SCREENING FOR OVARIAN CANCER: A SYSTEMATIC REVIEW 1998

EXECUTIVE SUMMARY

Background

Ovarian cancer is the seventh most common cancer in women world wide, and in England and Wales the mortality rate is 14.7 per 100, 000 females per year. Nearly half of all cases occur in women aged between 50 and 69, and in this age group the annual incidence of ovarian cancer is around 45 per 100, 000. The overall five-year survival rate is poor, at about 30%, with minimal improvement in this figure over the past 20-30 years. Survival is much better, around 75% at five years, for women whose disease is localised to the ovaries (FIGO stage I), but only about a quarter of cases in the UK are currently diagnosed at this stage. This has led to interest in methods to detect ovarian cancer in asymptomatic women, in the hope that population screening might result in earlier diagnosis and reduce mortality and morbidity from this disease. The tests which have been most extensively evaluated as screening methods include ultrasound scanning, and the measurement of serum levels of CA125, a tumour marker produced by most ovarian cancers. When used for screening, CA125 measurement is followed by ultrasound scanning in women with abnormal CA125 levels ('CA125 based screening'). Women with persistently abnormal findings are then referred for diagnostic laparotomy or laparoscopy under general anaesthesia for removal of the ovaries.

The current status of the effectiveness of screening, and trials in progress

Deciding whether a screening programme should be established depends on the balance between the potential benefits of screening in terms of improved

outcome for women with ovarian cancer; the harms of screening resulting from testing and investigating healthy women; and the resources required. The impact of screening on ovarian cancer mortality can only be reliably assessed by a randomised controlled trial (RCT) comparing similar groups of screened and unscreened women.

No RCTs of screening for ovarian cancer have been completed. This means that there is currently no reliable evidence that screening can improve outcomes for women with ovarian cancer (including those at higher risk from the disease). In the absence of evidence of effectiveness, it would be premature to establish any kind of screening programme.

Three large RCTs are currently in progress, two of which are based in the UK. One of the UK based trials is evaluating transvaginal ultrasound as a screening test, and the other CA125 based screening, with ultrasound as a follow-up test for women with elevated or rising CA125 levels. If successfully completed, in about 5-7 years' time these trials will provide evidence as to whether or not screening can reduce ovarian cancer mortality. They will also provide an estimate of the perioperative mortality and complication rates in women referred for diagnostic surgery. However, the trials currently plan to provide little additional information concerning potential harms of screening, in particular the psychological impact of screening and the broader effects on morbidity of diagnostic surgery in false positives. One trial currently plans to provide an estimate of the cost-effectiveness of screening. The value of the information provided by these trials would be enhanced if an assessment of the relative cost-effectiveness of their different screening strategies was undertaken.

Screening test performance

Evidence relating to the performance of ultrasound scanning and CA125 as screening tests can be obtained from prospective screening studies. Such studies can provide information on intermediate outcomes such as detection rates, false positive rates and the stage at diagnosis of screen detected cancer. They cannot provide reliable evidence about the effectiveness of screening, which depends further on whether earlier detection and treatment results in improved outcomes. It should also be noted that the cut-off points and protocols

used in these prospective studies varied widely; studies using the same screening test did not necessarily use the same criteria for defining positive results, and not all the studies specified the definitions of abnormal results; the full screening protocol was frequently not described fully. In addition, the conclusions regarding the impact of screening on stage distribution are based on the initial screening rounds (the prevalence screen); in these the proportion of early cancers may be lower than in subsequent screening rounds, and this may underestimate any benefit of screening.

Available evidence however suggests that both CA125 based screening and ultrasound screening can detect a higher proportion of ovarian cancers at stage I compared with that currently observed in the UK - around 50% diagnosed at stage I in CA125 based screening studies and around 75% in ultrasound screening studies. These data should be interpreted cautiously, however, because they are based on small numbers of cancers detected in diverse studies carried out mainly on self-selected women.

From the limited follow-up reported in published screening studies, annual screening with ultrasound appears to have a sensitivity close to 100%. The reported sensitivity of annual CA125 based screening is around 80%. The precision of these estimates is low, however, as they are based on small numbers of cancers.

The effect of different screening intervals on the detection rate and false positive rate has not been investigated. Less frequent screening may reduce the proportion of cancers detected at screening, but may also reduce the number of unnecessary investigations and the cost of screening. Intervals for ultrasound scanning of between one and three years are under investigation in the RCTs, while CA125 based screening has been carried out annually.

About 1.2-2.5% of women screened by ultrasound scanning have persistently abnormal findings resulting in referral for diagnostic surgery, but are found not to have ovarian cancer. The figure is lower for CA125 based screening, around 0.1-0.6%. Such diagnostic surgery carries a risk of complications such as infection, excessive bleeding, and more seriously, damage to the bladder or bowel. There is also a small risk of death. These risks are difficult to quantify and give only a limited picture of the impact of false positive screening results,

but perhaps 0.5-1% of women undergoing diagnostic surgery will suffer a significant complication. Most women referred for surgery who do not have ovarian cancer will be found to have a benign ovarian tumour or other benign gynaecological condition. The extent to which surgical intervention may benefit these women, by averting future clinical problems or perhaps reducing ovarian cancer risk, is unknown. There is however 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. Further research is required to determine whether this is the case.
The number of women finally classified as positive on screening and referred for diagnostic surgery is low compared with the number who initially have abnormal or equivocal test results. Perhaps 3-12% of screened women are recalled for further testing and assessment, resulting in potential distress and anxiety to otherwise healthy women, before they finally receive the reassurance of a negative result. There may be a lengthy period before this final decision is made.

The potential impact of screening for ovarian cancer

Typically, annual ultrasound screening of 10, 000 women aged 50-69 at average risk might result in 700 women being recalled for further assessment, 130 undergoing diagnostic surgery and 4 cancers detected (assuming 100% sensitivity) of which 2 - 4 may be stage I; a positive predictive value (PPV) of 3% for surgery and 0.6% for initial recall. Annual CA125 based screening might typically result in 300 women being recalled, 20 women undergoing diagnostic surgery and 3 cancers detected (assuming 75% sensitivity), of which 1-2 may be stage 1. This implies a PPV of 15% for surgery and 1% for initial recall.
The relatively low prevalence of ovarian cancer may limit the potential cost-effectiveness of general population screening. Compared to breast cancer, ovarian cancer causes around one third the number of deaths. This implies that to achieve comparable cost-effectiveness, screening for ovarian cancer would need to result in much greater relative reduction in mortality than breast screening (which reduces mortality by around 40% in screened women), or would need to be much less costly. If the optimum screening interval for ovarian

cancer is less than 3 years, then the overall costs of any screening programme may be greater.

Comparing the performance of screening tests involves consideration of the balance between the detection rate, the false positive rate, and the costs. Evidence from prospective screening studies suggests that ultrasound screening is more sensitive than CA125 based screening, but that the latter method may result in a smaller proportion of false positives and hence a higher positive predictive value. However, a less sensitive test must be repeated more frequently to achieve the same overall detection rate of ovarian cancers, and this may reduce the apparent advantages of CA125 based screening. The screening method and interval resulting in the best overall balance of potential benefits, harms and costs is currently unknown, but modelling studies suggest that annual CA125 based screening may provide lower overall benefits but at greater cost-effectiveness than annual ultrasound screening.

A number of potential improvements to screening tests are under development. It is suggested that the addition of colour Doppler imaging (CDI) to ultrasound screening may reduce the false positive rate, but mixed results have been reported. The additional impact of CDI on the sensitivity, false positive rates and costs of ultrasound screening requires clarification. The use of mathematical models, incorporating epidemiological data and CA125 levels, to determine thresholds for defining abnormal results has also been proposed, and is being evaluated in one of the RCTs in progress. Further developments of methods based on serological testing depend on the further evaluation of newer tumour markers. At present, none has been shown to improve overall performance compared with CA125 alone.

It is important that newly developed tests or screening strategies are evaluated in such a way that the findings can be related to the results of RCTs. This may mean increased use of study designs which directly compare the performance of different screening methods in the same group of women, since this increases the validity of the results. The impact of any newer methods on the overall cost-effectiveness of screening also needs to be considered.

Screening a higher risk population

A family history of ovarian cancer is one of the strongest risk factors for developing ovarian cancer. However, only about 7% of women with ovarian cancer report a family history of the disease. Most of these women have only one affected relative and are at only modestly increased risk, on average 2-3 times that of a woman with no family history. A small minority report more than one affected close relative. These women are at substantially increased risk, around 10 times that of the general population on average. This is equivalent to a 15% lifetime risk of developing the disease. Screening is currently being offered as a service in some UK centres to this latter group of women.

Screening women at higher risk does not alter the potential benefit of screening for each woman with ovarian cancer. The higher prevalence, however, means that fewer women need to be screened to detect each case of cancer, and there are fewer false positives for every case detected, increasing the positive predictive value. The balance of potential benefits, harms and costs may therefore be more favourable. The costs of identifying high risk women need to be taken into account, however, when considering the overall cost-effectiveness of this approach.

Until RCTs have been completed, there is no evidence as to whether screening women at higher risk is effective in reducing mortality. Further research is required before a full assessment of the potential benefits, harms and costs of screening can be made. Until such information is available, it is premature to establish a screening programme, including services to seek out women at higher risk in order to offer them screening.

The results from RCTs in the general population could be used to model the impact of screening in different risk groups, if the natural history of ovarian cancer is similar. If, however, the disease progresses at a different rate in women at higher risk, the results may not be applicable. They will also be relevant only to the screening methods evaluated in the RCTs, and if different screening methods are proposed for higher risk women, comparison of their performance with the methods used in the trials will be necessary. Research in the higher risk group should therefore be directed towards areas in which there may be differences with the general population, such as the natural history,

screening test performance, and the age-specific risks of developing ovarian cancer.

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. The identification of some such mutations raises the possibility of testing individuals in these families to determine whether they are carriers, potentially enabling more accurate assessment of risk. This is not yet possible for many families, but this is a rapidly evolving field. Carriers of some specific mutations may have a lifetime risk of developing ovarian cancer as high as 50-60%, although there is little epidemiological data on which to base such risk assessments. Mutations frequently predispose to risk at several cancer sites and for some of these, screening tests of proven effectiveness are available. The implications of genetic testing for cancer risk are therefore broad, and go beyond the scope of this review. Consideration should be given to specific research to address the policy implications of these developments.

Conclusions: implications for policy and research

1. Screening for ovarian cancer is currently unproven as a strategy for improving outcomes for women with ovarian cancer. Screening programmes should therefore not be considered until further research provides a better understanding of the potential benefits, harms and costs involved. While awaiting the results of the current trials, demand for screening is likely to increase, and a strong national lead on this is required.
2. RCTs currently underway should, in 5-7 years, give an estimate of any impact of screening on ovarian cancer mortality. Information from the trials would be enhanced by extending their investigation of the adverse effects of screening, and by ensuring that comparisons of the cost-effectiveness of the different screening strategies evaluated can be undertaken.
3. Research into screening test performance has frequently been poorly reported, and has made insufficient use of designs which enable assessment of the relative performance of different test methods. Future developments to screening tests should be compared with the tests being evaluated in the trials, to enable an assessment of their marginal impact on potential benefits, harms and costs. Test developments which require further evaluation include the marginal impact of adding colour Doppler imaging to ultrasound screening; the use of CA125 levels in multivariate algorithms to determine thresholds for ultrasound and surgical intervention; and the marginal value of adding CA125 measurement to ultrasound screening. It should also be noted that the screening modalitites reviewed in this report are continuously evolving; this makes evaluation difficult, and specification of the protocol particularly important. These

modalities will require continuous re-evaluation in line with technical developments.

4. The relatively low prevalence of ovarian cancer means that the positive predictive value of screening tests, even those with very high specificity, is low. Since the consequence of a false positive result is a surgical procedure, consideration of the overall impact of ovarian cancer screening, and not only the potential benefits, is important. The low prevalence also limits the potential cost-effectiveness of population screening.
5. The balance of potential benefits, harms and costs of screening may be more favourable in the small group of women who are at significantly increased risk due to a strong family history. However, benefit from screening has not been established and therefore there is no case for establishing a screening programme in this group. No RCTs are planned in a higher risk population, but a screening study has recently been established. This will provide some evaluation of intermediate outcomes of screening, but may also increase demand for screening services.
6. Evidence of potential effectiveness of screening in women at higher risk could be extrapolated from the results of trials on women recruited from the general population. However, this will only produce valid results if the natural history of ovarian cancer is similar for these women, and for the screening strategies used in the trials. Research efforts should be directed towards evaluating the effectiveness and cost effectiveness of screening strategies for this higher risk group. This includes investigation of any differences in the natural history; the performance of screening tests compared with the strategies used in the RCTs; investigation of age-specific risk of developing ovarian cancer, and investigation of the psychological impact and value of risk assessment.