by the development of precisely defined criteria to distinguish between abnormalities likely to be malignant and those likely to be benign and need no further investigation. However, there are difficulties with this strategy. Some abnormalities defined as 'benign' may, in fact, prove to be malignant, which may lead to a reluctance ignore them. This may result in large numbers of screened women being recalled for early rescreening and those abnormalities which

do not resolve may be operated on anyway. This is likely to increase the costs of screening, the anxiety and inconvenience of women, and may not ultimately reduce the intervention rate.

The natural history of benign ovarian abnormalities is not fully understood; there is some evidence that a proportion of benign ovarian neoplasms may undergo malignant change.[121] Furthermore, benign tumours may themselves cause clinical problems. It is likely, however, that benign tumours which become malignant or troublesome are in the minority and, thus, interventions for such tumours may occur in women who would never otherwise have been troubled by them. At present, there is a lack of information on this subject and further research is needed. Improving the specificity of ultrasound screening may thus depend on a greater understanding of the risks and benefits of surgical intervention for apparently benign screen-detected lesions. Should the removal of benign tumours be proved to reduce the subsequent risk of ovarian cancer, this will also increase the potential benefits of screening and reduce the potential negative outcome of unnecessary surgical intervention.

Tumour markers
For initial screening with tumour markers, the challenge is to improve the sensitivity, without compromising the high specificity which has been demonstrated. For the foreseeable future, any biochemical screening test for ovarian cancer will be based around CA 125. Likely developments will centre around identifying further markers which might complement CA 125 and increase its sensitivity, and the use of mathematical models using epidemiological information together with marker levels to define the risk of ovarian cancer.

The use of a model incorporating the rate of change of CA 125 level has been described,[122] and is being used in one RCT.[110] The main drawback to this approach is the high proportion of women who must wait several weeks for repeat testing before a decision on whether or not they require a scan is made. This approach results in a higher false-positive rate than use of only a single measurement of CA 125 levels but the relative effect on sensitivity has not been published.[113]

A number of newer markers have been investigated to assess whether their use together with CA 125 might increase the sensitivity of biochemical screening tests. Some encouraging preliminary results have been described for the markers, OVX-1 and M-CSF.[123] Indeed, one RCT originally planned to use OVX-1 as a marker

but, unfortunately, problems with the assay led to it being dropped from the screening protocol.[110] Further evaluation of these markers is necessary.

A range of other markers has been investigated in screened cohorts: lipid-associated sialic acid (LASA), NB/70K, H-neu, and urinary gonado-protein (UGP).[75,124] The results obtained so far suggest that simple cut-off points which might discriminate between women with early ovarian cancer and healthy women have yet to be defined.

There is increasing interest in the development of complex algorithms for combining multiple-marker results to increase discrimination. However, the capabilities of such models are still dependent on the development of tumour markers which discriminate clearly between women who have early ovarian cancer and women who do not.

The establishment of serum banks in large cohorts of women for whom there is accurate subsequent ascertainment of ovarian cancer incidence will facilitate the investigation of new markers as they are developed. All three of the RCTs currently in progress propose this.

Reducing the harms of screening
Minimising the harms of screening depends partly on maximising the specificity of screening tests. However, methods to reduce the risks associated with a false-positive test may also be important. There is a lack of published information on the nature and magnitude of such adverse effects and, therefore, a need to characterise the harms of screening more precisely. These risks might be reduced by a clearer protocol for women referred for diagnostic interventions as a result of screening positive or by the development of less invasive diagnostic techniques, such as greater use of laparoscopy. One prospective screening study used fine needle biopsy or cytology as a secondary test; this may be worthy of further investigation.[64] Finally, if the removal of benign abnormalities proves to be beneficial, this will offset the potential hazards of intervention. Relevant information on this question should be provided by the RCTs in progress, in particular, by the Bart's trial, in which the rate of gynaecological intervention in screened and control groups is being investigated.

Reducing the costs of screening
Screening costs could potentially be reduced if some fixed costs could be shared with other screening programmes, for example, breast screening. However, such a joint exercise may not be practical as it may increase the risk of confusion or error,

given the large number of recalls involved. The opportunity to share fixed costs may also be limited by the fact that the relevant facilities such as computers, facilities and staff have limited capacity and may already be fully committed to the breast screening programme.

One of the major determinants of screening costs is the frequency of screening; therefore, screening strategies which involve less frequent screening may prove more cost-effective. However, ultrasound screening, which may need to be performed less frequently, is also likely to be much more costly than CA 125 screening. The most promising route to cost-effective screening may therefore lie in a screening method based on an initial blood test, which can be performed without the need for large capital or training investment. This requires strategies to increase the sensitivity of such a screening method.

There is no obvious way to improve the cost-effectiveness of screening for the general population through improved test performance. The key issue is the low prevalence of ovarian cancer. One potential way forward therefore is to target screening at women who are at increased risk of ovarian cancer.

Targeting screening on a higher-risk population

In a higher-risk population, a greater proportion of women will develop ovarian cancer. For any given test sensitivity, the same proportion of ovarian cancers will be detected at screening as in the general population but the number of cancers

will be greater because of the higher prevalence. Each woman with ovarian cancer has the same probability that screening will detect the cancer and the same potential benefit compared with general population screening. However, because each woman screened has a higher risk of ovarian cancer, the likelihood of the test detecting cancer is greater. The likelihood of harm, however, is the same as in general population screening, assuming the same test specificity. Any benefits of screening depend on the benefit arising from early detection and treatment, and this has not been established for any risk group. Despite this, many centres in the UK are already offering screening as a service to some women considered to be at high risk.

The potential effect on the outcomes of screening of selecting a population at higher risk is illustrated in *Table 17.* In this illustration, the 'transvaginal sonography scenario' from *Table 16* is applied to two different high-risk populations:

- women who have one relative with ovarian cancer, who are assumed to be at three times the risk of the general population
- women with two affected first-degree relatives, who are assumed to be at ten times the risk of the general population.[28,31]

The remaining assumptions are the same as those for *Table 16.*

The table shows that while the probability of harm is the same as for the general population, the probability of an individual benefiting from screening, if it is effective, is greater and, therefore, the balance of risks and benefits

TABLE 17 *Annual outcomes of screening in a hypothetical cohort of 10,000 women aged 50–64 years at higher risk, assuming 40% mortality reduction and bi-annual TVS screening*

	Three times risk (1 in 830 per year)	Ten times risk (1 in 250 per year)
Number of women participating in screening programme	10,000	10,000
Screening interval	Every 2 years	Every 2 years
Number of screening tests carried out per year	5000	5000
Number of women recalled for further assessment per year who do not have primary ovarian cancer	350 (7% of screens)	350 (7% of screens)
Number of women undergoing surgery per year who do not have primary ovarian cancer	65 (1.3% of screens)	65 (1.3% of screens)
Maximum number of cancers detected on screening per year (if 100% sensitivity)	12	40
Number of additional 5-year survivors per year	4.8	16
Predictive value of recall (if 100% sensitivity)	3.3%	10.3%
Predictive value of diagnostic surgery (if 100% sensitivity)	16%	38.1%

is more favourable. The number of additional survivors per 10,000 women screened rises in proportion to the increase in risk of ovarian cancer.

A number of the assumptions made for the general population may differ for the high-risk population, however. The screening protocol proposed by the UKCCCR may be taken as typical of the approach currently adopted in screening this group.[114] This protocol proposes annual screening from age 25 years, which means that the average incidence of ovarian cancer will be lower than that assumed in *Table 17,* reducing the relative advantage of screening a higher-risk population. Secondly, the screening protocol is designed to maximise sensitivity at the expense of specificity, resulting in a higher false-positive rate – 5% is suggested as an acceptable level in the study protocol. Maximising sensitivity may improve the proportion of cancers detected at an early stage and, if screening is effective, result in a greater mortality reduction. In *Table 18* these issues are incorporated by assuming annual screening, a 60% reduction in mortality, a baseline incidence of 25 per 100,000 per year, a recall rate of 15% and a false-positive rate of 5%.

With these assumptions, both the benefits and harms of screening are increased compared with the illustration in *Table 17.* For women at moderately increased risk, the increased harms are large compared with the increased likelihood of benefit and, even for women at substantially increased risk, the benefit:harm ratio is less favourable than for general population screening. This is due to the assumption of a much higher false-positive rate, and it illustrates the importance of maintaining, even in this group, the specificity of the screening process.

However, women at significantly increased risk may have a higher level of anxiety about ovarian cancer and, for them, the value of being reassured by a true-negative result may be more important. This means that sensitivity must be maximised to achieve a higher negative predictive value.

Screening a higher-risk group means that fewer women must be screened for every case of ovarian cancer detected, which may improve the cost-effectiveness of screening. However, if a screening programme were to be established for higher-risk women, this would require a system to identify eligible women and call them for screening. This is much more complex than targeting screening by age and, in effect, amounts to a two-stage screening process. If screening is targeted at the groups at very high risk, such identification requires the compilation of a detailed pedigree, which requires skill and time. The costs of such an identification process would greatly increase the total costs of the screening programme and might well result in reduced cost-effectiveness compared to general population screening. The small numbers of women eligible for screening is likely to increase the average costs; it is estimated that there are around 50,000 women at very high risk in England and Wales.[114]

A screening programme for women at higher risk may be difficult to sustain, because public awareness that screening is being offered to some women could result in pressure for general population screening.

Finally, many women at significantly increased risk of ovarian cancer may be offered prophylactic oophorectomy. This reduces the risk of ovarian

TABLE 18 *Annual outcomes of screening in a hypothetical cohort of 10,000 women aged 25–75 years at higher risk, assuming 60% mortality reduction and annual screening*

	Three times risk (1 in 1300 per year)	Ten times risk (1 in 400 per year)
Number of women participating in screening programme	10,000	10,000
Screening interval	Annual	Annual
Number of screening tests carried out per year	10,000	10,000
Number of women recalled for further assessment per year who do not have primary ovarian cancer	1500 (15% of screens)	1500 (15% of screens)
Number of women undergoing surgery per year who do not have primary ovarian cancer	500 (5% of screens)	500 (5% of screens)
Maximum number of cancers detected on screening per year (if 100% sensitivity)	7.5	25
Number of additional survivors per year	4.5	15
Predictive value of recall (if 100% sensitivity)	0.5%	1.6%
Predictive value of diagnostic surgery (if 100% sensitivity)	1.5%	4.7%

cancer but there remains a small residual risk of disseminated intra-abdominal carcinoma. There are also other potential adverse effects associated with surgery and subsequent reduced oestrogen levels. A full assessment of the benefits and harms of this strategy is beyond the scope of this review but this is a potential alternative intervention to screening for this group. The vast majority of cancers in women at higher risk, like the general population, occur after the age of 40 years.[28]

Chapter 7
Remaining research questions

What are the benefits of screening for ovarian cancer?

The key question still to be answered is whether the use of currently available screening tests to screen for ovarian cancer in asymptomatic women will result in more benefit than harm – and at an acceptable cost. This can only be investigated reliably in an RCT in which the mortality from ovarian cancer in the screened and control groups is compared. Without such evidence, debate about the overall balance of costs, harms and benefits can only be based on information from models, whose assumptions may be unreliable. The trials in progress appear to be well-designed and large enough to estimate the impact of screening on ovarian cancer mortality, although the results will not be available for at least 5 years and successful completion may be dependent on securing additional funding.

The trials will, however, only look at the effect of screening on ovarian cancer mortality and not on the morbidity or quality of life experienced by women diagnosed with ovarian cancer. There is also the possibility that screening may have an impact on morbidity due to benign ovarian conditions, since many of these will be detected and removed as a result of screening. This effect may be either positive or negative, depending on the balance between operative morbidity and the morbidity of these conditions if treated conservatively. One of the RCTs will provide some information in the form of the numbers of operations undergone by screened and control groups.[110]

What are the harms of screening?

To judge the overall impact of screening on the health of a population requires information about the adverse effects of screening. These may include:

- operative morbidity in false-positives undergoing diagnostic surgery
- anxiety in women initially screened positive or with equivocal results who are recalled for further assessment
- possible over-diagnosis and over-treatment of women with borderline tumours and benign

conditions which might not otherwise cause any morbidity
- false reassurance in women who develop ovarian cancer following a negative screen result (interval cancers).

The published screening studies reviewed have not investigated these issues in detail and have reported few data on adverse effects of screening, even though the design of these studies would allow assessment of these questions. There is currently remarkably little published information about the consequences for women recalled or referred for diagnostic surgery who do not have ovarian cancer.

The RCTs in progress could provide some information relating to these issues. Data on complication rates is being collected and it is important that this information is published in a timely manner to enable an assessment of the risks experienced by women entering these trials. On completion of follow-up, the RCTs will also allow comparison of the incidence of ovarian cancer and borderline tumours in screened and unscreened groups.

However, none of the trials have planned to conduct detailed investigations on the impact of screening on women who are recalled and referred unnecessarily. This may be particularly important in the Bart's trial, where large numbers of women must wait for repeat measurements of CA 125 levels, with up to five recalls for retesting before a final decision is made. It would seem sensible to consider the value of additional research as part of these trials to investigate these issues.

What is the overall impact and the cost-effectiveness of screening?

Assessing the overall impact of screening involves weighing-up the probability of benefit and the probability of harm resulting from screening. The balance of benefits and harms which are judged 'acceptable' may vary between women, clinicians and policy-makers, and between

individuals. Research that increased our knowledge of women's views of risk, and how they determine acceptable levels of risk, could be valuable in assessing the circumstances in which screening might be judged worthwhile.

It is also important to assess the resources required for screening, to determine whether there may be more effective ways of deploying these resources. The Bart's trial has not, so far, secured funding for an economic evaluation – an important consideration in deciding a policy for screening. It would add value to the results of these trials to establish a collaborative economic analysis to enable an assessment of the relative cost-effectiveness of the different strategies.

Developing improved screening strategies

The potential impact of ovarian cancer screening might be improved in the following ways.

- Improvements to the sensitivity and specificity of screening methods.
- A better understanding of the optimum management and natural history of screen-detected benign conditions to reduce unnecessary intervention.
- Less invasive techniques for the diagnosis of ovarian cancer.

With the results of RCTs some years away, it is important that the evaluation of the performance of potential new screening methods is undertaken in such a way that the results are capable of being related to the results of these trials. The serum banks being established as part of these trials represent one way to achieve this for serum-based algorithms.

Research into the optimum management of screen-detected benign lesions could be incorporated into screening trials. Such research could consist of randomised comparisons of active and conservative management of abnormalities where there is uncertainty regarding the value of operative intervention. This might enable a reduction in the false-positive rate for ultrasound scans, if methods can be developed to characterise scan abnormalities which are at low risk of malignancy.

Screening women at higher risk of developing ovarian cancer

Until the RCTs have been completed, the effectiveness of screening for ovarian cancer remains unproven, regardless of the underlying risk of ovarian cancer. Screening a higher-risk group only changes the potential balance of benefit and harm – it does not establish benefit. Results from RCTs on the general population can be applied to a population at high risk, so long as the natural history (i.e. the speed with which ovarian cancer develops and progresses) is similar in the two groups.

Research in this area should therefore concentrate on investigating issues in this group which may be different from those in the general population. These include:

- the natural history of ovarian cancer, including stage at diagnosis, histological type and grade
- the age-specific risk of developing ovarian cancer
- the psychological impact of risk assessment and screening in this group, who may have a different level of anxiety compared with the general population
- women's perception of risk, the value they attach to knowledge of their individual risk and the effect of presenting information about risk in different ways.

These issues are most relevant to women at significantly increased risk, who have a history of more than one affected close relative. Investigation of the impact of screening in this group also requires investigation of methods of identifying women at higher risk cheaply and accurately, and the effect of this on the cost-effectiveness of screening.

Finally, the possibilities of genetic testing to more accurately characterise risk in individuals are increasing as more mutations are discovered. Such testing is expensive and little is known about its consequences or the purposes of testing if no effective interventions can be offered to the individual at high risk. This issue goes beyond the problem of screening for ovarian cancer, not least because many genetic mutations confer increased risk of cancer at more than one site, implying screening for several types of cancer. Research is needed into the impact of such screening on health outcomes at a population level and the levels of demand for such services.

Acknowledgements

References

1. Tortolero Luna G, Mitchell MF, Rhodes-Morris HE. Epidemiology and screening of ovarian cancer. *Obstet Gynecol Clin North Am* 1994;**21**:1–23.

2. Office for National Statistics. 1994 mortality statistics: cause. England and Wales. London: HMSO, 1994.

3. Office of Population Censuses and Surveys. 1989 cancer statistics: registrations. England and Wales. London: HMSO, 1989.

4. Roth LM, Czernobilsky B. General aspects of ovarian cancer. In: Roth L, Czernobilsky B, editors. Tumors and tumorlike conditions of the ovary. New York: Churchill Livingstone, 1985:1–9.

5. Richardson G, Scully R, Nikrui N, Nelson J. Common epithelial cancer of the ovary. *N Engl J Med* 1985;**312**:415–24.

6. Ovarian cancer: screening, treatment and follow-up. NIH consensus statement. Washington DC: National Institutes of Health, 1994:1–30.

7. Black R, Macfarlane GJ, Maisonneuve P, Boyle P. Cancer incidence and mortality in Scotland 1960–89. Edinburgh: Information and Statistics Division, NHS Scotland, 1995.

8. Mant J, Vessey M. Ovarian and endometrial cancers. In: Cancer surveys. Volume 19: Trends in cancer incidence and mortality, 1994:287–306.

9. Black R, Sharp L, Kendrick S. Trends in cancer survival in Scotland 1968–90. Edinburgh: Information and Statistics Division, NHS Scotland, 1993.

10. Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinu T, Esteve J. Survival of cancer patients in Europe. Lyon: IARC Scientific Publications, 1995.

11. Kosary C. FIGO stage, histology, histologic grade, age and race as prognostic factors in determining survival for cancers of the female gynaecological system. *Semin Surg Oncol* 1994;**10**:31–46.

12. Hand R, Fremgen A, Chmiel J, Recant W, Berk R, Sylvester J, *et al.* Staging procedures, clinical management and survival outcome for ovarian carcinoma. *JAMA* 1993;**269**:1119–22.

13. Kehoe S, Powell J, Wilson S, Woodman C. The influence of the operating surgeon's specialisation on patient survival in ovarian carcinoma. *Br J Cancer* 1994;**70**:1014–17.

14. Annual report on the results of treatment in gynecological cancer: twenty-first volume. *Int J Gynaecol Obstet* 1991;**36** suppl:1–315.

15. Hogberg T, Carstensen J, Simonsen E. Treatment, results and prognostic factors in a population based study of epithelial ovarian cancer. *Gynecol Oncol* 1993;**48**:38–49.

16. Young R, Walton L, Ellenberg S, Homesley HD, Wilbanks GD, Decker DG, *et al.* Adjuvant therapy in stage I and stage II epithelial ovarian cancer. *N Engl J Med* 1990;**322**:1021–7.

17. Parazzini F, Francheschi S, La Vecchia C, Fasoli M. The epidemiology of ovarian cancer. *Gynecol Oncol* 1991;**43**:9–23.

18. Westhoff C. Ovarian cancer. *Annu Rev Public Health* 1996;**17**:85–96.

19. Whittemore AS, Harris R, Itnyre J, group Coc. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case–control studies. II. Invasive epithelial ovarian cancers in white women. *Am J Epidemiol* 1992;**136**:1184–1203.

20. The reduction in risk of ovarian cancer associated with oral-contraceptive use. The cancer and steroid hormone study of the Center for Disease Control and the National Institute of Child Health and Human Development. *N Engl J Med* 1987;**316**:650–5.

21. Hankinson S, Colditz G, Hunter D, Spencer T, Rosner B, Stampfer M. A quantitative assessment of oral contraceptive use and risk of ovarian cancer. *Obstet Gynecol* 1992;**80**:708–14.

22. Kerlikowske K, Brown JS, Grady DG. Should women with familial ovarian cancer undergo prophylactic oophorectomy? *Obstet Gynecol* 1992;**80**:700–7.

23. Hartge P, Whittemore AS, Itnyre J, McGowan L, Cramer D, group Coc. Rates and risks of ovarian cancer in subgroups of white women in the United States. *Obstet Gynecol* 1994;**84**:760–4.

24. Fathalla MF. Incessant ovulation – a factor in ovarian neoplasia? *Lancet* 1971;**ii**:163.

25. Stadel BV. The etiology and prevention of ovarian cancer. *Am J Obstet Gynecol* 1975;**123**:772–4.

26. Whittemore AS, Harris R, Itnyre J, group Coc. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case–control studies. IV. The pathogenesis of epithelial ovarian cancer. *Am J Epidemiol* 1992;**136**:1212–20.

27. Schildkraut JM, Thompson WD. Familial ovarian cancer: a population based case–control study. *Am J Epidemiol* 1988;**128**:456–66.

28. Ponder BAJ. Familial ovarian cancer. In: Eeles RA, Ponder BAJ, Easton DF, Horwich A, editors. Genetic predisposition to cancer. London: Chapman and Hall, 1996:290–96.

29. Kerber RA, Slattery ML. The impact of family history on ovarian cancer risk. *Arch Intern Med* 1995;**155**:905–12.

30. Parazzini F, Negri E, La Vecchia C, Restelli C, Francheschi S. Family history of reproductive cancers and ovarian cancer risk: an Italian case–control study. *Am J Epidemiol* 1992;**135**:35–40.

31. Easton DF, Matthews FE, Ford D, Swerdlow AJ, Peto J. Cancer mortality in relatives of women with ovarian cancer: the OPCS study. *Int J Cancer* 1996;**65**:284–94.

32. Auranen A, Pukkala E, Makinen J, Sankila R, Grenman S, Salmi T. Cancer incidence in the first-degree relatives of ovarian cancer patients. *Br J Cancer* 1996;**74**:280–84.

33. Ford D, Easton DF. The genetics of breast and ovarian cancer. *Br J Cancer* 1995;**72**:805–12.

34. Eddy D. How to think about screening. In: Eddy D, editor. Common screening tests. Philadelphia: American College of Physicians, 1991:1–21.

35. Bourne TH, Campbell S, Reynolds KM, Whitehead MI, Hampson J, Royston P, *et al.* Screening for early familial ovarian cancer with transvaginal ultrasonography and colour blood flow imaging. *BMJ* 1993;**306**:1025–9.

36. Parkes CA, Smith D, Wald NJ, Bourne TH. Feasibility study of a randomised trial of ovarian cancer screening among the general population. *J Med Screen* 1994;**1**:209–14.

37. Pavlik EJ, van Nagell JR, DePriest PD, Wheeler L, Tatman JM, Boone M, *et al.* Participation in transvaginal ovarian cancer screening: compliance, correlation factors, and costs. *Gynecol Oncol* 1995;**57**:395–400.

38. Fleischer AC, James AE, Millis JB, Julian C. Differential diagnosis of pelvic masses by gray scale sonography. *Am J Roentgenol* 1978;**131**:469–76.

39. Granberg S, Norstrom A, Wikland M. Tumors in the lower pelvis as imaged by sonography. *Gynecol Oncol* 1990;**37**:224–9.

40. DePriest PD, Varner E, Powell J, Fried A, Puls L, Higgins R, *et al.* The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. *Gynecol Oncol* 1994;**55**:174–8.

41. Kurjak A, Shalan H, Kupesic S, Kosuta D, Sosic A, Benic S, *et al.* An attempt to screen asymptomatic women for ovarian and endometrial cancer with transvaginal color and pulsed Doppler sonography. *J Ultrasound Med* 1994;**13**:295–301.

42. Tekay A, Jouppila P. Controversies in assessment of ovarian tumors with transvaginal color Doppler ultrasound. *Acta Obstet Gynecol Scand* 1995;**75**:316–29.

43. Bast RC, Klug TL, St John ERN. A radio-immunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med* 1983;**309**:883–7.

44. Carlson KJ, Skates SJ, Singer DE. Screening for ovarian cancer. *Ann Intern Med* 1994;**121**:124–32.

45. Jacobs I, Bast RC. The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 1989;**4**:1–12.

46. Grover S, Quinn MA, Weideman P, Koh H. Factors influencing serum CA 125 levels in normal women. *Obstet Gynecol* 1992;**79**:511–14.

47. Hakama M, Stenman UH, Knekt P, Jarvisalo J, Hakulinen T, Maatela J, *et al.* CA 125 as a screening test for ovarian cancer. *J Med Screen* 1996;**3**:40–2.

48. Helzlsouer KJ, Bush TL, Alberg AJ, Bass KM, Zacur H, Comstock GW. Prospective study of serum CA-125 levels as markers of ovarian cancer. *JAMA* 1993;**269**:1123–6.

49. Zurawski VRJ, Orjaseter H, Andersen A, Jellum E. Elevated serum CA 125 levels prior to diagnosis of ovarian neoplasia: relevance for early detection of ovarian cancer. *Int J Cancer* 1988;**42**:677–80.

50. Zurawski VRJ, Broderick SF, Pickens P, Knapp RC, Bast RCJ. Serum CA 125 levels in a group of nonhospitalized women: relevance for the early detection of ovarian cancer. *Obstet Gynecol* 1987;**69**:606–11.

51. Einhorn N, Sjovall K, Knapp RC, Hall P, Scully RE, Bast RCJ, *et al.* Prospective evaluation of serum CA 125 levels for early detection of ovarian cancer. *Obstet Gynecol* 1992;**80**:14–18.

52. Zurawski VRJ, Sjovall K, Schoenfeld DA, Broderick SF, Hall P, Bast RCJ, *et al.* Prospective evaluation of serum CA 125 levels in a normal population, phase I: the specificities of single and serial determinations in testing for ovarian cancer. *Gynecol Oncol* 1990;**36**:299–305.

53. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. York: NHS Centre for Reviews and Dissemination, University of York, 1996.

54. Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests. Recommended methods, updated 6 June 1996. Available from: http://som.flinders.edu.au/FUSA/COCHRANE/1996.

55. Tabor A, Jensen FR, Bock JE, Hogdall CK. Feasibility study of a randomised trial of ovarian cancer screening. *J Med Screen* 1994;**1**:215–19.

56. Vuento MH, Pirhonen JP, Makinen JI, Laippala PJ, Gronroos M, Salmi TA. Evaluation of ovarian findings in asymptomatic postmenopausal women with color Doppler ultrasound. *Cancer* 1995;**76**:1214–18.

57. Millo R, Facca MC, Alberico S. Sonographic evaluation of ovarian volume in postmenopausal women: a screening test for ovarian cancer? *Clin Exp Obstet Gynecol* 1989;**16**:72–8.

58. Andolf E, Svalenius E, Astedt B. Ultrasonography for early detection of ovarian carcinoma. *Br J Obstet Gynaecol* 1986;**93**:1286–9.

59. Adonakis GL, Paraskevaidis E, Tsiga S, Seferiadis K, Lolis DE. A combined approach for the early detection of ovarian cancer in asymptomatic women. *Eur J Obstet Gynecol Reprod Biol* 1996;**65**:221–5.

60. Jacobs I, Stabile I, Bridges J, Kemsley P, Reynolds C, Grudzinskas J, *et al.* Multimodal approach to screening for ovarian cancer. *Lancet* 1988;**i**:268–71.

61. Grover S, Quinn MA, Weideman P, Koh H, Robinson HP, Rome R, *et al.* Screening for ovarian cancer using serum CA 125 and vaginal examination: report on 2550 females. *Int J Gynecol Cancer* 1995;**5**:291–5.

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

63. Jacobs I, Davies AP, Bridges J, Stabile I, Fay T, Lower A, *et al.* Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. *BMJ* 1993;**306**:1030–4.

64. Schincaglia P, Brondelli L, Cicognani A, Buzzi G, Orsini LF, Bovicelli L, *et al.* A feasibility study of ovarian cancer screening: does fine-needle aspiration improve ultrasound specificity? *Tumori* 1994;**80**:181–7.

65. van Nagell JRJ, Gallion HH, Pavlik EJ, DePriest PD. Ovarian cancer screening. *Cancer* 1995;**76** (10 suppl):2086–91.

66. Bourne TH, Campbell S, Reynolds K, Hampson J, Bhatt L, Crayford TJ, *et al.* The potential role of serum CA 125 in an ultrasound-based screening program for familial ovarian cancer. *Gynecol Oncol* 1994;**52**:379–85.

67. Sato S, Hasuo Y, Ohta S, Maruyama H, Kagiya A, Saito Y. Mass-screening for ovarian cancer by means of transvaginal ultrasonography – study on size of ovary to select subjects requiring second screening. *Nippon Sanka Fujinka Gakkai Zasshi* 1992;**44**:1173–9.

68. Goswamy RK, Campbell S, Whitehead MI. Screening for ovarian cancer. *Clin Obstet Gynaecol* 1983;**10**:621–43.

69. Demidov VN, Krasnikova SP, Terskaia LV. Role of echography in early detection of ovarian tumors. *Vopr Onkol* 1990;**36**:1365–8.

70. Kurjak A, Predanic M. New scoring system for prediction of ovarian malignancy based on transvaginal color Doppler sonography. *J Ultrasound Med* 1992;**11**:631–8.

71. Holbert TR. Screening transvaginal ultrasonography of postmenopausal women in a private office setting. *Am J Obstet Gynecol* 1994;**170**:1699–703.

72. Weiner Z, Thaler I, Beck D, Rottem S, Deutsch M, Brandes JM. Differentiating malignant from benign ovarian tumors with transvaginal color flow imaging. *Obstet Gynecol* 1992;**79**:159–62.

73. Karlan BY, Raffel LJ, Crvenkovic G, Smrt C, Chen MD, Lopez E, *et al.* A multidisciplinary approach to the early detection of ovarian carcinoma: rationale, protocol design, and early results. *Am J Obstet Gynecol* 1993;**169**:494–501.

74. Muto MG, Cramer DW, Brown DL, Welch WR, Harlow BL, Xu H, *et al.* Screening for ovarian cancer: the preliminary experience of a familial ovarian cancer center. *Gynecol Oncol* 1993;**51**:12–20.

75. Schwartz PE, Chambers JT, Taylor KJ. Early detection and screening for ovarian cancer. *J Cell Biochem* 1995;**23**:233–7.

76. Dorum A, Kristensen GB, Abeler VM, Trope CG, Moller P. Early detection of familial ovarian cancer. *Eur J Cancer* 1996;**32A**:1645–51.

77. Belinson JL, Okin C, Casey G, Ayoub A, Klein R, Hart WR, *et al.* The familial ovarian cancer registry: progress report. *Cleve Clin J Med* 1995;**62**:129–34.

78. Cuckle HS, Wald NJ. The evaluation of screening tests for ovarian cancer. In: Sharp F, Mason W, Leake R, editors. Ovarian cancer. London: Chapman and Hall, 1990:229–41.

79. Weiner Z, Beck D, Shteiner M, Borovik R, Ben-Shachar M, Robinzon E, *et al.* Screening for ovarian cancer in women with breast cancer with transvaginal sonography and color flow imaging. *J Ultrasound Med* 1993;**12**:387–93.

80. DePriest PD, van Nagell JRJ, Gallion HH, Shenson D, Hunter JE, Andrews SJ, *et al.* Ovarian cancer screening in asymptomatic postmenopausal women. *Gynecol Oncol* 1993;**51**:205–9.

81. Akulenko LV, Garkavtseva RF, Zhordania KI, Samgina AA. Current status and perspectives for genetic consultation and prophylactic medical examination of risk groups with malignant neoplasms of the female reproductive system and breast. *Tsitol Genet* 1992;**26**:38–42.

82. van Nagell JRJ, Higgins RV, Donaldson ES, Gallion HH, Powell DE, Pavlik EJ, *et al.* Transvaginal sonography as a screening method for ovarian cancer. A report of the first 1000 cases screened. *Cancer* 1990;**65**:573–7.

83. Jones MH, Singer A, Jenkins D. The mildly abnormal cervical smear: patient anxiety and choice of management. *J R Soc Med* 1996;**89**:257–60.

84. Swanson V, McIntosh IB, Power KG, Dobson H. The psychological effects of breast screening in terms of patients' perceived health anxieties. *Br J Clin Pract* 1996;**50**:129–35.

85. Gram IT, Lund E, Slenker SE. Quality of life following a false-positive mammogram. *Br J Cancer* 1990;**62**:1018–22.

86. Sutton S, Saidi G, Bickler G, Hunter J. Does routine screening for breast cancer raise anxiety? Results from a three wave prospective study in England. *J Epidemiol Commun Health* 1995;**49**:413–18.

87. Ellman R, Angeli N, Christians A, Moss S, Chamberlain J, Maguire P. Psychiatric morbidity associated with screening for breast cancer. *Br J Cancer* 1989;**60**:781–4.

88. Bull AR, Campbell MJ. Assessment of the psychological impact of a breast screening programme. *Br J Radiol* 1991;**64**:510–15.

89. Andolf E, Jorgensen C, Uddenberg N, Ursing I. Psychological effects of ultrasound screening for ovarian carcinoma. *J Psychosom Obstet Gynecol* 1990;**11**:155–62.

90. Pernet AL, Wardle J, Bourne TH, Whitehead MI, Campbell S, Collins WP. A qualitative evaluation of the experience of surgery after false positive results in screening for familial ovarian cancer. *Psychooncology* 1992;**1**:217–33.

91. Wardle JF, Collins W, Pernet AL, Whitehead MI, Bourne TH, Campbell LS. Psychological impact of screening for familial ovarian cancer. *J Natl Cancer Inst* 1993;**85**:653–7.

92. Wardle J, Pernet A, Collins W, Bourne T. False-positive results in ovarian cancer screening: one year follow-up of psychological status. *Psychol Health* 1994;**10**:33–40.

93. Wardle J, Pernet A, Stephens D. Psychological consequences of positive results in cervical cancer screening. *Psychol Health* 1995;**10**:185–94.

94. Wolfe CDA, Raju KS. The attitudes of women and feasibility of screening for ovarian and endometrial cancers in inner city practices. *Eur J Obstet Gynecol* 1994;**56**:117–20.

95. Tymstra T, Bielman B. The psychosocial impact of mass screening for disease. *Fam Pract* 1987;**4**:287–90.

96. Wardle J. Women at risk of ovarian cancer. *J Natl Cancer Inst Monogr* 1995;**17**:81–5.

97. Smith PM, Schwartz PE. New fears in gynecologic cancer. *Cancer* 1995;**76**(10 suppl):2133–7.

98. Leetanaporn R, Tintara H. A comparative study of outcome of laparoscopic salpingo-oophorectomy versus open salpingo-oophorectomy. *J Obstet Gynaecol* 1996;**22**:79–83.

99. Daniell JF, Kurtz BR, Lee JY. Laparoscopic oophorectomy: comparative study of ligatures, bipolar coagulation, and automatic stapling devices. *Obstet Gynecol* 1992;**80**:325–8.

100. Possover M, Morawski A, Hettenbach A. Laparoscopic treatment of ovarian tumors in menopausal women. *J Gynecol Obstet Biol Reprod (Paris)* 1994;**23**:784–9.

101. Yuen PM, Rogers MS. Laparoscopic removal of dermoid cysts using endopouch. *Aust N Z J Obstet Gynaecol* 1994;**33**:397–9.

102. Papasakelariou C, Saunders D, de la Rosa A. Comparative study of laparoscopic oophorectomy. *J Am Assoc Gynecol Laparosc* 1995;**2**(4):407–10.

103. Reich H, Johns DA, Davis G, Diamond MP. Laparoscopic oophorectomy. *J Reprod Med* 1993;**38**:497–501.

104. Mage G, Wattiez A, Canis M, Manhes H, Pouly JL, Bruhat MA. Apport de la coelioscopie dans le diagnostic precoce des cancers ovariens. [Contribution of celioscopy in the early diagnosis of ovarian cancers]. *Ann Chir* 1991;**45**:525–8.

105. Minelli L. Ovarian cysts. *Eur J Obstet Gynecol Reprod Biol* 1996;**65**:81–9.

106. Canis M, Mage G, Pouly JL, Wattiez A, Manhes H, Bruhat MA. Laparoscopic diagnosis of adnexal cystic masses: a 12-year experience with long-term follow-up. *Obstet Gynecol* 1994;**83**:707–12.

107. Yuen PM, Rogers MS. Laparoscopic management of ovarian masses: the initial experience and learning curve. *Aust N Z J Obstet Gynaecol* 1994;**34**:191–4.

108. Urban N, Drescher C, Etzioni R, Colby C. Use of a stochastic model to identify an efficient protocol for ovarian cancer screening. *Control Clin Trials* 1997;**18**:251–70.

109. Cohen C, Jennings T. Screening for ovarian cancer: the role of noninvasive imaging techniques. *Am J Obstet Gynecol* 1994;**170**:1088–94.

110. Randomised trial of screening for ovarian cancer (protocol). London: Ovarian Cancer Screening Unit, The Royal Hospitals Trust, 1995.

111. European randomised trial of ovarian cancer screening (protocol). London: Wolfson Institute of Preventive Medicine, Department of Environmental and Preventive Medicine, 1995.

112. Kramer BS, Gohagan J, Prorok P, Smart C. A National Cancer Institute sponsored screening trial for prostatic, lung, colorectal and ovarian carcinomas. *Cancer* 1993;**71**(suppl):589–93.

113. Skates SJ, Jacobs IJ, Sjovall K, Einhorn N, Xu F-J, Yu Y-H, *et al.* High sensitivity and specificity of screening for ovarian cancer with the risk of ovarian cancer (ROC) algorithm based on rising CA 125 levels. *J Clin Oncol* 1996;**14**:2007.

114. Jacobs I, Mackay J, Skates S. UKCCCR national familial ovarian cancer screening study (protocol). London: UKCCCR Gynaecological Subcommittee, 1997.

115. Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening: a decision analysis model. *Ann Intern Med* 1993;**118**:838–43.

116. Skates SJ, Singer DE. Quantifying the potential benefits of CA 125 screening for ovarian cancer. *J Clin Epidemiol* 1991;**44**:365–80.

117. Westhoff C, Randall MC. Ovarian cancer screening: potential effect on mortality. *Am J Obstet Gynecol* 1991;**165**:502–5.

118. Jacobs IJ, Oram DH, Bast RCJ. Strategies for improving the specificity of screening for ovarian cancer with tumor-associated antigens CA 125, CA 15-3, and TAG 72.3. *Obstet Gynecol* 1992;**80**:396–9.

119. Department of Health Advisory Committee. Breast cancer screening 1991: evidence and experience since the Forrest report. London: NHS Breast Screening Programme, 1991.

120. Wright CJ, Mueller CB. Screening mammography and public health policy: the need for perspective. *Lancet* 1995;**346**:29–32.

121. Puls L, Powell D, DePriest P, Gallion H, Hunter JE, Kryscio RJ, *et al.* Transition from benign to malignant epithelium in mucinous and serous ovarian cystadenocarcinoma. *Gynecol Oncol* 1992;**47**:53–7.

122. Skates SJ, Xu FJ, Yu YH, Sjovall K, Einhorn N, Chang YC, *et al.* Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. *Cancer* 1995;**76**:2004–10.

123. Woolas RP, Xu FJ, Jacobs IJ, Yu YH, Daly L, Berchuck A, *et al.* Elevation of multiple serum markers in patients with stage I ovarian cancer. *J Natl Cancer Inst* 1993;**85**:1748–51.

124. Cane P, Azen C, Lopez E, Platt LD, Karlan BY. Tumor marker trends in asymptomatic women at risk for ovarian cancer. *Gynecol Oncol* 1995;**57**:240–5.

125. Tay SK. Comparison of the usefulness of serum CA 125 level and a risk scoring system in detecting malignancy in ovarian cysts. *Ann Acad Med Singapore* 1995;**24**:168–71.

126. Grover SR, Quinn MA. Is there any value in bimanual pelvic examination as a screening test. *Med J Aust* 1995;**62**:408–10.

127. Sato S, Sugo T, Maruyama H, Saito Y, Hasuo Y. Mass-screening for ovarian cancer by transvaginal ultrasonography – study on tumor markers at the second screening. *Nippon Sanka Fujinka Gakkai Zasshi* 1994;**46**:1247–53.

128. Sato S, Sugo T, Maruyama H, Kunugi K, Saito Y, Hasuo Y. Mass-screening for ovarian cancer by transvaginal ultrasonography – study on ultrasonographic findings. *Nippon Sanka Fujinka Gakkai Zasshi* 1994;**46**:1234–40.

129. Pardo J, Kaplan B, Rosenberg C, Ovadia Y, Neri A. A modified transvaginal sonographic technique for better ovarian evaluation. *J Clin Ultrasound* 1993;**21**:503–5.

130. Campbell S, Bourne T, Bradley E. Screening for ovarian cancer by transvaginal sonography and colour Doppler. *Eur J Obstet Gynecol Reprod Biol* 1993;**49**:33–4.

131. Kuznesov VV. Selective screening on hormono-dependent tumours of women's reproductive system organs. *Eur J Gynaecol Oncol* 1992;**13**(1 suppl):65–8.

132. Kuznetzov VV, Semiglazov VF, Maximov SY. Selective screening of hormonodependent tumours in women's reproductive system organs. *Eur J Gynaecol Oncol* 1993;**14**(2):95–8.

133. van Nagell JRJ, DePriest PD, Gallion HH, Pavlik EJ. Ovarian cancer screening. *Cancer* 1993;**71**(4 suppl):1523–8.

134. Kobayashi H, Terao T. Field trial for the early detection of patients with ovarian cancer. *Rinsho Byori* 1992;**40**:139–45.

135. Kobayashi H, Sumimoto K, Terao T, Kawashima Y, Okada K. Field trial for the early detection of patients with ovarian cancer – discrimination of ovarian cancer patients by the statistical analysis using Mahalanobis' generalized distance. *Nippon Sanka Fujinka Gakkai Zasshi* 1992;**44**:174–80.

136. Kurjak A, Schulman H, Sosic A, Zalud I, Shalan H. Transvaginal ultrasound, color flow, and Doppler waveform of the postmenopausal adnexal mass. *Obstet Gynecol* 1992;**80**:917–21.

137. Ohmura M. Screening for ovarian cancer. *Jpn J Cancer Chemother* 1992;**19**:1958–63.

138. Konno R, Sato S, Yajima A. Transvaginal ultrasound screening for ovarian tumor. *Nippon Sanka Fujinka Gakkai Zasshi* 1992;**44**:109–12.

139. Sato S, Hasuo Y, Ohta S, Maruyama H, Kagiya A. Mass-screening for ovarian cancer by means of transvaginal ultrasonography. *Nippon Sanka Fujinka Gakkai Zasshi* 1992;**44**:683–8.

140. Schwartz PE, Chambers JT, Taylor KJ, Pellerito J, Hammers L, Cole LA, *et al.* Early detection of ovarian cancer: preliminary results of the Yale Early Detection Program. *Yale J Biol Med* 1991;**64**:573–82.

141. Bourne TH, Whitehead MI, Campbell S, Royston P, Bhan V, Collins WP. Ultrasound screening for familial ovarian cancer. *Gynecol Oncol* 1991;**43**:92–7.

142. van Nagell JRJ, DePriest PD, Puls LE, Donaldson ES, Gallion HH, Pavlik EJ, *et al.* Ovarian cancer screening in asymptomatic postmenopausal women by transvaginal sonography. *Cancer* 1991;**68**:458–62.

143. Osmers R, Volksen M, Rath W, Kuhn W. Vaginal ultrasound as a screening method for detection of adnexa tumors in postmenopause. *Onkologie* 1990;**13**:268–70.

144. Osmers R, Volksen M, Rath W, Kuhn W. Vaginal sonography a screening method for early detection of ovarian tumors and endometrial cancers. *Arch Gynecol Obstet* 1989;**245**:602–6.

145. Makarov OV, Dobrokhotova IE, Solomatina EV, Neverov AA. Ultrasonic screening of a group at risk of developing ovarian tumors. *Akush Ginekol (Sofiia)* 1990;**4**:71–2.

146. Duda V, Rode G, Thein C, Schulz KD. Vaginal sonography: pilot study for using in ovarian screening procedures. *Geburtshilfe Frauenheilkunde* 1990;**50**:388–93.

147. Campbell S, Royston P, Bhan V, Whitehead MI, Collins WP. Novel screening strategies for early ovarian cancer by transabdominal ultrasonography. *Br J Obstet Gynaecol* 1990;**97**:304–11.

148. Westhoff C, Gollub E, Patel J, Rivera H, Bast RJ. CA 125 levels in menopausal women. *Obstet Gynecol* 1990;**76**:428–31.

149. Kobayashi H, Sumimoto K, Terao T, Kawashima Y. Serodiagnostic tests by factor analysis and stepwise discriminating analysis with tumor markers for the detection of ovarian cancer. *Nippon Sanka Fujinka Gakkai Zasshi* 1989;**41**:1903–10.

150. Kobayashi H, Sumimoto K, Terao T, Kawashima Y, Kouda M. Field trial of the early detection in patients with ovarian cancer. *Nippon Sanka Fujinka Gakkai Zasshi* 1989;**41**:1743–9.

151. Kobayashi H, Sumimoto K, Terao T, Kawashima Y. Diagnostic value of serological tumor marker tests in patients with ovarian cancer. *Nippon Sanka Fujinka Gakkai Zasshi* 1989;**41**:1501–6.

152. Bhan V, Amso N, Whitehead MI, Campbell S, Royston P, Collins W. Characteristics of persistent ovarian masses in asymptomatic women. *Br J Obstet Gynaecol* 1989;**96**:1384–91.

153. Besson P, Constant M, Besson-Proye F. Transvaginal echography in gynecologic oncology. *Ann Radiol (Paris)* 1989;**32**:435–40.

154. Higgins RV, van Nagell JRJ, Donaldson ES, Gallion HH, Pavlik EJ, Endicott B, *et al.* Transvaginal sonography as a screening method for ovarian cancer. *Gynecol Oncol* 1989;**34**:402–6.

155. Alberico S, Facca MC, Millo R, Radillo L, Mandruzzato GP. Tumoral markers (CA 125-CEA) in the screening of ovarian cancer. *Eur J Gynaecol Oncol* 1988;**9**:485–9.

156. Rodriguez MH, Platt LD, Medearis AL, Lacarra M, Lobo RA. The use of transvaginal sonography for evaluation of postmenopausal ovarian size and morphology. *Am J Obstet Gynecol* 1988;**159**:810–14.

157. Goswamy RK, Campbell S, Royston JP, Bhan V, Battersby RH, Hall VJ, *et al.* Ovarian size in postmenopausal women. *Br J Obst Gynaecol* 1988;**95**:795–801.

158. Schoenfeld A, Levavi H, Hirsch M, Pardo J, Ovadia J. Transvaginal sonography in postmenopausal women. *J Clin Ultrasound* 1990;**18**:350–8.

159. Oram DH, Jacobs IJ, Brady JL, Prys-Davies A. Early diagnosis of ovarian cancer. *Br J Hosp Med* 1990;**44**:320–4.

160. Loskutova GP, Vesnin AG. Value of dispensary examination for the earliest possible detection of malignant ovarian tumors. *Vopr Onkol* 1977;**23**(3):75–8.

161. Loskutova GP. Clinical evaluation of the complex diagnosis of ovarian tumors. *Vopr Onkol* 1978;**24**(3):50–4.

162. Andolf E, Jorgensen C, Astedt B. Ultrasound examination for detection of ovarian carcinoma in risk groups. *Obstet Gynecol* 1990;**75**:106–9.

Appendix I

Search strategies

Search strategy for screening studies

001 exp ovarian neoplasms/
002 (ovar$ adj4 (cancer$ or tumo?r$ or malignan$)).ab.
003 (ovar$ adj4 (oncolog$ or carcinoma$)).ab.
004 (ovar$ adj4 (cancer$ or tumo?r$ or malignan$)).ti.
005 (ovar$ adj4 (oncolog$ or carcinom$)).ti.
006 (adnexa$ adj mass$).tw.
007 1 or 2 or 3 or 4 or 5 or 6
008 exp mass screening/
009 (screen$ or test$ or imag$ or predict$ or surveillance).tw.
010 exp population surveillance/
011 (earl$ adj2 diagnos$).ab.
012 (earl$ adj2 detect$).ab.
013 (earl$ adj2 (treatment$ or therap$)).ab.
014 (earl$ adj2 diagnos$).ti.
015 (earl$ adj2 detect$).ti.
016 (earl$ adj2 (treatment$ or therap$)).ti.
017 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
018 7 and 17
019 limit 18 to human
020 letter.pt.
021 19 not 20
022 21

Search strategies for studies of adverse effects of surgery in false-positives

001 oophorectomy.tw.
002 laparoscop$.tw.
003 cystectomy.tw.
004 diagnos$.tw.
005 ovariectomy/ae,px
006 exp hysterectomy/ae,mo,px
007 laparoscopy/ae,mo,px
008 laparotomy/ae,mo
009 exp anesthesia/ae,mo,px
010 1 or 2 or 3 or 4 or 5 or 6 or 7or 8.ti,ab,sh. or 9
011 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
012 exp postoperative complications/
013 exp intraoperative complications/
014 complication$.tw.
015 12 or 13 or 14
016 11 or 15
017 exp ovarian neoplasms/

018 16 and 17
019 11 and 15
020 17 and 19
021 20

Strategy for literature on adverse effects (psychological literature)

#1:   SCREENING
#2:   SCREENING in DE
#3:   SCREENING-TESTS
#4:   SCREENING-TESTS in DE
#5:   DIAGNOSIS
#6:   DIAGNOSIS in DE
#7:   #2 or #4 or #6
#8:   OVAR*
#9:   CANCER*
#10: NEOPLASM*
#11: MALIGNAN*
#12: OVAR* near4 (CANCER* or NEOPLASM* or MALIGNAN*)
#13: OVAR*
#14: CARCINOMA*
#15: TUMO?R*
#16: OVAR* near4 (CARCINOMA* or TUMO?R*)
#17: #12 or #16
#18: #7 and #17
#19: FALSE
#20: POSITIVE
#21: FALSE POSITIVE
#22: FALSE
#23: NEGATIVE
#24: FALSE NEGATIVE
#25: #21 or #24
#26: NEOPLASMS
#27: NEOPLASMS in DE
#28: OVARIES
#29: OVARIES in DE
#30: #27 and #29
#31: #25 and #30
#32: #25 and #29
#33: ADVERSE
#34: EFFECTS
#35: ADVERSE EFFECTS
#36: #30 and #35
#37: #21
#38: #21 and #26
#39: #24 and #27
#40: #18 or #38 or #39

Search strategy for economic literature

#1   OVAR*

#2   explode OVARIAN NEOPLASMS/ all subheadings

#3   explode MASS SCREENING/all subheadings

#4   #2 and #3

#5   explode COST (searched Costs and Cost Analysis)/all subheadings

#6   #4 and #5

#7   COST-EFFECTIV*

#8   #4 and #7

#9   COST*

#10   BENEFIT

#11   COST* NEAR BENEFIT

#12   #4 and #11

#13   COSTS

#14   #4 and #13

#15   #6 or #8 or #12 or #14

Appendix 2

Data extraction form

Reviewer   ....

1.   Study number(s)   ....

2.   Bibliographic details: First author   ....
                                              Journal   ....
                                              Year   ....
                                                                                                              Y/N
3.   Inclusion criteria: 1. Subjects not suspected of having ovarian cancer
                                              2. Histological confirmation of ovarian cancer
                                              3. Tests performed prior to diagnosis
                                              4. Subjects at risk of ovarian cancer (i.e. not selected to be disease-free)

Continue if all four fulfilled

4.   Country/region   ....

5.   Dates of recruitment   ....

6.   Methods of recruitment (tick any that apply):
                                              Self-referred following publicity   ....
                                              Referred by doctor   ....
                                              Written invitation   ....
                                              Other/comments   ....
                                              ....
                                              ....

7.   Inclusion/exclusion criteria:
                                              Age   ....
                                              Menopausal status   ....
                                              Family history   ....

                                              Other important comments ....
                                              ....
                                              ....

8.   Screening protocol:
        Summary ....

        (a) Description of initial test(s) (performed on all subjects)
        ....
        ....
        ....

        Definition of positive result (i.e. threshold for recall for further test(s))
        ....
        ....
        ....
        ....
        ....

(b) Description of follow-up test(s) if done:

....
....
....

Definition of positive result (i.e. resulting in referral for definitive diagnosis)

....
....
....
....
....

(c) Number of screening rounds completed and interval between them

....
....
....

9. Description of reference standard for positive tests (i.e. how was the definitive diagnosis made:
    note if diagnosis/stage reviewed or carried out by one hospital/team; note if there was a protocol
    for diagnostic and staging procedures; note if the number of women undergoing different
    procedures is stated)

    ....
    ....
    ....
    ....
    ....

10. Description of reference standard for negative tests (i.e. method and length of follow-up)

    ....
    ....
    ....
    ....
    ....

    Results (give separately for each screening round if appropriate or state if results may come from
    different screening rounds; if different tests carried out on all women, give results separately for
    each test)

11. Number of women screened

Total	....
Age range	....
Number invited for screening	....

Numbers under 50/50+ (or nearest cut-point given: state)   ....
Numbers pre/postmenopausal   ....
Numbers with/without family history   ....

Other information given ....
....

12. Number positive on initial tests (i.e. proceeding to further tests)

Total   ....

Numbers under 50/50+ (or nearest cut-point given:state)   ..../..............................
Numbers pre/postmenopausal   ..../..............................
Numbers with/without family history   ..../..............................

13. Number positive after further testing (i.e. proceeding to definitive diagnosis)

   Total ....

   Numbers under 50/50+ (or nearest cut-point given:state) ..../.............................
   Numbers pre/postmenopausal ..../.............................
   Numbers with/without family history ..../.............................

14. Drop-outs ....
   Uninterpretable tests ....
   Equivocal tests ....

15. Number of ovarian cancers in screen-positive women:

   Total ....
   Breakdown by Stage I–IV ....

   Numbers under 50/50+ (or nearest cut-point given: state) ..../.............................
   Numbers pre/postmenopausal ..../.............................
   Numbers with/without family history ..../.............................

16. Number of ovarian cancers in screen-negative women:

   Total ....
   Breakdown by Stage I–IV ....
   Breakdown by time since last screen: 1 year/2 year/more ....

   Numbers under 50/50+ (or nearest cut-point given: state) ..../.............................
   Numbers pre/postmenopausal ..../.............................
   Numbers with/without family history ..../.............................

   Completeness of follow-up: ....

17. Calculations (do separately for each screening round and each test if appropriate)

	Disease		
	+	–	
Test +	a	b	a + b
–	c	d	c + d
	a + c	b + d	a + b + c + d

   a:
   b:
   c:
   d ($\sim$c + d):
   a + b:
   c + d:
   a + c:
   b + d:
   a + b + c + d:

   (a) Prevalence of screen-detected cancer ($a/(a + b + c + d)$) per 100,000 ....
   (b) Sensitivity ($a/(a + c)$) ....
   (c) Specificity ($d/(b + d)$): note if estimated ....
   (d) False-positive rate ($b/(b + d)$) ....
   (e) False-negative rate ($c/(a + c)$) ....
   (f) PPV ($a/(a + b)$) ....
   (g) % of women recalled for further tests ....
   (h) % of women requiring surgery ....

Can calculations be performed for these subgroups?   Age 50+   ....
                                                                      Postmeno   ....
                                                                     Family history ....

18.   Quality check list

	Poor/ not stated	Fair	Good
(a) Ref standard for positives (method and validity)			
(b) Ref standard for negatives (method and completeness)			
(c) Clarity of cut-off points			
(d) Completeness of result reporting			
(e) Drop-out rates (through the screening process)			
(f) Description of study population with respect to important risk factors			
(g) If tests performed at the same time, was there blinding?			

18.   Further information:
    (a) Results of surgery in false-positive women: No abnormality   ....
                                                                                     Benign pathology   ....
                                                                                     Malignant pathology ....
    (b) Any women undergoing surgery outside screening protocol? ....
    (c) Reported surgical complications ....
    (d) Information about psychological outcomes ....
    ....
    ....
    ....

Appendix 3

Studies excluded from review of test performance

Author	Reason for exclusion
Tay[125]	Study on women undergoing surgery for ovarian cysts.
Grover & Quinn[126]	Multiple publication of data in Grover *et al.,* 1995[61].
Cane *et al.*[124]	Retrospective analysis of tumour marker levels in women without ovarian cancer.
Sato *et al.*[127]	Retrospective analysis of bank of tumour markers.
Sato *et al.*[128]	Description of US imaging of ovaries – not a screening study.
Pardo *et al.*[129]	No surgical intervention to detect ovarian cancer in women with abnormal results.
Campbell *et al.*[130]	No data reported.
Kuznetzov[131]	Not a screening study.
Kuznetsov *et al.*[132]	Not a screening study.
van Nagell *et al.*[133]	Multiple publication of data presented in DePriest *et al.,* 1993[80].
Koboyashi & Terao[134]	Retrospective analysis of battery of tumour markers in a cohort of women.
Koboyashi *et al.*[135]	Multiple publication of Kobayashi & Terao, 1992[134].
Kurjak & Predanic[70]	Includes women presenting with clinical symptoms suspicious of ovarian cancer.
Kurjak *et al.*[136]	Includes women presenting with clinical symptoms suspicious of ovarian cancer.
Ohmura[137]	A survey only.
Einhorn *et al.*[51]	No definitive intervention to diagnose ovarian cancer in women with abnormal test findings.
Konno *et al.*[138]	Multiple publication – same data as Sato *et al.,* 1992[139].
Schwartz *et al.*[140]	Multiple publication of data presented in Schwartz *et al.,* 1995[75].
Bourne *et al.*[141]	Multiple publication of data reported in Bourne *et al.,* 1993[35].
van Nagell *et al.*[142]	Multiple publication of data reported in DePriest *et al.,* 1993[80].
Osmers *et al.*[143]	Multiple publication of Osmers *et al.,* 1989[144].
Makarov *et al.*[145]	Measurement of ovarian volume in healthy women – no intervention to detect ovarian cancer.
Duda *et al.*[146]	Included women who had presented with clinical symptoms of ovarian cancer.
Campbell *et al.*[147]	Retrospective analysis of data reported in Campbell *et al.,* 1989[62].
Westhoff *et al.*[148]	Measurement of CA 125 levels in women assumed to be free of cancer – no intervention to detect ovarian cancer.
Zurawski *et al.*[52]	Follow-up study – no definitive intervention to detect ovarian cancer in women with abnormal findings.

continued

Author	Reason for exclusion
Kobayashi et al.[149]	Multiple publication of data presented in Kobayashi & Terao, 1992[134].
Kobayashi et al.[150]	Retrospective analysis of tumour markers in healthy women and women with cancer.
Koboyashi et al.[151]	Not a screening study.
Bhan et al.[152]	Multiple publication of data presented in Campbell et al., 1989[62].
Besson et al.[153]	Not a screening study.
Osmers et al.[144]	Included women who had presented with clinical symptoms of ovarian cancer.
Higgins et al.[154]	Multiple publication of data presented in van Nagell et al., 1990[82].
Alberico et al.[155]	No surgical intervention to detect ovarian cancer in women with abnormal test results.
Rodriguez et al.[156]	A study comparing ultrasound findings with histology in women undergoing non-ovarian gynaecological surgery. No cut-off point defined.
Goswamy et al.[157]	Describes ovarian volume in healthy women – no intervention to detect ovarian cancer in women with abnormal findings.
Schoenfeld et al.[158]	Women referred for ultrasound examination for clinical indications (not asymptomatic).
Oram[159]	Multiple publication of data reported in Jacobs et al., 1993[63].
Loskutova & Vesnin[160]	Not a study of screening for ovarian cancer but of the outcomes of general clinical health checks for women. No screening protocol described.
Loskutova[161]	Not a study of screening for ovarian cancer but of the outcomes of general clinical health checks for women. No screening protocol described.
Andolf et al.[162]	Included women who presented with clinical symptoms of ovarian cancer.

Appendix 4

Details of prospective screening studies included in review of test performance

Studies using grey-scale ultrasonography alone

	Goswamy et al.[68] 1983 UK	Andolf et al.[58] 1986 Sweden	Millo et al.[57] 1989 Italy	Campbell et al.[62] 1989 UK
Test	Transabdominal sonography (TAS).	TAS.	Transvaginal sonography (TVS).	TAS.
Screening protocol	Referred for diagnosis if abnormal morphology: entirely cystic, cystic with locules, cystic with solid areas, solid with irregular outline.	Definition of positive results not given. Abnormal findings rescanned unless 'surgery immediately necessary'.	Recalled for repeat scan if abnormal pathological morphological features. Criteria for diagnostic intervention not stated.	Recalled for repeat scan if hyper-/hypoecho-genicity, or irregular outline, or volume > 20 ml. Referred for diagnosis if persistent abnormality. Three screenings approximately 18 months apart.
Study population	Postmenopausal women aged 39–78 years. Self-referred.	Women aged 40–70 years contacting gynaecological outpatients department for 'a variety of reasons'.	Women aged 45+ years or post-menopausal; mean age, 54 years. Invitations to women attending for cervical screening – about 50% uptake.	Women aged over 45 years or with family history (4%): (age range, 18–78 years; mean age, 52 years). Self-referred.
Number of women	1084	805	500	**Screen 1:** 5479 **Screen 2:** 4914 **Screen 3:** 4201
Number positive on initial test (%) **Number attending for further tests (%)** **Number still being retested (%)**	Not stated.	83 (10.3)	28 (5.6)	334 (6.1) 346 (7.0) 312 (7.4)
Number positive after further tests (%)	34 (3.1)	50 (6.2)	Not clear.	– – –
Number undergoing diagnostic tests (%)	15 (1.4)	39 (4.8)	6 (1.2)	195 (3.6) 92 (1.9) 51 (1.2)
Cancers detected (rate per 1000) Total Stage I invasive Borderline	1 (0.9) 1 (0.9) –	3 (3.7) – 2 (2.5)	0 0 0	2 (0.4) 3 (0.6) 0 1 (0.2) 1 (0.2) 0 1 (0.2) 2 (0.4) 0
Probability of having ovarian cancer at diagnostic intervention (PPV:%) Any Stage I invasive	6.7 6.7	7.7 –	– –	1.0 3.3 – 0.5 1.1 –
Cancers arising in screen-negative women	Not stated.	Not stated.	Not stated.	0 at 1 year.
Follow-up of women screened negative	No information.	No information.	No information.	86% complete follow-up 1 year after last scan: method not stated.
Sensitivity at 1 year **Specificity**	– –	– –	– –	100 100 100? 96.5 98.2 98.8
Details of outcome in women undergoing diagnostic procedures	Benign disease: 14	No abnormality:4 Benign disease:32	Benign disease: 6	Malignant:4 No abnormality:12 Benign disease: remainder
Details of diagnostic procedures	Laparotomy or laparoscopy.	Laparotomy.	Not given.	Laparotomy or laparoscopy; abnormal tissues removed for histology. Histology reviewed centrally; classified according to WHO criteria.

Studies using grey-scale ultrasonography alone (continued)

Author / Year / Country	Demidov et al.[69] 1990 Russia	van Nagell et al.[82*] 1990 USA	DePriest et al.[80*] 1993 USA	van Nagell et al.[65*] 1995 USA	Tabor et al.[55] 1994 Denmark
Test	Ultrasound (not further specified).	TVS.	TVS.	TVS.	TVS.
Screening protocol	Women classified as positive on the basis of ovarian size – no further details given.	Recalled for repeat scan if ovarian volume > 18 ml (premenopausal). Referred for diagnosis if persistent enlargement (premenopausal) or volume > 8 ml (post-menopausal) or complex/solid areas seen at any scan.	First half of study: referred for diagnosis if ovarian volume > 8 ml or complex/solid areas. Second half: referred if abnormality persisted on repeat scan.	Recalled for repeat scan if ovarian volume > 10 ml (postmenopausal) or > 20 ml (premenopausal) or papillary projection into a cystic tumour. Referred for diagnosis if persistent abnormality.	Recalled for rescan if ovarian volume > 14 ml (premenopausal) or > 8 ml (postmenopausal), or irregular outline or hyper/hypoechogenicity. Referred for diagnosis if persistent, unless uni-locular cyst < 6 cm diameter with smooth walls and no septations.
Study population	Women aged 18 years and over. Invited for screening (sampling frame not stated).	Women aged 40 years and older. Mean age, 45 years. Recruitment method not stated.	Postmenopausal women aged 33–90 years. Mean age, 60 years. Recruitment method not stated.	Women aged either 50+ years and postmenopausal, or 25+ years with family history. Recruitment method not stated.	Women aged between 46 and 65 years invited to enter trial. 56% uptake.
Number of women	11,996	1000	3220	8500	435
Number positive on initial test (%)	Not stated.	–	Not stated.	Not stated.	54 (12.4)
Number attending for further tests (%)	"	54 (5.4)	"	"	54 (12.4)
Number still being retested (%)	"	–	"	"	–
Number positive after further tests (%)	262 (2.2)	31 (3.1)	–	–	9 (2.1)
Number undergoing diagnostic tests (%)	259 (2.2)	24 (2.4)	44 (1.4)	121 (1.4)	9 (2.1)
Cancers detected (rate per 1000): Total	11 (0.9)	0	3 (0.9)	8 (0.9)	0
Stage I invasive	4 (0.3)	0	2 (0.6)	6 (0.7)	0
Borderline	–	0	0	0	0
Probability of having ovarian cancer at diagnostic intervention (PPV:%) Any	4.2	–	6.8	6.7	–
Stage I invasive	1.5	–	4.5	5.0	–
Cancers arising in screen-negative women	Not stated.	Not stated.	Not stated.	One discovered incidentally (Stage II) within 1 year.	Not stated.
Follow-up of women screened negative	No information.	Invited for rescreening at 1 year: details of follow-up not stated.	Invited for rescreening at 1 year: details of follow-up not stated.	Invited for rescreening at 1 year: details of follow-up not stated.	No information.
Sensitivity (at 1 year)	–	–	–	88% maximum	–
Specificity	–	–	–	98.7	–
Details of outcome in women undergoing diagnostic procedures	Benign disease: 248 Malignant disease: 1	Benign disease: 23	Benign disease: 41 Malignant disease: 1	Benign disease: 112	Benign disease: 9
Details of diagnostic procedures	Surgery – no further details.	Exploratory laparotomy with sampling of ovarian tumours.	Exploratory laparotomy with removal and histology of ovarian tumour.	Laparotomy or laparoscopy and removal and histology of ovarian tumour.	Laparoscopy or laparotomy.

These three studies were undertaken by the same research team and the study populations overlap: they are presented separately because each reports slightly different information. Note that the reported study protocol is slightly different for each study.

Studies using ultrasonography with colour Doppler imaging

	Weiner et al.[72] 1993 Israel	Kurjak et al.[41] 1994 Croatia	Vuento et al.[56] 1995 Finland
Test	TVS and colour Doppler imaging (CDI).	TVS and CDI.	TVS and CDI.
Screening protocol	Recalled for repeat scan if ovarian volume > 20 ml, or cyst or mass present, or low impedance intra-ovarian blood vessels. Referred for laparotomy if persistent enlargement, complex cyst or low impedance blood flow. Patients with simple cysts followed-up with repeat scans.	Recalled for repeat scans if 'enlarged' ovary or cyst > 2.5 cm or resistance index < 0.4. Referred for diagnosis if persistent enlargement or cyst > 5 cm, or resistance index < 0.4 at any scan. Morphology score also calculated but role in decisions not clear.	Recalled for repeat scan if ovarian volume > 8 ml or not uniformly hypoechogeneic, or pulsatility index ≤ I. Referred for diagnosis if malignancy suspected. Other abnormalities followed-up, except simple cysts < 20 mm diameter.
Study population	Women with previous breast cancer referred from oncology clinic.	Women aged 40–71 years. Mean age, 45 years. Self-referred.	Women aged 56–61 years eligible for mammography screening invited. Mean age, 59 years. 74% uptake.
Number of women	600	5013	1364
Number positive on initial test (%) **Number attending for further tests (%)** **Number still being retested (%)**	– 100 (16.7) 46 (7.7)	– 424 (8.5) 316 (6.3)	– 160 (11.7) 23 (1.7)
Number positive after further tests (%) **Number undergoing diagnostic tests (%)**	18 (3.0) 12 (2.0)	– 38 (0.8)	5 (0.4) 5 (0.4)
Cancers detected (rate per 1000): Total Stage I invasive Borderline	3 (5.0) 1 (1.7) 0	4 (0.8) 4 (0.8) 0 (0)	1 (0.7) – 1 (0.7)
Probability of having ovarian cancer at diagnostic intervention (PPV:%) Any Stage I invasive	25.0 8.3	10.5 10.5	20.0 –
Cancers arising in screen-negative women	Not stated.	Not stated.	None at 1 year; one at 2 years (Stage I).
Follow-up of women screened negative	Screen-negatives recalled for rescan 1 year later – details of follow-up not given.	No information.	Women screened negative followed-up through Finnish cancer registry for 2.5 years.
Sensitivity (at 1 year) **Specificity**	– –	– –	100 99.7
Details of outcome in women undergoing diagnostic procedures	Benign disease: 8 Malignant disease: 1	No information.	Benign disease: 4
Details of diagnostic procedures	Exploratory laparotomy.	A minority TAH and BSO, the rest had 'less major' procedures. 'Equivocal' results followed-up at discretion of patient's own doctor.	Three laparotomies and two cyst aspirations.

Studies using grey-scale ultrasonography as an initial test, with CDI as a secondary test

	Bourne et al.[35] 1993 UK		Parkes et al.[36] 1994 UK
Test	TVS then CDI.		TVS then CDI.
Screening protocol	Rescanned if TVS showed areas of hyper- or hypoechogenicity. First 1000 referred for diagnosis if persistent morphological changes, unless volume reduced to < 63% of initial scan.	Rescanned if TVS showed areas of hyper- or hypoechogenicity. Second 601 referred for surgery if CDI showed pulsatility index < 1 or morphology score ≥ 5.	Recalled for CDI if ovarian volume more than 4 multiples of the age-specific median (premenopausal) or more than 3 multiples (postmenopausal). Referred for diagnosis if pulsatility index < 1 and peak systolic velocity > 10 cm/s in an area of abnormal morphology, or ovary or cyst > 5 cm diameter, or very abnormal morphology.
Study population	Women with a family history of ovarian cancer, age 17–79 years, mean age, 47 years. Self-referred.		Women aged 50–64 years attending breast screening (uptake 74% for randomisation).
Number of women	1000	601	2953
Number positive on initial test (%)	–		Not stated.
Number attending for further tests (%)	Altogether: 909 (56.8)		"
Number still being retested (%)	–		"
Number positive after further tests (%)	52 (5.2)	9 (1.5)	15 (0.5)
Number undergoing diagnostic tests (%)	52 (5.2)	9 (1.5)	14 (0.5)
Cancers detected (rate per 1000):			
Total	3 (3)	5.0 (3)	1 (0.3)
Stage I invasive	2 (2)	0	1 (0.3)
Borderline	1 (1)	3.3 (2)	0
Probability of having ovarian cancer at diagnostic intervention (PPV:%)			
Any	5.8	33.3	7.0
Stage I invasive	3.8	–	7.0
Cancers arising in screen-negative women	None at 6–16 months; four at 48 months – 1 Stage II, 3 Stage III.	None at 6–16 months.	One at 19 months: Stage I.
Follow-up of women screened negative	Patients contacted 6–16 months and asked about their health; 89% response rate. Completeness of follow-up at 4 years not stated.		No information.
Sensitivity (at 1 year)	100	100	100
Specificity	95.1	99.0	99.5
Details of outcome in women undergoing diagnostic procedures	No abnormality: 7 Benign disease: 48		Benign disease: 13
Details of diagnostic procedures	57 had laparotomy and bilateral oophorectomy; four had laparoscopy. Eight further patients had prophylactic oophorectomy.		Eight women underwent laparotomy, two laparoscopy and four had cysts aspirated. Ten further women referred for surgery outside the protocol.

Studies using grey-scale ultrasonography as an initial test, with other secondary tests

	Sato et al.[139] 1992 Japan	Schincaglia et al.[64] 1994 Italy	Holbert[71] 1994 USA
Author / Year / Country	Sato et al.[139] 1992 Japan	Schincaglia et al.[64] 1994 Italy	Holbert[71] 1994 USA
Test	TVS then computer tomography and magnetic resonance imaging scans and combination of tumour markers (CA 125, CA 19-9, TPA).	TAS then aspiration cytology or biopsy.	TVS then CA 125 test.
Screening protocol	Recalled for follow-up tests if TVS showed 'abnormal ultrasound findings' such as ovarian tumour > 30 mm diameter or ascites. Referred for diagnosis if tumour > 50 mm or complex in nature, or if tumour marker algorithm abnormal.	Initial scan: < 9 ml and cystic, repeat in 6 months. If enlarged, referred for biopsy/cytology. If 9–15 ml: repeat in 3 and 6 months; if same size, referred for biopsy/cytology. If > 15 ml: referred for biopsy/cytology. Biopsy/cytology: referred for diagnostic procedures if malignant cells seen, or enlarging complex/solid mass, or inadequate biopsy of complex/irregular mass, or cystic lesion recurring twice after aspiration.	Recalled for repeat scan and CA 125 if cystic ovaries present. Referred for diagnosis if cyst enlarging, or CA 125 > 135 U/ml, or patient choice.
Study population	Women aged over 30 years. Recruitment method not stated.	Postmenopausal women aged 50–69 years attending a breast clinic.	Postmenopausal women aged 30–89 years attending for annual routine examinations.
Number of women	15,282	3541	478
Number positive on initial test (%)	–	347 (9.8)	29 (6.1)
Number attending for further tests (%)	838 (5.5)	347 (9.8)	11 (2.3)
Number still being retested (%)	–	249 (7.0) underwent repeat scans, 98 (2.8) referred for biopsy/cytology.	–
Number positive after further tests (%)	–	19 (0.5)	–
Number undergoing diagnostic tests (%)	48 (0.3)	19 (0.5)	10 (2.1)
Cancers detected (rate per 1000):			
Total	1 'malignant', 1 'premalignant'	2 (0.6)	1 (2.1)
Stage I invasive	Not stated.	0	1 (2.1)
Borderline	Not stated.	0	0
Probability of having ovarian cancer at diagnostic intervention (PPV:%)			
Any	4.2	10.5	10.0
Stage I invasive	–	–	10.0
Cancers arising in screen-negative women	Not stated.	None at 1 year.	Not stated.
Follow-up of women screened negative	No details given.	Contacted at 12 months by questionnaire; also followed through cancer registry.	Women screened negative called for repeat screening at 1 year; 6.7% follow-up.
Sensitivity (at 1 year)	–	100	–
Specificity	–	99.5	–
Details of outcome in women undergoing diagnostic procedures	Benign disease: 46	Benign disease: 17	Benign disease: 9
Details of diagnostic procedures	Laparotomy.	Laparotomy.	Eight underwent laparoscopic oophorectomy; one, laparotomy and BSO; one, TAH and BSO.

Studies using CA 125 followed by ultrasonography

	Jacobs et al.[60] 1988 UK	Jacobs et al.[63] 1993 UK	Grover et al.[61] 1995 Australia
Test	CA 125 test followed by TAS.	CA 125 test followed by TAS.	CA 125 test followed by TAS or TVS.
Screening protocol	Recalled for TAS if CA 125 ≥ 30 U/ml. Referred for diagnosis if ovarian volume > 8.8 ml.	Recalled for TAS if CA 125 ≥ 30 U/ml. Referred for diagnosis if ovarian volume > 8.8 ml or non-uniform echogenicity or persistent morphological abnormality.	Recalled for ultrasonography and repeat CA 125 if initial CA 125 > 35 U/ml or if two measurements > 35 U/ml (premenopausal). Referred for diagnosis if scan in postmenopausal women showed cyst, or enlarged/asymmetrical ovaries; in premenopausal women if cyst > 6 cm or with unusual features; also if CA 125 level rising or very high.
Study population	Postmenopausal women aged 45+ years. Mean age, 54 years. Self-referred.	Postmenopausal women aged 45+ years. Median age, 56 years. Self-referred.	Women aged over 40 years or with family history (3%). Median age, 51 years. Self-referred.
Number of women	1010	22,000	2550
Number positive on initial test (%)	31 (3.1)	340 (1.6)	101 (4.0)
Number attending for further tests (%)	31 (3.1)	339 (1.5)	101 (4.0)
Number still being retested (%)	–	–	–
Number positive after further tests (%)	–	41 (0.2)	8 (0.3)
Number undergoing diagnostic tests (%)	3 (0.3)	41 (0.2)	8 (0.3)
Cancers detected (rate per 1000):			
Total	1 (1.0)	11 (0.5)	0
Stage I invasive	1 (1.0)	4 (0.2)	0
Borderline	0	0	0
Probability of having ovarian cancer at diagnostic intervention (PPV:%)			
Any	33.3	26.8	–
Stage I invasive	33.3	9.8	–
Cancers arising in screen-negative women	None after 1 year.	Three after 1 year; eight after 2 years (5 Stage I, 3 Stage III).	One after 10 months (discovered on repeated CA 125 testing).
Follow-up of women screened negative	Negative screens followed up by annual postal questionnaire.	Negative screens followed-up by annual postal questionnaire: 98.7% response at 1 year, 57% at 2 years.	Questionnaire at 12 months (results and completeness of follow-up not stated).
Sensitivity (at 1 year)	100	78.6	–
Specificity	99.8	99.9	99.7
Details of outcome in women undergoing diagnostic procedures	Benign cyst 1. No abnormality: 1. Those with negative ultrasound scan (44.6/1000) followed-up with CA 125 test every 3 months for 1 year.	Benign pathology: 25. Malignancy: 3. No abnormality: 2	Benign disease: 6. No abnormality: 2
Details of diagnostic procedures	Two laparotomies; one not stated.	Laparotomy or laparoscopy.	Three laparoscopies; five laparotomies.

Studies using CA 125 followed by ultrasonography (continued)

	Adonakis et al.[59] 1996 Greece	Bourne et al.[66] 1994 UK				
Author Year Country						
Test	CA 125 test followed by TVS.	CA 125 and TVS then CDI: all subjects received both tests : retrospective analysis of effects of different levels of CA 125 as an initial screen.				
Screening protocol	Recalled for TVS if CA 125 ≥ 35 U/ml. Referred for diagnosis if ovarian volume > 18 ml (premenopausal) or > 8 ml (postmenopausal), or hyper-/hypoechogenicity or irregular outline.	Recalled for TVS: various values of CA 125. Rescanned if TVS showed areas of hyper-/hypoechogenicity. First 1000: referred for surgery if persistent morphological changes, unless the volume had reduced to < 63% of initial scan. Next 601: referred for surgery if CDI showed pulsatility index < 1 or morphology score ≥ 5.				
Study population	Women aged over 45 years. Mean age, 58 years. Self-referred/invited.	Women with a family history of ovarian cancer; age 17–79 years; mean age, 47 years. Self-referred.				
Number of women	2000	1601 (TVS) 1502 (CA 125 available)				
		TVS	CA 125 cut-off point			
			20	25	30	35
Number positive on initial test (%)	18 (0.9)	1601 (100.0)	379 (25.2)	242 (16.1)	129 (8.5)	83 (5.5)
Number attending for further tests (%)	18 (0.9)					
Number still being retested (%)						
Number positive after further tests (%)	14 (0.7)	61 (3.8)	21 (1.1)	17 (1.1)	13 (0.9)	10 (0.7)
Number undergoing diagnostic tests (%)	14 (0.7)					
Cancers detected (rate per 1000):						
Total	2 (1.0)	6 (3.7)	5 (3.3)	4 (2.7)	3 (2.0)	3 (2.0)
Stage I invasive	1 (0.5)	2 (1.2)	2 (1.3)	2 (1.3)	1 (0.7)	1 (0.7)
Borderline	1 (0.5)	3 (1.9)	2 (1.3)	1 (0.7)	1 (0.7)	1 (0.7)
Probability of having ovarian cancer at diagnostic intervention (PPV:%):						
Any	14.3	9.8	23.8	23.5	23.0	30.0
Stage I invasive	7.1	3.3	9.5	11.8	7.7	10.0
Cancers arising in screen-negative women	None after 1 year.	0	1	2	3	3
		4 arising 2–4 years after screening (not detected by ultrasound): 1 Stage II, 3 Stage III; one had CA 125 level > 35 U/ml, rest < 20 U/ml.				
Follow-up of women screened negative	Negative screens followed-up at second screening 1 year later. Completeness of follow-up not stated.	Negative screens contacted 6–16 months after screening to enquire about health, with 89% response. Also asked to inform if cancer developed. Completeness of follow-up at 4 years not given.				
Sensitivity (at 1 year)	100*	100*	83.3*	66.7*	50.0*	50.0*
Specificity	99.4	96.6	98.9	99.1	99.3	99.5
Details of outcome in women undergoing diagnostic procedures	Benign pathology: 12	See results given on page 70 (Studies using ultrasonography with CDI).				
Details of diagnostic procedures	2 laparoscopies; 12 laparotomies.	Most (58/62) underwent laparotomy and BSO. Remaining four had no abnormality on laparoscopy.				

* Compared with ultrasound screening

Studies using CA 125 test and ultrasonography

	Akulenko et al.[81] 1992 Russia	Karlan et al.[73] 1993 USA	Muto et al.[74] 1993 USA
Test	CA 125 and ultrasonography and CA 19-9 and REA.	TVS and CDI and CA 125.	CA 125 and TVS.
Screening protocol	No details given of criteria for recall or for referral for diagnosis.	Tests repeated if adnexal mass ≥ 5 cm, abnormal ovarian architecture, resistance index < 0.4, or CA 125 > 35 U/ml. Referred for diagnosis if persistent abnormality on scan.	Recalled for repeat testing if CA 125 ≥ 35 U/ml, or complex or large (> 2 cm) cyst (premenopausal), or any ovarian mass (postmenopausal). Referred for diagnosis if CA 125 doubled or rose by > 95 U/ml, or scan showed persistent mass.
Study population	Women aged over 18 years responding to an invitation to attend for a health check with a family history of breast, ovarian or endometrial cancer.	Women aged over 35 years with family history of ovarian, breast, endometrial or colon cancer. Self- or physician-referred.	Women aged 25 years and over with a family history of ovarian cancer. Self- or physician-referred.
Number of women	1003	597	384
Number positive on initial test (%) **Number attending for further tests (%)** **Number still being retested (%)**	– Not stated. –	– 115 (19.2) –	89 (23.2) scan, 42 (10.9) CA 125 55 (14.3) scan, 32 (8.3) CA 125 34 (8.8) scan, 10 (2.6) CA 125
Number positive after further tests (%) **Number undergoing diagnostic tests (%)**	– 14	10 (1.7) – not clear whether follow-up completed. –	15 (3.9) scan, 8 (2.1) CA 125 15 (3.9)
Cancers detected (rate per 1000): Total Stage I invasive Borderline	 1 (1.0) Not stated. –	 1 (1.7) 0 1 (1.7)	 0 0 0
Probability of having ovarian cancer at diagnostic intervention (PPV:%) Any Stage I invasive	 7.1 –	 10.0 –	 – –
Cancers arising in screen-negative women	Not stated.	Not stated.	Not stated.
Follow-up of women screened negative	No information given.	Those testing negative recalled for second screen after 6 months – 25% follow-up completed.	Women screened negative followed-up with repeat screening; completeness of follow-up unclear.
Sensitivity (at 1 year) **Specificity**	– –	– –	– –
Details of outcome in women undergoing diagnostic procedures	Benign disease: 11 No abnormality: 2	Benign: 8 Malignant: 1 (endometrial carcinoma)	Benign disease: 13 No abnormality: 1 Malignant disease: 1
Details of diagnostic procedures	No information.	Laparotomy or laparoscopy and oophorectomy. Nine further women underwent prophylactic oophorectomy.	Eight TAH and BSO; four bilateral or unilateral oophorectomy; one cyst aspiration; one cystectomy; one laparoscopy. Four women had surgery outside the protocol and 19 had prophylactic oophorectomy.

Studies using CA 125 test and ultrasonography (continued)

Author Year Country	Schwartz et al.[75] 1995 USA	Belinson et al.[77] 1995 USA	Dorum et al.[76] 1996 Norway
Test	CA 125 and TVS and CDI.	CA 125 and TVS and CDI.	CA 125 and TVS.
Screening protocol	Recalled for repeat tests if CA 125 > 35 U/ml, or resistance index < 0.5, or other abnormality. Criteria for diagnostic intervention not stated.	Recalled for repeat test if CA 125 > 35 U/ml, or ovarian volume > 18 ml (premenopausal) or > 8 ml (postmenopausal), or abnormal morphology, or abnormal resistance index. Criteria for diagnostic intervention not stated.	Recalled for repeat TVS if simple uni- or bilocular cyst > 2 cm diameter. Recalled for repeat CA 125 if > 35 U/ml. Referred for diagnosis if tumour observed at first scan, or if persistent cyst on repeat scanning, or if CA 125 > 35 U/ml at repeat testing.
Study population	Women aged over 30 years with family history of ovarian cancer. Median age, 42.5 years. Self-referred.	Women aged over 23 years with family history of ovarian cancer. Mean age, 43 years. Self- or physician-referred.	Women aged over 18 years (mean age, 43 years), with two relatives with breast or ovarian cancer or one relative with breast and ovarian cancer. Self- or physician-referred.
Number of women	247	137	180
Number positive on initial test (%) Number attending for further tests (%) Number still being retested (%)	Not clear (initial screening not complete). – –	Not clear. – –	Not stated. – –
Number positive after further tests (%) Number undergoing diagnostic tests (%)	Not clear. 1 (0.4)	Not clear. 2 (1.5)	16 (8.9) 16 (8.9)
Cancers detected (rate per 1000): Total Stage I invasive Borderline	0 0 0	1 (7.3) 0 0	7 (39) 0 (0) 3 (17)
Probability of having ovarian cancer at diagnostic intervention (PPV:%) Any Stage I invasive	– –	50.0 –	44 –
Cancers arising in screen-negative women	Not stated.	Not stated.	None
Follow-up of women screened negative	Women screened negative recalled for further screening at 6 months; no details of completeness of follow-up.	No information.	56% returned for annual screening.
Sensitivity (at 1 year) Specificity	– –	– –	– –
Details of outcome in women undergoing diagnostic procedures	Benign disease: 1	No information.	Benign conditions: 9 Further 13 women underwent oophorectomy as treatment for breast cancer; two had ovarian cancer not detected at TVS.
Details of diagnostic procedures	One woman underwent laparotomy for suspected malignancy; not clear if this the only woman undergoing surgery as a consequence of the screening.	One laparoscopy and one laparotomy.	14, laparotomy and oophorectomy; two, laparoscopy.

Studies using pelvic examination

	Andolf et al.[58] 1986 Sweden	Jacobs et al.[60] 1988 UK	Grover et al.[61] 1995 Australia	Adonakis et al.[59] 1996 Greece
Test	Pelvic examination followed by TAS.	Pelvic examination followed by TAS.	Pelvic examination followed by TAS or TVS.	Pelvic examination followed by TVS.
Screening protocol	Definition of positive results not given. Abnormal findings recalled for TAS. Criteria for diagnosis not stated.	Recalled for TAS if palpable pelvic mass. Referred for diagnosis if ovarian volume > 8.8 ml.	Recalled for ultrasonography if adnexal mass (postmenopausal) or abnormally large ovary (premenopausal). Referred for diagnosis if scan showed cyst, enlarged or asymmetrical ovaries (postmenopausal), or if cyst > 6 cm or with unusual features (premenopausal).	Recalled for TVS if palpable adnexal mass or inadequate pelvic examination. Referred for diagnosis if ovarian volume > 18 ml premenopausal or > 8 ml postmenopausal, or hyper-/hypoechogenicity or irregular outline.
Study population	Women aged 40–70 years contacting gynaecological outpatients department for 'a variety of reasons'.	Postmenopausal women aged 45+ years. Self-referred.	Women aged over 40 years or with family history of ovarian cancer (3%).	Women aged over 45 years; age range, 45–80 years; mean, 58 years. Self-referred.
Number of women	805	1010	2550	2000
Number positive on initial test (%)	13 (1.6)	28 (2.8)	40 (1.6)	Abnormal: 59 (3.0) · Abnormal or ambiguous: 174 (8.7)
Number undergoing diagnostic tests (%)	8 (1.0)	11 (1.1)	10 (0.4)	27 (1.4) · 33 (1.7)
Cancers detected (rate per 1000):				
Total	0	1 (1.0)	0	1 (0.5) · 2 (1.0)
Stage I invasive	0	1 (1.0)	0	1 (0.5) · 1 (0.5)
Borderline	0	0	0	0 · 1 (0.5)
Probability of having ovarian cancer at diagnostic intervention (PPV:%):				
Any	—	9.1	—	3.7 · 6.1
Stage I invasive	—	9.1	—	3.7 · 3.0
Cancers arising in screen-negative women	Three detected by TAS.	None.	None.	1 (found on CA 125 testing) · 0
Details of outcome of women undergoing diagnostic procedures	Benign disease: 8	Benign cyst: 8 · No abnormality: 1 · Declined surgery: 1	Benign disease: 9 · No abnormality: 1	Benign pathology: 27 · Malignant: 1 · No abnormality: 3
Details of diagnostic procedures	Laparotomy or laparoscopy.	Not given.	Laparotomy or laparoscopy.	Laparoscopy or laparotomy.

Appendix 5
Details of modelling studies

Author	Aim and design of study	Population; screening procedure and assumptions	Assumptions about benefits and harms of screening	Cost data	Findings and comments
Westhoff & Randall[117]	To illustrate the potential effect of screening on ovarian cancer mortality.	Cohort of women screened annually between the ages of 45 and 74 years. 80% sensitivity of test; all screen-detected cancers assumed to be at Stage I. Test specificity varied.	Stage-specific survival rates taken from observed survival rates in the USA; benefits calculated as the number of extra 5-year survivors per year. No account taken of potential harms of screening.	None.	6000 screening tests needed to produce one extra 5-year survivor (if 80% of cancers diagnosed at Stage I). Screening test specificity of 98% results in 50 positive test results per cancer detected. Comment: similar methodology to Parkes *et al.* Cannot calculate number of life-years gained.
Parkes *et al.*[36]	To illustrate the potential effect of screening on ovarian cancer mortality. Comparison of expected cost-effectiveness of screening for ovarian cancer compared with breast and cervical cancer.	Cohort of women screened every 3 years between the ages of 50 and 64 years. Mortality reduction at 5 years estimated on the basis of varying proportions of cancers diagnosed at Stage I in the screened population.	5-year survival at Stage I, 75%; at later stages, 16%. Proportion of cancers diagnosed at Stage I in unscreened population, 33%. Life-expectancy of each additional survivor at 5 years, 19.3 years. No account taken of potential harms of screening.	Cost per screen assumed to be £20 for ovarian cancer and same for breast cancer.	If screening increases proportion of Stage I cancers to 80%, mortality reduction of 43% at 5 years. Cost per life-year saved about twice that for breast cancer screening and a little less than that for cervical screening. Comment: cost-effectiveness comparisons depend on screening interval of 3 years which may be over-optimistic.
Schapira *et al.*[115]	To illustrate the net benefit in terms of average life-expectancy resulting from screening for ovarian cancer, using a decision-analysis model.	Cohorts of 40-year-old and 65-year-old women resident in the USA. Screening occurs once only at age 40 or 65 years. Screening test sensitivity 45% for early disease and 81% for late disease; 50% of prevalent cases in early stage. Test specificity, 99.95%.	Survival from early disease, 26.8 years at age 40 years, 18.3 years at age 65 years; survival from late disease 3.4 and 2.7 years, respectively. Probability of death following diagnostic laparotomy, 0.2% at age 40 years and 1.5% at age 65 years.	None.	Screening increased average life-expectancy by one-third of a day in 40-year-olds and by three-quarters of a day in 65-year-olds. Screening reduced average life-expectancy at a screening specificity of 98.5%. Comment: the probability of death after laparotomy is probably an overestimate and is based on mortality from staging laparotomies. Use of average benefit makes the size and distribution of benefits and harms unclear.
Skates & Singer[116]	To illustrate potential benefit of CA 125 screening using a stochastic model.	Annual CA 125 screening for women aged between 50 and 75 years. Models the detection of ovarian cancer assuming CA125 levels rise exponentially with time and that a level over 35 U/ml results in detection. Natural history modelled on the basis of clinical opinion – Stage I assumed to last 9 months on average.	Survival following diagnosis based on observed stage-specific survival curves. No account taken of potential harms of screening.	None.	44% of cases detected at an earlier stage; 7.7 years of life gained for each case detected at screening, and 3.4 for each ovarian cancer diagnosed. Comment: use of stochastic simulation enables more sophisticated modelling of the dynamic nature of cancer development. Gains predicted are lower than Parkes *et al.* and slightly greater than Schapira *et al.* but of the same order of magnitude.

continued

Author	Aim and design of study	Population; screening procedure and assumptions	Assumptions about benefits and harms of screening	Cost data	Findings and comments
Urban[108]	To compare the relative cost-effectiveness of different screening strategies, measured as cost per life-year saved.	Screening women aged 50–79 years with a variety of strategies. CA 125 model similar to that developed by Skates & Singer; TVS model based on reported false-positive rates and sensitivity equivalent to 88% for annual screening.	Survival post-diagnosis based on currently observed stage-specific survival curves. Probability of death at laparoscopy assumed to be 0.1%; this is combined with benefits to give average life-years saved with each strategy.	Based on charges in the USA: $40 for CA 125, $150 for TVS, and $3000 for laparoscopy. Treatment costs also incorporated.	Found that annual CA 125 test followed by TVS if level elevated or rising was the most cost-effective strategy; it resulted in fewer life-years saved compared with annual TVS but the cost per life-year saved was less. This finding was robust over a range of sensitivity analyses. Comment: the total life-years gained was equivalent to only about 1 year for every case of ovarian cancer. Charge data may be inaccurate compared with true costs.

HTA panel membership

This report was identified as a priority by the Population Screening Panel.

Acute Sector Panel
Chair: Professor John Farndon, University of Bristol †

Professor Senga Bond,
University of Newcastle-
upon-Tyne †

Professor Ian Cameron,
Southeast Thames Regional
Health Authority

Ms Lynne Clemence,
Mid-Kent Health Care Trust †

Professor Francis Creed,
University of Manchester †

Professor Cam Donaldson,
University of Aberdeen

Mr John Dunning,
Papworth Hospital,
Cambridge †

Professor Richard Ellis,
St James's University Hospital,
Leeds

Mr Leonard Fenwick,
Freeman Group of Hospitals,
Newcastle-upon-Tyne †

Dr David Field,
Leicester Royal Infirmary †

Ms Grace Gibbs,
West Middlesex University
Hospital NHS Trust †

Dr Neville Goodman,
Southmead Hospital
Services Trust,
Bristol †

Mr Ian Hammond,
Bedford & Shires Health &
Care NHS Trust

Professor Adrian Harris,
Churchill Hospital, Oxford

Professor Robert Hawkins,
University of Bristol †

Dr Chris McCall,
General Practitioner, Dorset †

Professor Alan McGregor,
St Thomas's Hospital, London

Mrs Wilma MacPherson,
St Thomas's & Guy's Hospitals,
London

Professor Jon Nicoll,
University of Sheffield †

Professor John Norman,
University of Southampton

Dr John Pounsford,
Frenchay Hospital, Bristol †

Professor Gordon Stirrat,
St Michael's Hospital,
Bristol

Professor Michael Sheppard,
Queen Elizabeth Hospital,
Birmingham †

Dr William Tarnow-Mordi,
University of Dundee

Professor Kenneth Taylor,
Hammersmith Hospital,
London

Diagnostics and Imaging Panel
Chair: Professor Mike Smith, University of Leeds †

Professor Michael Maisey,
Guy's & St Thomas's Hospitals,
London *

Professor Andrew Adam,
UMDS, London †

Dr Pat Cooke,
RDRD, Trent Regional
Health Authority

Ms Julia Davison,
St Bartholomew's Hospital,
London †

Professor Adrian Dixon,
University of Cambridge †

Professor MA Ferguson-Smith,
University of Cambridge †

Dr Mansel Hacney,
University of Manchester

Professor Sean Hilton,
St George's Hospital
Medical School,
London

Mr John Hutton,
MEDTAP International Inc.,
London †

Professor Donald Jeffries,
St Bartholomew's Hospital,
London †

Dr Andrew Moore,
Editor, *Bandolier* †

Professor Chris Price,
London Hospital Medical
School †

Dr Ian Reynolds,
Nottingham Health Authority

Professor Colin Roberts,
University of Wales College
of Medicine †

Miss Annette Sergeant,
Chase Farm Hospital,
Enfield

Professor John Stuart,
University of Birmingham

Dr Ala Szczepura,
University of Warwick †

Mr Stephen Thornton,
Cambridge & Huntingdon
Health Commission

Dr Gillian Vivian,
Royal Cornwall Hospitals Trust †

Dr Jo Walsworth-Bell,
South Staffordshire
Health Authority †

Dr Greg Warner,
General Practitioner,
Hampshire †

Methodology Panel
Chair: Professor Martin Buxton, Brunel University †

Professor Anthony Culyer,
University of York *

Dr Doug Altman, Institute of
Health Sciences, Oxford †

Professor Michael Baum,
Royal Marsden Hospital

Professor Nick Black,
London School of Hygiene
& Tropical Medicine †

Professor Ann Bowling,
University College London
Medical School †

Dr Rory Collins,
University of Oxford

Professor George Davey-Smith,
University of Bristol

Dr Vikki Entwistle,
University of Aberdeen †

Professor Ray Fitzpatrick,
University of Oxford †

Professor Stephen Frankel,
University of Bristol

Dr Stephen Harrison,
University of Leeds

Mr Philip Hewitson,
Leeds FHSA

Professor Richard Lilford,
Regional Director, R&D,
West Midlands †

Mr Nick Mays, King's Fund,
London †

Professor Ian Russell,
University of York †

Professor David Sackett,
Centre for Evidence Based
Medicine, Oxford †

Dr Maurice Slevin,
St Bartholomew's Hospital,
London

Dr David Spiegelhalter,
Institute of Public Health,
Cambridge †

Professor Charles Warlow,
Western General Hospital,
Edinburgh †

* Previous Chair
† Current members

continued

continued

Pharmaceutical Panel
Chair: Professor Tom Walley, University of Liverpool †

Professor Michael Rawlins,
University of Newcastle-
upon-Tyne*

Dr Colin Bradley,
University of Birmingham

Professor Alasdair
Breckenridge, RDRD,
Northwest Regional
Health Authority

Ms Christine Clark,
Hope Hospital, Salford †

Mrs Julie Dent,
Ealing, Hammersmith
& Hounslow Health Authority,
London

Mr Barrie Dowdeswell,
Royal Victoria Infirmary,
Newcastle-upon-Tyne

Dr Desmond Fitzgerald,
Mere, Bucklow Hill,
Cheshire

Dr Alistair Gray,
Health Economics Research
Unit, University of Oxford †

Professor Keith Gull,
University of Manchester

Dr Keith Jones,
Medicines Control Agency

Professor Trevor Jones,
ABPI, London †

Ms Sally Knight,
Lister Hospital, Stevenage †

Dr Andrew Mortimore,
Southampton & SW Hants
Health Authority †

Mr Nigel Offen, Essex Rivers
Healthcare, Colchester †

Dr John Posnett,
University of York

Mrs Marianne Rigge,
The College of Health,
London †

Mr Simon Robbins,
Camden & Islington
Health Authority, London †

Dr Frances Rotblat,
Medicines Control Agency †

Mrs Katrina Simister,
Liverpool Health Authority †

Dr Ross Taylor,
University of Aberdeen †

Dr Tim van Zwanenberg,
Northern Regional Health
Authority

Dr Kent Woods, RDRD,
Trent RO, Sheffield †

Population Screening Panel
Chair: Professor Sir John Grimley Evans, Radcliffe Infirmary, Oxford †

Dr Sheila Adam,
Department of Health*

Ms Stella Burnside,
Altnagelvin Hospitals Trust,
Londonderry †

Dr Carol Dezateux, Institute of
Child Health, London †

Dr Anne Dixon Brown,
NHS Executive,
Anglia & Oxford †

Professor Dian Donnai,
St Mary's Hospital,
Manchester †

Dr Tom Fahey,
University of Bristol †

Mrs Gillian Fletcher,
National Childbirth Trust †

Professor George Freeman,
Charing Cross & Westminster
Medical School, London

Dr Mike Gill, Brent & Harrow
Health Authority †

Dr JA Muir Gray, RDRD,
Anglia & Oxford RO †

Dr Ann Ludbrook,
University of Aberdeen †

Professor Alexander Markham,
St James's University Hospital,
Leeds †

Professor Theresa Marteau,
UMDS, London

Dr Ann McPherson,
General Practitioner,
Oxford †

Professor Catherine Peckham,
Institute of Child Health,
London

Dr Connie Smith,
Parkside NHS Trust, London

Dr Sarah Stewart-Brown,
University of Oxford †

Ms Polly Toynbee,
Journalist †

Professor Nick Wald,
University of London †

Professor Ciaran Woodman,
Centre for Cancer
Epidemiology, Manchester

Primary and Community Care Panel
Chair: Dr John Tripp, Royal Devon & Exeter Healthcare NHS Trust †

Professor Angela Coulter,
King's Fund, London *

Professor Martin Roland,
University of Manchester *

Dr Simon Allison,
University of Nottingham

Mr Kevin Barton,
East London & City
Health Authority †

Professor John Bond,
University of Newcastle-
upon-Tyne †

Ms Judith Brodie,
Age Concern, London †

Dr Nicky Cullum,
University of York †

Professor Shah Ebrahim,
Royal Free Hospital, London

Mr Andrew Farmer,
Institute of Health Sciences,
Oxford †

Ms Cathy Gritzner,
The Patients' Association †

Professor Andrew Haines,
RDRD, North Thames
Regional Health Authority

Dr Nicholas Hicks,
Oxfordshire Health Authority †

Professor Richard Hobbs,
University of Birmingham †

Professor Allen Hutchinson,
University of Sheffield †

Mr Edward Jones,
Rochdale FHSA

Professor Roger Jones,
UMDS, London

Mr Lionel Joyce,
Chief Executive, Newcastle City
Health NHS Trust

Professor Martin Knapp,
London School of Economics
& Political Science

Professor Karen Luker,
University of Liverpool

Professor David Mant,
NHS Executive South & West †

Dr Fiona Moss, North Thames
British Postgraduate
Medical Federation †

Professor Dianne Newham,
King's College London

Professor Gillian Parker,
University of Leicester †

Dr Robert Peveler,
University of Southampton †

Dr Mary Renfrew,
University of Oxford

Ms Hilary Scott,
Tower Hamlets Healthcare
NHS Trust, London †

* Previous Chair
† Current members

National Coordinating Centre for
Health Technology Assessment, Advisory Group

Chair: Professor John Gabbay, Wessex Institute for Health Research & Development [†]

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639   Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta