University of Sheffield

ScHARR

School of Health and Related Research

An economic evaluation of the cost-effectiveness of screening for ovarian cancer amongst post-menopausal women who are not at high risk of ovarian cancer.

Report for the National Screening Committee.

May 2016.

Benjamin Kearns
Jim Chilcott
Sophie Whyte
Neill Calvert
Louise Preston

School of Health and Related Research (ScHARR)
Regent Court, 30 Regent Street, Sheffield, S1 4DA.
Telephone: 0114 222 6380.

Contents

Executive Summary

Background

Amongst women in the United Kingdom, ovarian cancer is the fourth highest cause of cancer deaths, and has the highest mortality of all the gynaecological cancers. Survival is highly dependent upon stage at diagnosis with five year survival rates ranging from less than 20% for women diagnosed with the most advanced stage of disease to 90% for women diagnosed with the least advanced stage of disease. The majority of women present with advanced disease, which has been attributed to the lack of disease-specific symptoms. Hence screening has the potential to detect ovarian cancers earlier, which may in turn lead to improvements in mortality.

The UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is the largest ovarian cancer screening trial to date. It was designed to evaluate the performance of two screening strategies compared to no screening. The screening strategies involved first line screening with either the CA-125 blood test interpreted using a risk of ovarian cancer algorithm (ROCA) or with a transvaginal ultrasound scan (TVS). Both strategies included second-line screening with TVS. The CA-125 screening group was referred to as multimodal screening (MMS) group, with the TVS group referred to as the ultrasound screening (USS) group. Women were randomised to one of the three trial arms between April 2001 and October 2005, with results based on a median follow-up of 11.1 years published in December 2015.

The aim of this project was to conduct an economic evaluation of the cost-effectiveness of screening for ovarian cancer amongst postmenopausal women who are not at high risk of ovarian cancer. To support this economic evaluation a detailed systematic review of the published health economic evidence relating to screening and treatment for ovarian cancer was conducted. As the largest, most recent and UK-based ovarian cancer screening trial, the UKCTOCS was envisaged as one of the key evidence sources on both the effectiveness and costs of ovarian cancer screening.

Methods

Three systematic reviews were conducted. The searches for all three reviews were conducted between September and October 2014, with no date limit on the studies retrieved. Searches were

performed in Medline, Embase, CINAHL, Web of Science, Cochrane NHS EED, and Econlit. The three systematic reviews were designed to identify evidence on the following:

1) Ovarian cancer screening trials (to inform estimates of effectiveness),
2) Ovarian cancer economic evaluations (to inform the model structure, along with possible sources for utility, cost and resource use estimates), and
3) Ovarian cancer utility studies (to inform estimates of health-related quality of life relating to ovarian cancer screening, symptoms and treatment).

The systematic reviews were supplemented by evidence known by the study team, or by their clinical advisors. Costs and resource use were derived from a mixture of national sources (such as clinical guidelines), literature, data from the English Cancer Registries, and estimates provided by multidisciplinary teams responsible for the management and treatment of ovarian cancer.

Results from the systematic reviews, along with clinical guidelines and input from clinical advisors were used to develop conceptual models. These conceptual models covered the natural history of ovarian cancer, its treatment, and screening for ovarian cancer, and were used to inform the development of the health economic model. A cohort-level Markov model with the perspective of the NHS and Personal Social Services and a lifetime horizon was developed. Detailed evidence on the age and stage breakdowns of ovarian cancer incidence and mortality for the three trial arms (no screening, MMS and USS) was not available from UKCTOCS. This analysis relies on UKCTOCS results in the public domain to estimate both the incidence of ovarian cancer and mortality from ovarian cancer for all three screening arms. Costs were reported in 2013/14 pound sterling. The primary measure of effectiveness was the incremental cost per incremental quality-adjusted life years (QALYs) gained, summarised as the incremental cost-effectiveness ratio (ICER).

Results.

Results of systematic reviews

An existing systematic review and meta-analysis of screening trials was identified. This considered publications up to February 2012. There were no relevant trials identified beyond those included in the existing systematic review, although more recent publications from the UKCTOCS were available. There were concerns that non-English trials would not be generalisable to the English healthcare setting, primarily due to differences in the approach taken to treatment, which may impact on subsequent mortality. Further, it was felt that the UKCTOCS trial results were likely to supersede

those of previous UK-based screening trials, as the latter were forerunners of the UKCTOCS design with considerable overlap in the study teams, and due to evolutions in the use of blood-tests for screening. Hence the main effectiveness data were taken from the UKCTOCS.

Three existing economic evaluations of screening for ovarian cancer amongst general (or low-risk) populations were identified. One of these was considered to be of very poor methodological quality, and so was not considered further. The remaining two evaluations were used to inform the conceptual modelling and model structure. As these two economic evaluations both used the perspective of the US healthcare system it was deemed that their estimates of cost and resource use would not be applicable within the English healthcare setting. Ovarian cancer utility values were not estimated in either evaluation.

The systematic review of utility studies identified one relevant systematic review and two relevant studies. The systematic review focused on the treatment of ovarian cancer. There was a lack of evidence on if women's health related quality of life (HRQoL) varied with the treatment strategy received. However, there was a statistically significant increase in HRQoL after completing treatment, compared with at the start of treatment. The two identified studies provided some evidence to suggest that there is no impact of screening on HRQoL. This finding was supported by the existing meta-analysis of screening trials, which reached the same conclusion. One of the identified studies also elicited utility values for early and advanced ovarian cancer, suggesting that the latter was associated with a reduced HRQoL.

Health economic results

Both active screening strategies were associated with increased QALYs, but also increased costs. The estimated total lifetime average costs per woman for no screening, MMS and USS were £179 (95% confidence interval £137 to £242), £598 (£434 to £758), and £824 (£566 to £1,154) respectively, whilst the average QALYs accrued per woman were 14.290 (5.159 to 15.907), 14.357 (5.168 to 15.959) and 14.297 (5.147 to 15.926) respectively. The ICER comparing MMS with no screening was £8,864 per QALY (£2,600 to £51,576). Use of USS was dominated by MMS, being both more expensive and less effective. Under the base-case an estimated 3.19% of the simulated cohort of postmenopausal women would die from ovarian cancer under no screening. For MMS and USS this proportion was estimated to reduce to 1.41% and 2.35%, respectively, resulting in increased life expectancies of 1.7 (0.4 to 2.9) and 0.8 (-0.8 to 2.0) months, respectively.

The key uncertainties in the health economic results were the long-term estimates of the effect of screening on ovarian cancer mortality. Different assumptions about how to model this led to two-fold and four-fold increases in the ICER, to £18,372 and £36,769. Value of information analyses suggested that at a willingness to pay of £20,000 per QALY it was worth spending approximately £20 million to eliminate all of this long-term uncertainty. The other key uncertainty in the model inputs that drove the decision uncertainty relate to the HRQoL of women with ovarian cancer, along with the impact of treatment on this. Other uncertainties (in the cost of USS screening, the long-term effect of USS on ovarian cancer mortality, long term estimates of incidence, and the cost of diagnosis, treatment and palliative care) had a very small impact on decision uncertainty.

Discussion

The modelled results suggest that both screening strategies are likely to result in health benefits when compared to no screening, but at increased costs. Screening using MMS is estimated to be both more effective and cheaper than USS.

Strengths of this study were the detailed systematic reviews undertaken to identify input parameters and to inform the model structure and the application of advanced methodologies that allowed for alternative estimates of long-term effectiveness and the impact of these on decision uncertainty. It is also the first economic evaluation of screening for ovarian cancer amongst women who are not at high risk of ovarian cancer to use a UK-healthcare perspective, and the first to consider HRQoL as an outcome. A major limitation of this study was the lack of age and stage breakdowns for both the incidence of, and mortality from, ovarian cancer. This limited the analysis in that it was not possible to use the trial evidence to understand the natural history of ovarian cancer and thus to estimate the potential cost-effectiveness of alternative screening strategies (such as different screening intervals or different age-ranges), nor was it possible to evaluate the impact on cost-effectiveness of improvements in the screening test characteristics (such as sensitivity and specificity) or changes in compliance. The results are also limited by the uncertainty in a number of the model inputs, in particular the HRQoL of women with ovarian cancer, and the long-term effects of screening on ovarian cancer mortality. Further research into reducing these uncertainties would be beneficial.

In conclusion, results from the UKCTOCS demonstrated both that screening for ovarian cancer did not increase the incidence of ovarian cancer, and that screening was associated with a non-

statistically significant effect on ovarian cancer mortality, based on a median of eleven years of follow-up. Cost-effectiveness modelling of the lifetime health outcomes associated with ovarian cancer screening are promising, with an estimated ICER comparing MMS with no screening of £8,864 per QALY (95% confidence interval £2,600 to £51,576). However, there is substantial uncertainty in the long-term effectiveness of MMS in reducing ovarian cancer mortality, which is a key driver of cost-effectiveness. The other screening strategy considered by UKCTOCS was USS, which is estimated to be more expensive and less effective than MMS, and hence unlikely to be cost-effective.

1. Introduction

Ovarian cancer has the highest mortality of all gynaecological cancers in the UK and is the fourth highest cause of cancer deaths amongst women. Based on data reported by the English Cancer Registries[1,2], long-term survival rates are relatively low with five-year survival of 44%, and prognosis is worse for older women and those with a later stage of cancer at diagnosis. Stage one cancers have five-year survival rates of 90% following diagnosis, but less than a third of patients are diagnosed with this stage of cancer. At diagnosis, 60% of women have cancer which has already reached stage III or IV, with five-year survival rates of less than 20%.

Ovarian cancer can be diagnosed late because its symptoms (including abdominal pain, bloating, lack of appetite) are often shared with other, more common and benign disorders[3]. Screening could potentially reduce mortality rates, if it allowed earlier identification of cancers, ideally at less advanced stage.

The most commonly used screening approaches are blood tests for CA-125 (a protein biomarker for cancer) and trans-vaginal ultrasound imaging[4]. The ability of screening to drive health benefits depends on the sensitivity and the specificity of the tests. Good sensitivity will lead to better cancer detection and fewer women going undiagnosed. Good specificity will lead to fewer 'false positives' and fewer women having to undergo unnecessary tests and treatment.

A 2013 review of ten randomised control trials (RCTs) of ovarian cancer screening in asymptomatic women showed no reductions in mortality[5]. It found some evidence to suggest that screening did not lead to earlier stage at diagnosis, and that screening did not have an impact on mortality. However, the authors identified heterogeneity in these results. There was also evidence for some dis-benefits of unnecessary further interventions in women without cancer as a result of screening, although no impact on health-related quality of life was identified.

The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)[6] is a large (over 200,000 women) RCT running over ten years, which aims to develop definitive evidence on ovarian cancer screening. It tests two regimes: a) multimodal screening (MMS; CA-125 and ultrasound) and b) ultrasound screening (USS) with transvaginal ultrasound (TVS). For the purposes of this report TVS is used to refer to the general screening test, whilst USS is used to refer to the specific screening strategy employed in the UKCTOCS. The trial recruited post-menopausal women aged 50-74 with normal (low) risk of, and no previous history of, ovarian cancer. Results, based on a median of 11.1 years follow-up, showed that compared to no screening, MMS produced a 15% reduction in all-

cause mortality whilst USS produced a 11% reduction[7]. The authors also noted a potential 'late-effect' of screening, with the MMS mortality reduction being 8% in the first seven years and 23% in the subsequent seven years – for USS the reductions were 2% and 21% respectively.

The aim of this project is to conduct an economic evaluation of the cost-effectiveness of screening for ovarian cancer amongst postmenopausal women who are not at high risk of ovarian cancer. In order to achieve this objective, this project has two main objectives. The first objective is to conduct a detailed assessment of the available health economic evidence relating to screening for ovarian cancer. The second objective is the development of a mathematical model to estimate the healthcare costs and health-benefits of alternative screening options for ovarian cancer in England.

2. Epidemiology of ovarian cancer

2.1. Defining ovarian cancer.

There are a variety of different ways that ovarian cancer may be defined. One definition is by type of ovarian cancer, with approximately 90% of ovarian cancers being classified as epithelial and approximately 10% as either borderline or non-epithelial tumours[8]. Borderline tumours are slow growing, with low malignant potential. There are different types of non-epithelial ovarian cancer, such as germ cell cancer or sex-cord stromal tumours. These other types of cancer are all rare (occurring in less than 5% of cases of ovarian cancer). Within this study, both borderline and non-epithelial cancers are considered collectively, unless otherwise stated.

Epithelial ovarian cancers may be further sub-divided based on their morphological type. Data from the English cancer registries showed that, between the years of 2008 to 2010 (n = 14,827), 39% (5,749) of ovarian cancers had an unclassified or unspecified morphology. Of the remaining ovarian cancers the majority (64%; 5,773) were serous carcinomas. The proportions of the other morphologies were all similar: mucinous (11%; 992), endometriod (10%; 947), clear cell (8%; 726) and other classified epithelial tumours (7%; 640).

Cancers may also be defined by their stage at diagnosis. There are a variety of different grading systems that may be used[9]. The FIGO staging system is frequently used within England[6,10] and is also used for international comparisons[11]. There is also a strong association between age and stage at diagnosis, as shown in chapter 2.2. For this study, stage is used to sub-divide ovarian cancers. There are four FIGO stages, of increasing severity. These are defined as:

- FIGO stage I: the tumour is only present in the ovaries or fallopian tubes.
- FIGO stage II: the tumour is present in the ovaries or fallopian tubes, and has also grown into the pelvis or the peritoneum.
- FIGO stage III: the tumour has spread outside the area surrounded by the pelvis into the abdominal cavity. Tumours of nearby lymph nodes are also stage III tumours.
- FIGO stage IV: the tumour has spread to another organ (become metastatic), such as the liver, brain or lungs.

The FIGO staging system applies to both epithelial and non-epithelial ovarian cancers.

A noted limitation of the FIGO staging system is that it does not incorporate the grade of the tumour. This is an important limitation because it has been theorised that the natural history of ovarian cancer may vary by grade, with high grade cancers growing quicker than low grade cancers, irrespective of morphology[12]. However, a recent study showed that there was no difference in time to diagnosis between high grade and low grade cancers[13], although this study was limited in that it only considered the time since the onset of symptoms (and not since the onset of occult signs that may be screen-detectable). In addition, the reporting of grade data has historically been poor, with reporting levels amongst English cancer registries being consistently between 43% and 44% of all epithelial cancers between the years 2008 and 2012 (for borderline cancers the rate is between 4% and 7%)[14]. Moreover, it is noted that there is a high correlation between stage of ovarian cancer and grade. Data from the Eastern and South West regions of England for 2012 (for which 46% and 43% of records had complete data for both stage and grade, respectively) showed a statistically significant correlation (Pearson correlation coefficient: 0.4219, $p < 0.0001$)[14]. Because of this, differences by grade were not explicitly considered in this study, as it is anticipated that these will be covered by considering differences by stage. However, there remains uncertainty in this component of the natural history of ovarian cancer.

The International Classification of Diseases (ICD) provides a standardised set of codes that may be used to classify a disease[15]. Currently the tenth version of ICD (ICD-10) is in use. However, ICD-10 codes do not distinguish between epithelial and borderline ovarian cancers.

As this study is concerned with the potential impacts of implementing screening for ovarian cancer, the working definition of ovarian cancer adopted for this study is that it is any cancer that would be both identified as a result of screening and diagnosed as an ovarian cancer (as opposed to, say, a metastatic cancer from a different site). Based on discussions with clinical experts, this working definition was further refined to the following:

- ICD-10 codes C56 (ovarian cancer), C57 (fallopian tube cancer) and C48 (peritoneal cancer).
- All histological types and all grades.
- Borderline (and non-epithelial) cancers are to be treated as false-positive results (although an additional analysis that treats them as true positives shall also be conducted).

Screening for ovarian cancer may result in more borderline cancers being diagnosed. However, there is uncertainty about whether or not borderline ovarian cancers represent 'true' ovarian cancers. This is motivated by the observation that borderline tumours have a very favourable prognosis[16], over 95% of women newly diagnosed with borderline ovarian cancers will live for over 10 years[17]. Hence,

a diagnosis by screening of a borderline ovarian cancer may represent an over-diagnosis of a tumour of such low malignancy (or slow growth) that it would never become malignant within a patient's lifetime had they not been screened[18], and so a diagnosis of a borderline cancer may represent a false-positive finding[6]. However, there is some evidence to suggest that borderline cancers do have the potential to become malignant[19], so it may be appropriate to include them in the definition of ovarian cancers. For the primary outcomes of this study borderline cancers were excluded from the definition of ovarian cancers. The impact of this was assessed by including borderline cancers in the definition in a secondary analysis.

2.2. Incidence and prevalence.

Epithelial ovarian cancer is more common in economically developed countries[20], where it is the fourth most common cause of cancer mortality amongst females[21]. The risk of developing ovarian cancer is associated with the lifetime number of ovulatory cycles; pregnancy or oral contraception (which reduces the number of cycles) can lead to a protective effect from ovarian cancer[22]. Women can also have a genetic predisposition to ovarian cancer; this includes a history of BRCA mutation, family history of breast or ovarian cancer, and history of breast cancer[22].

The data presented below and in the subsequent section are based on data supplied by the English Cancer Registries[14]. This represents more detailed and more recent data than that presented in the publically available report *Overview of Ovarian Cancer in England: Incidence, Mortality and Survival* (http://www.ncin.org.uk/view?rid=1740)

In England, the incidence of epithelial ovarian cancer has remained relatively stable over time. Between 2008 and 2012 approximately 5,000 women (aged 15 and over) were diagnosed each year, resulting in a yearly incidence rate of approximately 23 diagnoses per 100,000 female population. The incidence of either borderline or non-epithelial cancers is much lower, with a combined incidence rate of approximately 3.5 diagnoses per 100,000 female population per year.

The incidence of ovarian cancer varies with both age and stage. However, in England not all incident ovarian cancers are staged. The proportion of un-staged cancers has been decreasing over time, for epithelial cancers this decrease has been from 65% in 2008 to 20% in 2013. For borderline cancers the decrease is from 72% to 15% respectively. The data used for this study are for the years 2008 to 2011 combined, as missing stage data for these years have been previously imputed, as described in Barclay *et al*[23]. It should be noted that only stage data for epithelial cancers were imputed. The available data for borderline (or non-epithelial) cancers suggest that the majority present as Stage I

cancers (91% of incident cancers staged in 2012). Hence a stage-breakdown of borderline cancers was not used for this study.

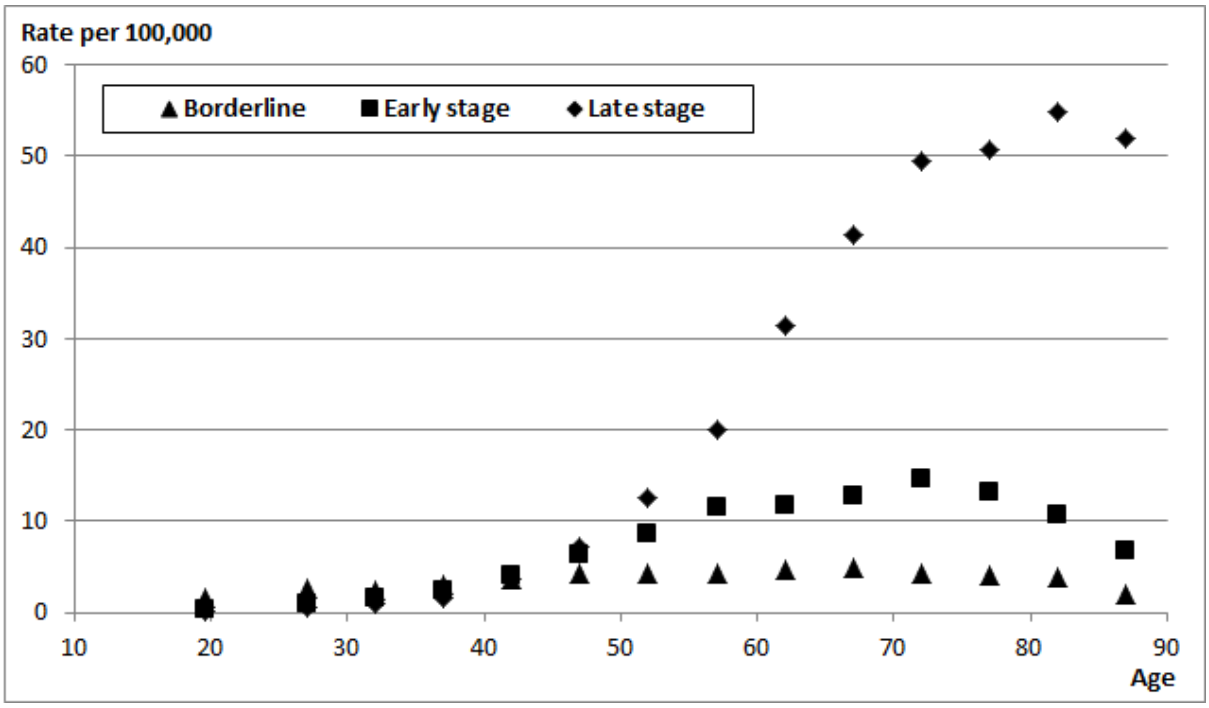
The incidence of ovarian cancer by age and stage for the combined years of 2008 to 2011 is presented in Table 2.1. There is little variation in the incidence of borderline (and non-epithelial) cancer by age. In contrast, epithelial ovarian cancers show a strong age-trend. Below the age of about 45 stage-specific rates of epithelial ovarian cancers are lower than that of borderline cancers. Above the age of about 45, there is an increase in the incidence of epithelial ovarian cancer and by the age of 55 the incidence rate of stage-specific epithelial ovarian cancers is (with the exception of Stage II) greater than that of borderline ovarian cancers. Stages of epithelial ovarian cancer may be loosely grouped into early-stage (Stages I and II) and late-stage (Stages III and IV), with a stronger age-trend in the latter age-group as shown in Figure 2.1. Above the age of 70, the incidence of late-stage cancers is more than ten times that of borderline cancers.

Table 2.1 Annual incidence of ovarian cancer (2008 to 2011), by stage and age.

Age Group	Borderline		Stage I		Stage II		Stage III		Stage IV	
	Count	Rate	Count	Rate	Count	Rate	Count	Rate	Count	Rate
15-24	169	1.7	36	0.4	5	0.0	13	0.1	6	0.1
25-29	137	2.5	41	0.8	9	0.2	23	0.4	8	0.1
30-34	123	2.4	67	1.3	11	0.2	25	0.5	19	0.4
35-39	169	3.0	105	1.9	33	0.6	62	1.1	32	0.6
40-44	219	3.7	192	3.2	50	0.8	153	2.6	61	1.0
45-49	246	4.3	276	4.9	80	1.4	294	5.2	112	2.0
50-54	212	4.3	334	6.8	93	1.9	414	8.5	200	4.1
55-59	196	4.3	380	8.3	143	3.1	619	13.6	291	6.4
60-64	220	4.6	399	8.4	154	3.3	988	20.8	505	10.7
65-69	179	4.9	333	9.1	135	3.7	993	27.2	521	14.2
70-74	139	4.3	331	10.3	138	4.3	973	30.4	609	19.0
75-79	110	4.0	253	9.2	112	4.1	789	28.7	606	22.0
80-84	83	3.8	162	7.4	70	3.2	609	27.9	589	26.9
85+	46	2.0	106	4.5	53	2.3	559	24.0	654	28.1

Borderline cancers include non-epithelial cancers. Stage-specific values are for epithelial cancers only. Rates are per 100,000 age specific female population.

Figure 2.1 Incidence of ovarian cancer (2008 to 2011), by broad stage and age.



Borderline cancers include non-epithelial cancers. Stage-specific values are for epithelial cancers only. Rates are per 100,000 female population.

The most recent prevalence estimates for ovarian cancer are for 2006. It was estimated that at the end of this year there were approximately 23,000 women in England who were alive and had a diagnosis of ovarian cancer in the past 20 years. Of these women, approximately 21,000 were diagnosed in the past 10 years and 13,000 in the past 5 years[24].

2.3. Survival and mortality

Mortality (and hence survival) varies by both age and stage. For example, mortality rates in the first year of diagnosis range from 11% for women under the age of 50 to 71% for women over the age of 85. By stage the rates vary from 5% (Stage I) to 63% (Stage IV).

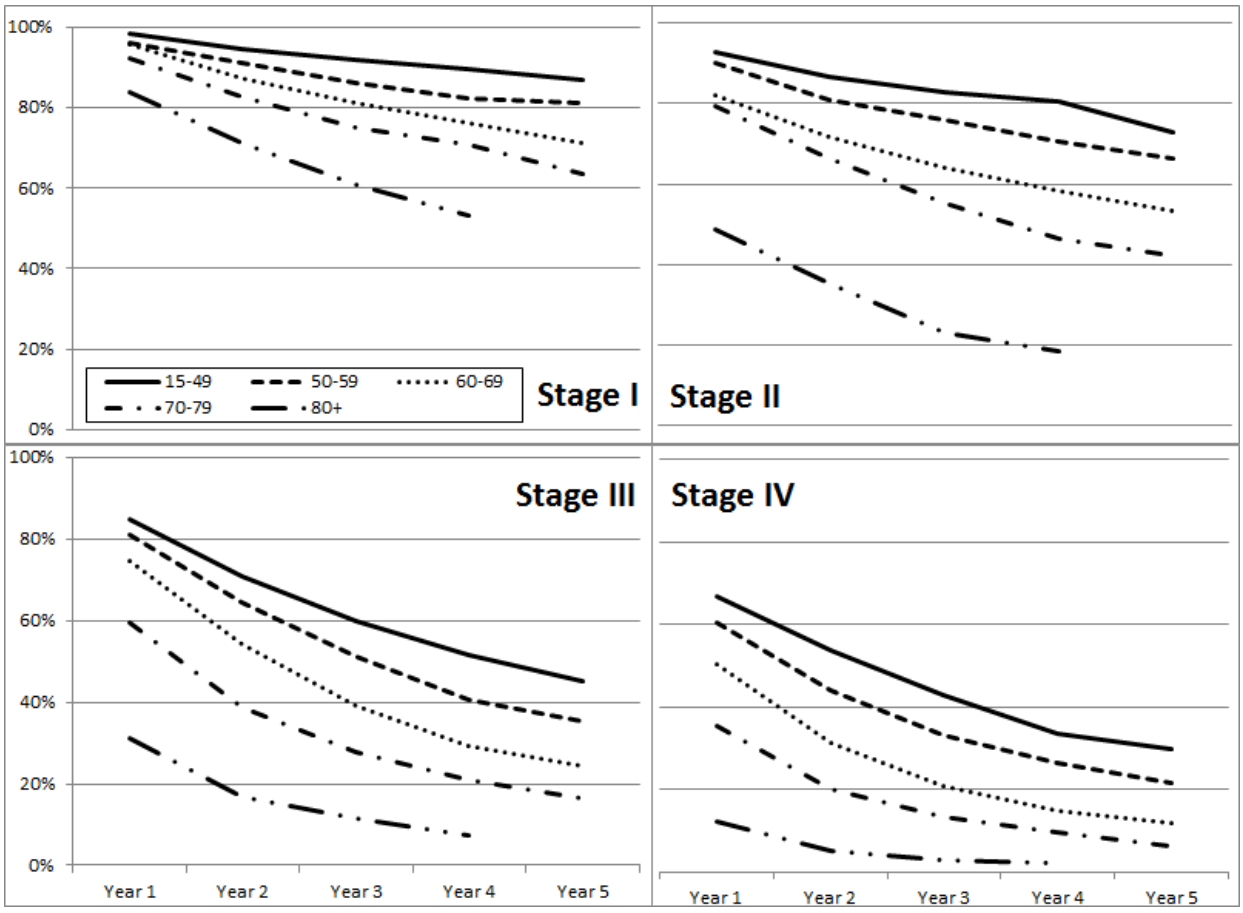
Of women diagnosed with borderline (or non-epithelial) ovarian cancer, 2.3% died within a year. This death rate is similar to that which would be observed in the general population[25], highlighting the relatively benign status of these tumours.

For this study data for the years 2008 to 2011 were used, as improved stage data were available for these years (as described in the previous section). Follow-up data are available until the end of 2013,

so survival data for up to five years is available (for individuals diagnosed in 2008). These are displayed in Figure 2.2, which highlights the variation by both age and stage. As an example, for the age-group '50 to 59', the percent of women who survive for five years following diagnosis is: 81% (Stage I), 66% (Stage II), 35% (Stage III) and 22% (Stage IV) .The figure also indicates that there is little evidence of any stage-by-age interaction with respect to survival. That is, for any given stage, the relative effects of age on survival are approximately constant over time.

Due to small numbers, an age-breakdown of survival for individuals diagnosed with borderline cancers is not presented. The overall percent surviving decreased from 97.7% after one year to 91.0% after 5 years.

Figure 2.2 **Percent surviving by age, stage, and length of follow-up: epithelial ovarian cancers diagnosed between 2008 and 2011.**



Due to small numbers, five-year follow-up for individuals aged 80 or over has been combined with that for individuals aged 70-79; this combined value is only displayed for the younger age-group.

2.4. Routes to diagnosis

In 2010 the National Cancer Intelligence Network (NCIN) analysed data on the routes by which patients with cancer were diagnosed, and their survival following diagnosis[26]. English patients diagnosed in 2007 were included in the analysis, which was performed to support the National Awareness and Early Diagnosis Initiative. Seven different possible routes to diagnosis were considered. For ovarian cancer, data on the percent diagnosed via each route and survival at one, two and three years are summarised in Table 2.2 (of note, these data were taken from the published workbook version 3.2, numbers differ from those given in the main data briefing).

Table 2.2 Routes to diagnosis for ovarian cancer (2007) and survival.

Route to diagnosis	Number in cohort	Percent presenting	Survival at: 1 year	2 years	3 years
Emergency presentation	8,699	31.1	43%	32%	25%
GP referral	7,135	25.5	80%	72%	66%
Two Week Wait	6,766	24.2	85%	73%	64%
Other outpatient	3,574	12.8	81%	70%	62%
Unknown	1,217	4.3	70%	61%	53%
Inpatient elective	606	2.2	80%	68%	61%
Screening	0	0	N/A	N/A	N/A

The Two Week Wait scheme is a type of GP referral, so collectively this route to diagnosis was the most common, accounting for half (49.7%) or all referrals. The next most common route to diagnosis was via emergency presentation with almost a third (31.1%) of all ovarian cancers diagnosed in this way. Women diagnosed following either GP referral or the Two Week Wait scheme contributed to half (49.7%) of all diagnoses.

Women diagnosed following emergency presentation had the lowest survival of all the routes considered. Women diagnosed via either GP referral or the Two Week Wait scheme were almost twice as likely to survive for one year, and more than 2.5 times more likely to survive for three years compared to women diagnosed following emergency presentation.

These data show that there is scope for improvement in the diagnosis of ovarian cancer, and that this could in-turn improve rates of survival. For example, if half of all emergency presentations were instead diagnosed via GP referral, then overall survival at one year would improve from 70% to 75%, whilst survival at three years would improve from 52% to 58%.

3. Screening for ovarian cancer

3.1. General screening issues.

3.1.1. The aim of ovarian cancer screening.

Screening for cancer can have one of two broad aims. Screening for prevention aims to identify precursors of cancer, so that preventative strategies may be employed to stop the cancer developing. Screening for detection aims to identify existing cancers earlier than they would otherwise be identified, with the expectation that earlier treatment will lead to improved survival.

Unlike other cancers such as colorectal and cervical, there is no established pre-cursor lesion for ovarian cancer. Some studies have posited that borderline cancers may represent pre-cursors[27]. However, the evidence base for this is weak and derived from mainly circumstantial evidence that borderline cancers may become malignant. In addition, a screening trial carried out during the 1980s showed that the removal of borderline cancers identified via ultrasound did not impact on mortality, strengthening the evidence base against the status of borderline tumours as pre-cursor lesions[28]. There is on-going research into the origins of ovarian cancer that may provide additional insight into the prospects of identifying pre-cursor lesions. The current state of art of this research has been discussed by a number of authors[29,30]. There is currently evidence to suggest that ovarian cancer consists of two important sub-sets. The first, known as Type I tumours, consists of low-grade serous cancers along with endometrioid, clear cell and mucinous cancers. Type II tumours consist of high-grade serous cancers and undifferentiated or poorly differentiated carcinomas. The two types differ in both the types of mutations witnessed and their prognosis, with Type II tumours more likely to be diagnosed at an advanced stage (and have poorer survival) and also more likely to harbour mutations in p53, BRCA1 and BRCA2. It is believed that the relatively benign Type I tumours may originate from borderline tumours (hence removal of these is unlikely to have a noticeable survival impact), whilst the more aggressive Type II tumours may originate from the fallopian tube; hence by the time they are identified in the ovaries they have already become metastatic.

To date, screening trials for ovarian cancer have sought to detect ovarian cancer at an earlier stage (screening for detection), with the rationale that because there is a strong stage-effect on mortality (as demonstrated in Section 2.3) then earlier diagnosis should lead to a mortality benefit. For example, using the data supplied by PHE (for the time period 2008 to 2011 inclusive, as reported in Section 2.3), if 50% of stage III & IV epithelial ovarian cancers had instead been diagnosed at stage I,

and there was no change in stage-specific survival, then 5-year survival for the entire cohort would increase from 33.1% to 51.5%.

As reported in the subsequent sections (3.2 and 3.3), to date four trials have reported on the effects of screening on stage-at diagnosis[7,31-33], of these three have also reported on the effects on mortality. These are the pilot trial by Jacobs *et al* carried out in the UK[31], the USA-based PLCO trial[33] and most recently UKCTOCS. The design of these trials along with the reported mortality effects are discussed in detail in section 3.3, whilst the effects of screening on stage at diagnosis are discussed in section 3.4.3. In brief, the PLCO trial reported that screening did not impact on either ovarian-cancer mortality or other-cause mortality, with relative risks of 1.18 and 1.01 respectively (neither value was statistically significantly different to 1, values below 1 indicate a beneficial effect of screening). In contrast, the study by Jacobs *et al* identified a statistically significant relative risk for ovarian-cancer mortality of 0.50 (the impact on other-cause mortality was not reported). Results from UKCTOCS, the most recent and largest of the screening trials, demonstrated a mortality reduction due to screening although this did not reach traditional levels of statistical significance. The UKCTOCS authors noted a potential late-effect of screening on mortality and suggested that further follow-up was required to assess the potential for a significant mortality reduction. A review by Reade *et* al[5] (conducted prior to the publication of the UKCTOCS mortality results) considered the differences in outcomes between the Jacobs trial and the PCLO, and noted that these different results were not explained by differences in the risk status or menopausal status of the participants, nor were they due to the screening modality employed or the length of follow-up.

3.2. Review of existing ovarian cancer screening trials.

To inform the choice of screening strategies to consider in the heath economic modelling, a systematic search of randomised controlled trials (RCTs) of ovarian cancer screening was conducted. The study authors were already aware of an existing systematic search (and review) of screening RCTs which searched for publications between 1st January 1979 and 5th February 2012[5]. This review was extended to consider publications up to 24th September 2014. Full details of the review methods are provided in Appendix 1. This extended review initially identified 2,233 unique publications. Of these, 2,104 were excluded from consideration based on the contents of their title and a further 121 were excluded based on their abstract. Of the remaining eight publications, only one was an RCT[34], however this was of women at elevated risk of ovarian cancer and so was excluded. Hence, no additional RCTs beyond those reported by Reade *et al*[5] were identified. However, whilst the Reade *et*

al study included UKCTOCS (as results from the prevalence screen had been published[6]), subsequent UKCTOCS publications (including mortality[7] and psychological morbidity[35]) were not available. Hence the available evidence relating to the types of ovarian cancer screening tests and their effectiveness are taken from both the Reade *et al* study and subsequent UKCTOCS publications.

Reade *et al*[5] identified ten RCTs of ovarian cancer screening amongst asymptomatic women[31,36-41]. Of these ten trials, nine recruited only women at low-risk of ovarian cancer, whilst one (the QUEST trial, a USA-based study) included both low-risk and high-risk women (whilst excluding women with a suspected BRCA mutation). The QUEST study is included in the following review, but it is noted that its relevance to the current work may be limited.

Key details for each of the trials are presented in Table 3.1. There are four key elements of the screening modalities presented: the screening tests used first line, the screening tests used second line (if applicable), the length of screening, and the time period between screens (if applicable). Differences in any one of these four elements may affect comparisons across different screening modalities. For example, both the SCSOCS and the PLCO trials used the same first line screening tests (TVS and CA-125) and the same time period between screens (one year). However, they differed with respect to both the second line tests used; one of TVS alone, TVS with CA-125 or referral to a gynaecologist in the SCSOCS trial, compared to no secondary screen in the PLCO trial and the length of screening (5 and 6 years for SCSOCS and PLCO respectively).

Table 3.1 Overview of RCTs evaluating screening methods for ovarian cancer (adapted from Reade *et al*).

Study	Setting and enrolment period.	First-line screen	Secondary screens	Subjects randomised
Parkes 1994[36]	UK, Sept 1989–Feb 1993	One-off TVS	Colour Doppler. Referral to gynaecologist if this is abnormal.	Screening: 3,562 Control: 3,562
Tabor 1994[37]	Denmark, Nov 1990	One-off TVS	Repeat TVS in 3 to 8 weeks. Referral to gynaecologist if this is abnormal.	Screening: 474 Control: 476
Jacobs 1999[31]	UK, 1989	Annual CA-125 for 3 years	TVS	Screening: 10,958 Control: 10,977
Taylor 2004[38]	US, May–Dec 1998	One-off TVS and CA-125	None	Screening: 215 Control: 217
ROCA trial[39]	UK, 1995–2000	One-off CA-125(ROCA)	CA-125(ROCA) with or without TVS depending on outcomes of first-line screen.	Screening: 6,682 Control: 6,790
Johnson 2006[40]	US, 1995–2000	Annual TVS and CA-125 for 4 years	None	Screening: 284 Control: 238
QUEST trial[41]	US, Dates not specified	Alternative CA-125 and TVS every 6 months for 18 months, with or without risk counselling (RC).	TVS	Screening + RC: 152 Screening, no RC: 140 RC alone: 150 Usual care: 150
SCSOCS trial[32]	Japan, Sept 1985–1999	Annual TVS and CA-125 for 5 years	*Either* repeat TVS every 3 to 6 months *or* repeat TVS with CA-125 at 6 months *or* referral to gynaecologist depending on outcomes of first-line screen.	Screening: 41,688 Control: 40,799
UKCTOCS trial – multimodal arm[6,7]	UK, 2001–Oct 2005	Annual CA-125(ROCA) for up to 11 years.	TVS	Screening: 50,640 Control: 101,359
UKCTOCS trial – ultrasound arm[6,7]	UK, 2001–Oct 2005	Annual TVS for up to 11 years.	TVS repeated by more experienced sonographer.	Screening: 50,639 Control: 101,359
PLCO trial[33,42]	US, Nov 1993–Jul 2001	Annual TVS and CA-125 for up to 6 years	None	Screening: 39,105 Control: 39,111

The control arm was usual care for all of the studies. Secondary screens apply if the first-line screen is abnormal.

TVS: transvaginal ultrasound, ROCA: risk of ovarian cancer algorithm.

3.3. Impact of ovarian cancer screening on mortality.

To date, three trials have reported on the impact of ovarian cancer screening on mortality. As this is a key outcome measure, these trials shall be discussed in detail.

3.3.1. The Prostate, Lung, Colorectal and Ovarian Cancer Screening trial[33].

The US-based Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening RCT enrolled participants between 1993 and 2001. As the name suggests, the trial had four separate disease-specific screening arms. The ovarian arm enrolled 78216 women aged between 55 to 74 years (inclusive) with no previous history of ovarian, lung or colorectal cancer. Being post-menopausal was not an inclusion criteria (although given the age of the subjects it is likely that the majority were). Women were randomised 1:1 to receive either annual screening with both TVS and CA-125 or no screening (usual care). A positive (abnormal) screen was defined as a positive result from either the TVS or the CA-125 screen. A cut-point of 35 U/mL was used to define a positive CA-125 result. Women received up to 6 years of screening, although screening with TVS was only available for up to 4 years. Women were followed-up for up to 13 years (median follow-up 12.4 years). Of the 39,105 women randomised to receive screening, 5,388 received a positive screen after four rounds of screening, and 212 cancers were identified[33,42]. Amongst the 39,111 women not screened 176 cancers were identified. There was no difference in the proportion of early-stage cancers identified (22% amongst screened women, 21% amongst not-screened women), although screening detected more cancers at Stage III (57%) and less at Stage IV (20%) compared to not screening (47% and 31% respectively). After follow-up, screening was found to not have any impact on ovarian-cancer mortality, with 118 deaths amongst the screened population and 100 amongst the usual care population, giving a relative risk of 1.18 (95% CI: 0.82 to 1.71 – values greater than one suggest that screening leads to increased mortality). Screening did not have an impact on other cause mortality either (which excluded ovarian, colorectal or lung cancer deaths), with 2,924 deaths amongst the screened population and 2,914 amongst the usual care population, giving a relative risk of 1.01 (95% CI: 0.96 to 1.06).

The mortality results reported from the PLCO have been criticised[4,6]. The trial had a large healthy volunteer effect[43] with mortality rates amongst women 62% lower than expected. This necessitated the long follow-up which in turn diluted the screening effect, with over 40% of cancers amongst the screening arm population being identified after screening had finished. Another limitation of the trial was the lack of a central protocol for the management of positive results. Finally, it has been

speculated that CA-125 screening without using the ROCA may not be as efficient as using the ROCA[44]. Some of these limitations have been addressed, as PLCO results show that screening did not have a mortality impact when follow-up was restricted to two years after the end of screening[45] and a post-hoc analysis suggested that any improvements due to incorporating the ROCA would not be sufficient to demonstrate a mortality impact due to screening[46]. However, the generalisability of the results to the English setting remains unclear.

3.3.2. The Jacobs et al trial[31].

Jacobs *et al* (1999) report the results of a UK-based RCT screening post-menopausal women aged 45 years or older. The first-line screen consisted of CA-125, with a second-line screen of pelvic ultrasound if the CA-125 result was 30 U/mL or greater. A total of 22,000 women were invited to be randomised 1:1 to either annual screening for three years or usual care. Follow-up was seven years, at the end of which 16 cancers were diagnosed amongst the screened population and 20 amongst the non-screened population. The study was not powered to detect a significant mortality difference, the observed relative risk for ovarian cancer mortality was 0.50 (95% confidence interval 0.19 to 1.28, p = 0.083) based on nine deaths amongst the screened population and 18 amongst the non-screened population (it should be noted that the definition of relative risk used by Jacobs *et al* was the inverse of that used in the PLCO trial). A significant effect of screening on survival since randomisation amongst patients developing ovarian cancer was identified. Median survival was 72.9 months amongst the screened ovarian cancer cases compared to 41.8 months amongst non-screened ovarian cancer cases (p = 0.0112), although this finding was treated with caution. There was some evidence of a stage-shift due to screening, with the proportion of diagnosed Stage I or II cancers being 31.3% for the screened population and 10.0% for the usual care population. However, this comparison did not reach statistical significance (p = 0.171).

In addition to being under-powered, the study by Jacobs *et al* may share some of the limitations of the PLCO trial. The effect of screening was diluted by long follow-up, with over 60% of all cancers amongst the screened population (10/16) diagnosed after the end of screening. As with the PLCO, the results of the CA-125 screen were also interpreted using a fixed cut-point, instead of a longitudinal algorithm such as the ROCA. There have been no published details regarding any potential healthy volunteer effect.

Reade *et al* [5] pooled results from both the PLCO and Jacobs *et al* trials to estimate a combined relative risk for the impact of screening on ovarian cancer mortality of 1.08 (95% CI 0.84 to 1.38), suggesting that screening does not have any impact on mortality. However, due to the previously mentioned limitations, this should not be taken as definitive evidence about the impact of screening on ovarian cancer mortality.

3.3.3. The UK Collaborative Trial of Ovarian Cancer Screening trial[6].

The UKCTOCS study (ISRCTN22488978) is an RCT that, between April 2001 and October 2005, randomised post-menopausal women aged 50 to 74 to one of three treatment arms: (1) annual multimodal screening with first-line CA-125 screening (interpreted using ROCA) TVS as the second-line test (MMS group, n = 50,640), (2) annual screening with TVS as both the first and second-line test (USS group, n = 50,639), or (3) no screening (n = 101,359). Exclusion criteria were: history of bilateral oophorectomy, active malignancy, previous history of ovarian cancer, participation in other ovarian cancer screening trials, or increased risk of familial ovarian cancer.

Women receiving a positive screen (indicated by a positive result for both first line and second line tests) underwent diagnostic investigations. These investigations included repeat TVS, measurement of CA-125 (if this had not been performed during screening) and computerised tomography or magnetic resonance imaging scans. The investigations were performed to confirm or refute the positive screen result, and hence decide if the woman required surgery. When surgery was carried out, the recommended approach was to remove both the ovaries (bilateral oophorectomy) along with the fallopian tube (salpingectomy), in a combined procedure known as bilateral salpingo-oophorectomy (BSO). In some cases a hysterectomy (removal of the womb and cervix) were also performed. The type of surgery performed depended on the degree to which ovarian cancer was suspected. A low suspicion led to laparoscopy with the intention to perform a laparoscopic BSO. A laparotomy was undertaken if either the diagnostic investigations or the initial laparoscopy led to a strong suspicion of ovarian cancer.

Women participating in the trial received annual screens up until 31st December 2011, and were followed up until 31st December 2014. Hence, women received up to 11 screens and were followed-up for between 13 years 8 months and 9 years 2 months, depending on when they were enrolled. The primary outcome measure was a comparison of ovarian cancer mortality between screened and not screened populations.

Results based on a median follow-up of 11.1 years (inter-quartile range 10.0 to 12.0 years) showed that ovarian cancer mortality was 0.34%, 0.29% and 0.30% in the no screening, MMS and USS groups respectively[7]. The main analysis estimated mortality reductions relative to no screening of 15% (95% CI –3% to 30%; p=0·10) with MMS and 11% (95% CI –7% to 27%; p=0·21) with USS. Whilst these mortality reductions did not reach traditional levels of statistical significance (taken at p = 0.05), a pre-specified analysis which excluded prevalent cases of ovarian cancer (defined based on the concentration profile of CA-125) estimated a statistically significant mortality reduction of 20% with MMS (95% CI –2%% to 40%; p=0.021). Further analyses suggested a potential late-effect of screening, with mortality reductions of 8% in the first seven years for MMS (2% for USS) and 23% in years 7 to 14 (21% for USS). When excluding prevalent cases the reductions were 8% in the first seven years for MMS and 28% in years 7 to 14.

As the UKCTOCS evaluated different screening regimes to the earlier two screening studies (which did not include the ROCA), results are not directly comparable.

3.4. Performance of ovarian cancer screening modalities.

3.4.1. Number of surgeries required to detect one cancer and false positive rate.

Data on the performance of the screening modalities were not presented for all of the trials. Further, when data were presented there was a lack of uniformity with regards to which findings represented false positives. In particular, there is uncertainty over whether or not borderline ovarian cancers detected by screening would ever be diagnosed in the absence of screening and hence whether or not these findings represent true or false positives[6,18]. There is also inconsistency over the reporting of other neoplasms, such as peritoneal cancers or ovarian neoplasms of uncertain behaviour[6].

A definitive diagnosis of ovarian cancer following a positive screen result requires surgery. Because of this, the number of surgeries required to detect one cancer is directly linked to the false positive rate. For a screening test, if p denotes the number of positive results, and c the number of these with cancer, the number of surgeries required to detect one cancer is p/c, whilst the false positive rate is $1 - c/p = 1 - (1/\text{ number of surgeries required})$.

Reade *et al* [5] presented a meta-analysis of number of surgeries required to detect one cancer for ovarian cancer screening. The definition of ovarian cancer used was invasive ovarian, peritoneal or fallopian tube carcinoma excluding borderline ovarian cancers. Data were available for all of the

trials listed in Table 3.1, with the exception of Taylor 2004[38] and Johnson 2006[40]. However, these reported results from a sub-population of the PLCO trial[33], so their exclusion is appropriate. The one trial that included high-risk women[41] was also included in this meta-analysis. However, this only screened 292 women (of which there were seven surgeries and no cancers), so its inclusion is unlikely to have a noticeable impact on the results. Pooling across all of the trials, the number of surgeries required to detect one cancer was 9 (95% CI 5.5 to 17.0). Sub-group analyses by the first-line screen showed variation in the number of surgeries required; with a value of 38 for TVS screening (95% CI 15.7 to 178.1), 13 for screening with both TVS and CA-125 (95% CI 15.7 to 178.1), and 4 for CA-125 screening (95% CI 15.7 to 178.1). These correspond to false-positive rates for the three screening modalities of 97% (95% CI 93.6% to 99.4%), 92% (95% CI 90.7% to 93.4%), and 75% (95% CI 63.0% to 79.6%) respectively (confidence intervals were derived by applying the formula '1 – (1/ number of surgeries required)' to the confidence intervals previously quoted). The CA-125 screening group may be further sub-divided as it contains one study that interpreted CA-125 results using a fixed score (Jacobs *et al* [31]), for which 4.8 surgeries were required to detect one cancer and two studies that interpreted CA-125 results using the ROCA; the ROCA study[39] and the multimodal screening arm of the UKCTOCS[6]. The numbers of surgeries required in these studies were 4.3 and 3.1 respectively.

The UKCTOCS data included in the meta-analysis are from the prevalence screen. Data from all screening rounds (between seven and eleven for women depending on when they were enrolled) have subsequently been published[7]. For the MMS arm 3.30 surgeries (95% confidence interval 2.96 to 3.70) were performed for every ovarian cancer identified (212 cancers out of 700 surgeries). For the USS arm 10.96 (95% confidence interval 9.49 to 12.70) surgeries were required (164 cancers out of 1798 surgeries), with corresponding false-positive rates of 70% and 91%.

3.4.2. Sensitivity and specificity

Sensitivity and specificity are both key performance characteristics of a screening test. However, these require knowledge of the true disease status for all subjects, and so their estimation can be difficult, and has only been attempted for a handful of screening studies. As with other performance measures, estimates also depend on whether or not borderline ovarian cancers are defined as true positive results or false positives. Sensitivity and specificity estimates treating borderline cancers as false positives are available for the ROCA, PLCO and UKCTOCS studies. The lowest reported sensitivity was 68.2% for the PLCO trial (specificity 98.4%). Sensitivities of 75.0% and 89.5% (with

respective specificities of 98.2% and 99.9%) were estimated for prevalent ultrasound screening and prevalent multimodal screening within the UKCTOCS trial, respectively. UKCTOCS estimates for all screening rounds were: MMS (sensitivity 84.0%, specificity 99.03%) and USS (sensitivity 72.9%, specificity 96.75%)[7]. A sensitivity estimate was not explicitly reported for the ROCA trial, but it was noted that there were no interval cancers one year after screening, implying a sensitivity of 100%. Specificity amongst followed-up individuals was 99.7%.

Sensitivity estimates which treat borderline cancers as true positives are available for the SCSOCS, PLCO and UKCTOCS studies. Estimated values were 69.5% (PLCO), 77.1% (SCSOCS), 84.9% (UKCTOCS prevalent ultrasound) and 89.4% (UKCTOCS prevalent multimodal). Corresponding specificities were only estimated for the UKCTOCS trial, and were 98.2% and 99.8% for ultrasound screening and multimodal screening, respectively. Estimates for all screening rounds of the UKCTOCS trial were: MMS (sensitivity 77.1%, specificity 98.98%) and USS (sensitivity 68.5%, specificity 96.73%).

3.4.3. Earlier diagnosis.

Within the systematic review of Reade *et al*, earlier diagnosis was defined as a shift in the stage at diagnosis from advanced stage (FIGO stages III and IV) to early stage (FIGO stages I and II). Data on stage at diagnosis was presented in three different studies as well as in a subsequent UKCTOCS publication. These four studies are discussed in turn.

The UK-based trial of Jacobs *et al* reported that 31% (n = 5) of cancers amongst the screened arm were diagnosed at an early stage, compared to 10% (n = 2) of cancers amongst the control arm. There was no statistically significant stage-shift identified by the study authors (p = 0.171), although the authors noted that the study was under-powered to detect a survival difference between the two trial arms (hence it is unlikely to be powered to detect a difference in stage distributions, as only 36 cancers were diagnosed). In addition, the screening effect may have been diluted due to the length of follow-up employed in the trial (seven years). If the analysis is restricted to only cancers identified by screening (n = 6), then 50% of cancers were identified at an early stage.

The Japanese-based SCSOCS trial identified 27 cancers amongst the screened-arm, of which 67% were at an early-stage. In comparison, 32 cancers were identified amongst the control-arm, of which 44% were at an early-stage. As with the Jacobs *et al* study, the authors tested the overall distribution of FIGO stages for any difference, the authors also found no statistically significant difference. However, the SCSOCS trial is likely to suffer from the same limitations as the Jacobs *et al* study, with

a mean follow-up of 9.2 years, and a likely lack of statistical power to detect a stage-shift in diagnosis.

The USA-based PLCO trial reported the stage-of cancers at diagnosis for two separate time-periods: during screening (the first five years of the trial) and during follow-up (years six to twelve of the trial). During the first five years, 24% (n = 30) of cancers were diagnosed at an early stage amongst the screened group, compared to 26% (n = 26) amongst the control arm. During follow-up the values are 20% (n = 17) for the screened arm and 15% (n = 12) for the control arm. No statistical tests of significance were performed, however given the relatively small number of cancers diagnosed and the similarity of the proportions, it is unlikely to reach statistical significance. The separation of results by screening-period and follow-up shows that for this trial the length of follow-up (median 12.4 years) did not dilute the stage-results. However, as with the other two trials, the PLCO is unlikely to be powered to detect a statistically significant difference in stage at diagnosis.

The UK-based UKCTOCS trial gives the stage-breakdown of ovarian cancers by trial arm[7]. The authors included Stage IIIa in their definition of earlier stage (referred to as 'low volume disease') with a statistically significantly higher proportion identified by MMS (40%) than by no screening (26%; $p<0.001$). There was no change in proportions for USS (24%; $p=0.57$). It is unclear if this classification of low volume disease was prospective or retrospective, although similar proportions were identified at Stage I or II (MMS: 38%, no screening 24%, USS 23%). However, a stage-breakdown of screen-detected cancers (as opposed to cancers amongst the screening arms) was not provided. It is unclear what impact this would have on earlier diagnosis; 41% and 49% of cancers were not screen detected amongst the MMS and USS groups.

The meta-analysis of Reade *et al* failed to identify a statistically significant effect of screening on diagnosis at an earlier stage, with a pooled relative risk of diagnosis at advanced stage of 0.86 (95% CI 0.68 to 1.11). However, the authors noted that the results were both inconsistent and imprecise. These inconsistencies (variation) amongst the trial results are highlight in Table 3.2, which summarises the proportions diagnosed at an early stage for the three trials. Data from the two UKCTOCS screening arms are also provided. Reade *et al* noted that this variation was not explained by differences in the risk status or menopausal status of the participants, nor was it due to the screening modality employed or the length of follow-up. The imprecision of the estimate may be due to a lack of statistical power.

Table 3.2 Percent of cancers identified at an early stage (n). Early stage defined as FIGO stages I or II.

	Jacobs *et al*		SCSOCS trial		PLCO trial		UKCTOCS: MMS		UKCTOCS: USS	
Screen (all)	31%	(5)	67%	(18)	22%	(47)	38%	(107)	23%	(58)
Screen-detected	50%	(3)	N/A	N/A	24%	(30)	N/A	N/A	N/A	N/A
Control	10%	(2)	44%	(14)	21%	(38)	24%	(136)	*Same as MMS*	

Screen (all): all cancers amongst the screening-group, including follow-up. Screen-detected: cancers identified amongst the screening-group during screening.

3.5. Risks of ovarian cancer screening.

Reade *et al* [5] considered three different types of screening-related harm: surgical complications for women without ovarian cancer who received a (false) positive screen result; distress due to worry about the risk of ovarian cancer; and the impact of screening on health-related quality of life (HRQoL).

3.5.1. Surgical complications following a false positive result.

A pooled analysis estimated that the risk of a severe complication during surgery for false positive results was 6.0% (95% CI 1.0% to 11.0%). This was based on data from the ROCA, PLCO and UKCTOCS studies [6,33,39]. However, the authors noted substantial variation in the reported trial results that was not explained by differences in the risk status or menopausal status of the participants, nor was it due to the screening modality employed or the length of follow-up. This unexplained variation may be due to international differences in healthcare settings[44]. For example, treatment in the USA may be more intensive, as treatment costs can be re-imbursed by private medical insurance. Because of this, it may be more appropriate to consider UK-based settings. Within the ROCA trial, one woman (out of 11 women with false positive results) developed a post-operative bowel obstruction. Results from the UKCTOCS trial reported that 3.1% of 488 women in the MMS group and 3.5% of 1,634 women in the USS group experienced a complications related to screen-positive surgery[7]. There were no deaths, with the most common complications being 'injury to hollow viscus' (n = 14) and 'haemorrhage' (n = 13). In addition, the number of serious complications following surgery is reported by Jacobs *et al* (1999)[31], with no complications amongst the 23 women with a false-positive result; this evidence appears to have been omitted from the Reade *et al* [5] analysis. Pooling the evidence from these three UK-based trials results in a risk of severe complications during surgery of 3.4%.

3.5.2. Distress due to a perceived risk of ovarian cancer.

Of the RCTs considered for this study, only Taylor 2004 (reporting initial results for a single site of the PLCO trial)[38] provided information on this, along with the QUEST trial[41], which includes high-risk women and so is not considered relevant to this study but was included in the meta analyses of Reade *et al* [5]. Evidence from both the individual studies were consistent in suggesting that there was no difference in cancer-specific distress, as measured by the Impact of Event Scale, between screened and not-screened groups.

3.5.3. Impact on health-related quality of life.

HRQoL was measured in both of the PLCO sub-sites[38,40], along with the excluded QUEST trial[41]. The George-town PLCO sub-site measured HRQoL using the SF-12, the other two studies used the SF-36. All three studies were consistent in showing that there was no impact of screening on either the physical or mental domains of HRQoL.

Because HRQoL parameters are a key input to the health economic model, evidence on the impact of screening on HRQoL was also taken from non-RCTs. These studies were identified as part of a broader systematic search for studies on HRQoL amongst women with ovarian cancer, as described in Section Five.

3.5.4. Surgical complications following a true positive result.

In addition to the screening-related risks reported by Reade *et al*, rates of surgical complications for women with ovarian cancer are also of interest for this study. Only Jacobs *et al* (1999)[31] and the PLCO trial[33] report rates for these. For the former trial there were no complications amongst the 6 women with a true-positive screen result. Results from the PLCO trial showed that 45% (95/212) of women whose cancer was diagnosed by screening and 52% (91/176) of women who developed cancer in the usual care arm experienced a major complication. There is a substantial difference between the rate reported in the UK-based trial and the two rates observed in the US-based trial. Whilst the rate reported by Jacobs *et al* (1999) is only based on six cancer cases (and so will be highly uncertain), it may be that international comparisons of surgical complications are not appropriate, as was suggested when considering rates of surgical complications following a false positive result.

3.6. Rates of screening participation.

The uptake of, and compliance with, cancer screening can be critical determinants of both effectiveness and cost-effectiveness[47,48]. For this study, the uptake rate of a screening test is defined as the proportion of invited women who agree to being screened, whilst the compliance rate is defined as the proportion of women who receive screening, amongst those women who are eligible to be screened. Hence, for any screening test there is a single uptake rate, but separate compliance rates for each round of screening. This study is concerned with the cost-effectiveness of ovarian cancer screening if it is implemented as a national screening programme. Hence rates of uptake and compliance should be based on rates observed amongst the general population. However, not all of the screening trials recruited women from the general population. For example, Parkes *et al* (1994)[36] recruited from a breast cancer screening clinic, whilst Jacobs *et al* (1999)[31] sent invitations to women who had participated in a previous ovarian cancer screening study. Only Tabor *et al* (1994)[37] and the UKCTOCS trial[6] recruited women from the general population (recruitment information is not provided for the PLCO trial). Tabor *et al* (1994) reported an uptake rate of 64.3% (950/1,477). Of the 474 women randomised to receive screening, 8% (38) did not attend. Detailed information on recruitment, uptake and compliance have been reported for the UKCTOCS[6,49]. Estimated values can vary depending on the definitions used. For the values reported here, any women who would be ineligible to receive screening are excluded from both numerator and denominator. The uptake rate was 24.8% (288,955/1,165,057, removing those who were ineligible). Of those who accepted an invitation for screening, a further 26.7% (73,965/276,603) withdrew or refused to participate prior to randomisation. Following randomisation, a proportion of women were not screened due to changing their mind. The rates of these were 2.9% for the ultrasound arm (1,490/50,639) and 1.0% for the multimodal screening arm (483/50,640). Data on compliance rates by round of screening are also available (web table 4)[7], with average compliance across all screens of 80.8% for MMS and 78.0% for USS. Compliance rates showed a monotonic decrease with each screening round (with approximately 4% attrition per round), falling from 98.4% for the first MMS screen (94.9% USS) to 47.2% for the 11th MMS screen (35.9% USS).

Uptake rates amongst a specialised population (such as women from a breast cancer screening clinic) are unlikely to generalise to the general population. However, compliance rates may as both are conditional on women having accepted the screening test. Both the SCSOCS and the PLCO trial reported compliance rates across multiple screening rounds. For the SCSOCS trial, compliance with screening was 82%, 71%, 67%, and 56% at the second, third, fourth, and fifth screen, respectively. For the PLCO trial, compliance was reported for the first four rounds[42]. Compliance with screening

varied with both round and type of screening. Compliance with CA 125 was greater than compliance with TVU for all rounds. Compliance rates also decreased with each screening round. Compliance rates for CA-125 and TVA decreased from 83.9% and 83.1% during the first screen (respectively) to 79.0% and 77.7% during the fourth screen (respectively). Compliance with both screening tests decreased from 83.1% in the first round to 77.6% in the fourth round.

3.7. Costs associated with ovarian cancer screening.

Of the trials identified by Reade *et al* [5], only the feasibility study of Parkes *et al* [36] reported any evidence on costs. In this study, the cost of a transvaginal screen was stated to be £20. Neither the source for this cost nor the year were reported. Alternative estimates of the costs of screening, derived from alternative evidence sources including national reference costs and guidance, are described in section six.

3.8. Summary of the available evidence on ovarian cancer screening.

There is limited evidence concerning the impact of screening on ovarian cancer mortality, as it has only been reported in three studies. All three studies considered screening with CA-125 and TVS, although the actual implementation of these varied. The two earliest studies interpreted CA-125 using a fixed cut-off. Of these, one found a positive impact of screening on mortality, but was under-powered to do so. The other trial was sufficiently powered, but failed to find any effect. Pooling the two results also failed to find any effect of screening on mortality. The more recent UKCTOCS trial interpreted CA-125 using the ROCA. It found that screening led to a mortality reduction, albeit a non-statistically significant reduction. Evidence of a possible delayed (or late) effect of screening on mortality was noted, with further follow-up required to further clarify the impact of screening on mortality.

Of the two screening strategies, those involving CA-125 have better test characteristics than those using TVA alone; with higher rates of sensitivity, specificity and lower rates of false-positive surgery. Different healthcare systems appeared to have different rates of surgical complications; UK-based evidence suggests that complication rates are about 3.4%. Trial-based evidence about the impact of screening on distress or health-related quality of life is limited, but consistent in indicating no effect. The evidence on whether screening leads to earlier diagnosis is inconsistent, and varies with screening regimen.

4. Review of the cost-effectiveness of population screening for ovarian cancer

The objective of this section is to present a systematic review of the evidence on the cost-effectiveness of population screening for ovarian cancer. The methods and results of searching and sifting are summarised and the included economic evaluations reviewed and compared. The included economic evaluations are discussed in terms of their quality, and their usefulness for informing UK decision making regarding NHS resources.

4.1. Methods and Search Results

The databases searched were Medline, Embase, CINAHL, Web of Science, Cochrane NHS EED, and Econlit. The searches were designed to identify both health economic evidence (for this review) and quality of life studies (for the review described in Section 5). The search terms used are described in Appendix 1. The searches were performed during the end of September and the start of October 2014, with no date limit on the studies retrieved. The search for economic evaluations was combined with a search for evidence of quality of life outcomes in ovarian cancer. The combined search resulted in 2,088 published studies being found. During the initial sift of papers, an existing review of economic evaluations by Suzcs, Wyss and Dedes was identified (see below for more details)[50]. As such, this review was subsequently modified to only consider studies published since the year 2000.

After the sifting process to exclude papers which were not economic evaluations and/or ovarian cancer screening papers based on titles and abstracts, all but five unique economic evaluations were excluded from further consideration. Of these five, after a sift of the full text, one economic evaluations published in 2013 was excluded as it was concerned with BRCA mutation testing in a high risk population.[51] In addition, the economic evaluations by Ding et al 2010[52] was excluded, as it was only available as a conference abstract. This left three unique economic evaluations that were identified for inclusion and full review.[53-55] Details on these, along with an overview and critique of the three unique economic evaluations are provided in the next section. The consolidated health economic evaluation reporting standards (CHEERS) checklists were completed for each of the evaluations, and are provided in Appendix 5.

4.2. Review of Published Evidence

The three papers selected for inclusion in this review are compared in tabular form in Table 4.1 followed by a more detailed review of the 3 papers individually. The two evaluations by Havrilesky [53] and Drescher[54] are modelling studies based in the USA. The third purports to be an economic evaluation for screening of fertile women in Thailand.

Table 4.1: Overview of post 2000 Cost-effectiveness of Ovarian Cancer Screening Analyses

	Havrilesky[53] 2008	Drescher[54] 2012	Wiwanitkit[55] 2013
Setting and population	USA hypothetical population at average and high risk of ovarian cancer. Women aged 20-death.	Hypothetical population of 1,000,000 USA post-menopausal women age 45-85 at average risk of epithelial ovarian cancer.	Fertile Thai women.
Screening Interventions	Hypothetical test with sensitivity / specificity and screening intervals (3-36mths) for women aged 50 to 85 years. Base-case values from recent biomarker studies. Base-case sensitivity = 85%; specificity = 95%.	Annual 2-step screening using ROCA[56] for CA-125 and TVS follow-up compared with hypothetical new tests based on imaging and biomarkers, CA-125 and TVS.	Comparison of CA-125 versus CA-125 and TVS using the risk of malignancy index algorithm.
Type of Analysis and approach	Cost-effectiveness using Markov model disease progression and hypothetical parameter values for modelled variables for women aged 20 years and older.	Existing lifetime micro-simulation model[57] clinical progression and cost variables.	Cost-effectiveness analysis. Validity of ICERs cannot be determined from data presented. Lack of clarity.
Key clinical model inputs and sources	Age specific Incidence and staging based on Surveillance, Epidemiology, and End Results database. Imputed values used to estimate Markov transition probabilities. Hysterectomy and oophorectomy rates from literature.	Appears to be selectively identified data from literature (trials), experimental analysis, and US databases supplemented by author opinion.	CA-125 test sensitivity = 70% Risk of malignancy index algorithm sensitivity = 85.4% Other clinical assumptions not stated but selective sources presented.
Costing and Perspective	2007 US$ Lifetime costs with healthcare sector perspective. Discount rate 3% per year. Direct costs of screening and treatment. Source: Medicare and USA guidelines. Screening costs assumed by authors. Resource use based on literature and assumptions.	2010 US$ discounted at 3% per year. Unit cost of screening, laparoscopy, oophorectomy, and ongoing care taken from previously reported Medicare reimbursement rates, although not all costs were explicitly stated. The perspective of the analysis was not explicitly stated, but appears to be the healthcare sector.	Thai baht (year not stated). Some unit costs of screening and treatment and savings are presented in Table 2. Costs per from lives and 'disability' saved are presented. It is unclear if some of these are model inputs or outputs.
Effectiveness Measures	Years of life saved discounted at 3%.	Years of life saved discounted at 3%.	Unclear.
Sensitivity Analyses	Extensive but deterministic one-way based on published ranges and author assumptions. For	Extensive but deterministic one-way sensitivity analyses only.	None.

	example: high risk prevalence for patients with family history.		
Effectiveness Results	Base case screen mortality reduction of 43%. Positive predictive value = 0.55%.	13% mortality reduction for ROCA.	Not presented.
CE Results	Base case ICER compared with no screen = $73,500 per life year gained reducing to $36k per life year gained for the high risk population.	The ICER for ROCA v no screen = $124,300 per life year gained.	Unclear measure of effectiveness and unclear if analysis is incremental. No results presented for the 'no screen' option.
Sensitivity Analyses Results	The ICER was sensitive to the screen cost, the sensitivity and specificity of screening, and the frequency of screening.	The ICER was sensitive to screen costs and test characteristics.	None presented
Author's Conclusions	Annual screening has the potential to be cost-effective, especially in high risk populations. Ideally the test specificity should exceed 99%.	ROCA achieves modest mortality gains with the currently accepted ICER. The ICER of 'better' tests will produce favourable mortality gains but costs would need to be held below $500 per test for an acceptable ICER. Cost-effectiveness improves if screening higher risk populations.	CA-125 is of 'increased' cost-effectiveness compared with the risk of malignancy index algorithm. However, it is not clear that this conclusion is valid based on the analysis presented by the author.

ICER: incremental cost-effectiveness ratio. ROCA: Risk of Ovarian Cancer Algorithm. TVS: transvaginal ultrasound.

4.2.1. Wiwanitkit 2013[55]

This, the most recent of the reviewed papers, was included based on the information available in the title and abstract. However, based on the full text of the article, this economic evaluation appears to be of poor quality and possibly even invalid. The description of the economic evaluation is lacking and it is unclear how it has been conducted, with the result that it is not reproducible. There is lack of clarity in how costs have been constructed or modelled. There is no clarity on any valid measure of effectiveness. The variable presented in the table summarising 'cost-effectiveness' (table 1) appears to be screening test sensitivity, but this is not clear. There is no evidence that the authors considered the costs and benefits of a no screening policy, however despite this they report the 'cost-effectiveness' of both considered screening options. Even if the 'cost-effectiveness' ratios were meaningful, at least one of the reported synthesised values (if not both) appears to be an average and not a marginal value, and as such is not relevant for decision making based on health economic evaluation. Further, there is no evidence of any sensitivity analysis being attempted. Given the weaknesses of this publication from a health economics perspective, this evaluation is not considered further.

4.2.2. Havrilesky et al 2008[58]

The model used for this economic evaluation was original developed as part of a technology assessment that considered the use of genomic tests for ovarian cancer screening, diagnosis, and treatment[59]. Within this assessment, the role of the model was to assess the sensitivity of results to different assumptions concerning the natural history of ovarian cancer. The model was later adapted twice, the first time to estimate the likely impact of different screening features (including interval lengths and test characteristics) on ovarian cancer screening[53]. The second adaption was to subdivide ovarian cancer into two types: aggressive and indolent[60]. The economic evaluation reported by Havrilesky *et al* in 2008[53] was deemed to be the most relevant to this study, and so is described in the following sections.

The objective of the study was to assess the likely cost-effectiveness of potential population screening strategies for US women at average risk of ovarian cancer. The screening tests under consideration were therefore hypothetical in order to assess the sensitivity of the cost-effectiveness results to changes in test costs and performance assumptions. Given the lack of screening trials (which were only ongoing, or had not started at the time of this publication), this was a laudable application of health economic modelling to test out the potential value of undertaking large

extensive trials. A well-described Markov model is presented, with transition probabilities based on well recognised and references sources (such as the Surveillance, Epidemiology, and End Results [SEER] national cancer database and US life tables). The model structure and base case assumptions are well described and referenced, thus making reproducibility of the model highly probable, although the studies selected to populate the model may be limited and or selective. The hypothetical population of women enter the model at age 20 and are followed until death (lifetime horizon). Women are assumed to receive a screen between the ages of 50 and 85 at varying modelled intervals. Costs and outcomes were discounted at an annual rate of 3%. The study perspective was the US healthcare system.

The model allows direct transition from stage 1 to advanced stage.[53] Unobservable data relating to annual rates of progression between stages and annual rates of detection (by stage) were estimated by manually varying the parameters until a good fit to the SEER stage-specific incidence data was obtained. Rates of progression were assumed to be the same for both undiagnosed and diagnosed cancers. As only hypothetical screening strategies were considered, test characteristics were manually chosen. The base-case test sensitivity and specificity values used were 85% and 95% respectively and were based on the results of studies involving ROCA[61] and CA-125[62]. Other base case model assumptions including probabilities of oophorectomy, hysterectomy, stage specific five-year survival, and chemotherapy were all comprehensively presented with either referenced sources or an explicit statement when the authors had used assumptions. These are detailed in Table 1 of the original paper, which also presents the ranges used in sensitivity analysis along with their sources. The base case screen test cost was an assumed $50, which was varied between $25 and $100 in sensitivity analysis. False positives were assumed to accrue an additional $100. Only direct medical costs were included in the model, with costs taken from 2007 Medicare reimbursement data (for the 2008 publication, costs were not considered in the 2011 publication). Separate treatment costs were calculated for each stage of (diagnosed) cancer, based on the type and length of chemotherapy received.

All three versions have the same thirteen health states:

- Well (people without ovarian cancer).
- Benign oophorectomy (assumed to occur for one quarter of all false positives).
- Undiagnosed ovarian cancer (four health states, relating to FIGO stages I to IV).
- Diagnosed ovarian cancer (four health states, relating to FIGO stages I to IV).

- Ovarian cancer survivors (originally people who are alive five years after a diagnosis of ovarian cancer, this value was changes to ten years in the two adaptations).
- Mortality (two health states: one for deaths due to diagnosed ovarian cancer, one for all other deaths).

For the second adaption of the model, the eight ovarian cancer health states were replicated; one for aggressive disease and one for indolent disease.

The impact of screening for ovarian cancer was included in the model as a stage shift in diagnosis, with an increased probability of being diagnosed at an earlier stage. This in turn leads to improved survival, as earlier stages have an increased probability of survival. It should be noted that diagnosis of ovarian cancer leads to a slight increase in the probability of mortality (compared to not being diagnosed) as individuals are then at risk from both non-ovarian and ovarian cancer mortality. Hence there may be some individuals who would not be diagnosed with ovarian cancer under no screening, but are diagnosed under screening. For these individuals, screening would be modelled as having a detrimental impact. However, the impact of this on the model results is likely to be small or negligible.

The model results estimate a lifetime risk of ovarian cancer of 1.38% (which compared with a risk of 1.42% observed in SEER). The modelled lifetime probability of death from ovarian cancer was 0.95% compared to 1.11% in SEER. Probabilities of mortality by stage at diagnosis were not compared. There was some evidence that the age-specific distributions for these two outcome variables under-estimated the risk of cancer and of death from cancer for woman aged above 75 years (as shown in Figure 2 of the original paper)[53]. In the base-case, risk of death from ovarian cancer was reduced by 43% using annual screening (compared with a strategy of no screening). Results presented by the show how this mortality risk reduction varied with changes in either screen frequency or test performance. Base-case positive predictive value was 0.55%, with each woman receiving an average of 1.06 false positive tests over their lifetime. The authors demonstrated that these outputs varied with specificity and frequency of screening[53]. Briefly, the base-care scenario predicted increases in average life expectancy of 2.92 days per woman at a cost of $589 per woman screened, resulting in an incremental cost per incremental life-year gained of $73,500. Based on the results of an extensive range of one-way deterministic sensitivity analysis (as presented in Table 4 of the original paper), the authors indicated that this value was sensitive to the cost of the screen, screening test

characteristics and the frequency with which women were screened.[53] The authors also explored the sensitivity of model outputs to compliance and age of first screen.

In their discussion the authors indicate that the main cost-effectiveness outcome measure (incremental cost per incremental life-year gained) was more sensitive to screen frequency than it was to the test sensitivity characteristic. The authors suggested that this could be expected in light of the fact that the model allowed for rapid progression from stage 1 to advanced stage disease. They further acknowledged that a potential weakness of their model was that it appeared to under-estimate risks of ovarian cancer and ovarian cancer mortality in women aged 75 years or older, and indicated that this may be due to issues around disease risk and calibration of cohort populations with cross-sectional data. Alternatively, this phenomenon could have been due to lower detection rates and / or faster progression of ovarian cancer amongst older patients. The impact of this under-estimation is that the cost-effectiveness of screening may be under-estimated, although this potential bias will be offset by the use of discounting. Finally, the authors indicated that the base-case screening costs may have over-estimated the costs of screening using CA-125 but under-estimated the costs of TVS. Given the sensitivity of the cost-effectiveness results to screening cost, it is important for future modelling work to ensure that screen cost is accurately estimated in future modelling work, and that such work should also consider quality of life in addition to survival when modelling cost-effectiveness.

There are some limitations with the model that should be noted. It is also how time since diagnosis is captured within the Markov model. No attempt was made to conduct a probabilistic sensitivity analysis, so the overall uncertainty in the model results cannot be quantified. The impact of screening on health-related quality of life is not considered. Calibration was performed manually; the authors mention that parameters were varied over clinically plausible range, and present good fit to SEER data (with the exception of women aged greater than 75). However, by not using a numerical calibration method, the authors risk missing potentially more relevant fits. There is also no way to quantify the fits obtained. Finally, the economic evaluation has a US healthcare perspective, with both natural history data and costs based on US data. Hence there may be limited generalisability of the model inputs and results to a UK setting.

4.2.3. Drescher et al 2012[12]

The model used for this economic evaluation started as a model for the natural history of ovarian cancer, along with the impact of CA 125 screening on this natural history, as reported in 1991[63]. A 1997 publication extended this model by also considering TVS screening and including a new survival component[64]. The model was further extended and refined in a 2012 publication examining the cost-effectiveness of multimodal screening for ovarian cancer[12]. The following summary relates to the model as reported in 2012 because this appeared to have important differences from previous versions of the model, and was judged to be the most relevant to this study.

The model is a stochastic microsimulation (time to event) model. The aim of the updated 2012 study was to compare the cost-effectiveness of a two-step annual screening strategy for a hypothetical cohort of 1,000,000 US women screened between the ages of 45 and 85. Four screening strategies were considered and are depicted in Table 4.2. Option 1 was a two-step algorithm using Ca-125 as first step, and TVS imaging as second step if triggered by CA-125 levels above a certain threshold. The other three options considered were two-step combinations involving either CA-125, or a hypothetical new biomarker as the first step, and TVS or a hypothetical new imaging technology as the second step. Screen option four in Table 4.2 comprised the sensitivity, specificity and costs of the hypothetical biomarker as step 1, and the hypothetical imaging technology as step 2. The hypothetical biomarker was assumed to have 2-fold improved sensitivity compared to CA-125 and the hypothetical imaging technology was assumed to have a 50% improved sensitivity compared to TVS (90% versus 63% respectively). The economic evaluation was a cost-effectiveness study using life years saved or gained as the measure of effectiveness. The costing perspective of the analysis was not stated, but appeared to be the healthcare sector, with costs primarily taken from Medicare reimbursement data. Separate treatment costs were used based on stage of cancer, and if treatment was in the first, last, or other year of cancer. The costs of CA 125 and of TVS were taken from the 2011 publication of the Durham model, and were $31 and $111 respectively. However, these costs could not be identified in the Durham publication (which did not appear to report any costs). Costs were indexed to 2010 US$. Both costs and life years were discounted at 3% per year.

Table 4.2: Screening strategies evaluated in *Drescher et al 2012*[12]

	TVS as step 2	Hypothetical step 2 scan
CA-125 as step 1	Screen Option 1	Screen Option 2
Hypothetical biomarker as step 1	Screen Option 3	Screen Option 4

The key input assumptions used in the model are presented in Table 1 of the original paper.[54] Only single sources were quoted for assumptions derived from the literature, which suggests that the authors may have been selective in their sources. In particular, sensitivities and specificities for CA-125 and TVS come from separate studies, which may hamper comparisons between the two screening modalities. Some of the assumed costs come from the Havrilesky *et al*[53] study reviewed above. A simple approach was taken for sensitivity analysis, with one-way deterministic sensitivity analyses often using (the slightly arbitrary) halve and double of input values.

Upon entering the model, an age at death and age at diagnosis of ovarian cancer (along with stage at diagnosis) is sampled for each woman. From age (and stage) at diagnosis, age at inception of ovarian cancer is derived by subtracting stage-specific estimates of disease duration. Women are then classified into one of four groups:

- Healthy: no ovarian cancer.
- Case: symptomatic ovarian cancer.
- Benign: non-malignant tumours.
- Latent: ovarian cancer that is not diagnosed before death.

Disease progression post-diagnosis does not appear to modelled. Women who live for fifteen years with a diagnosis of cancer, are assumed to be cured, this appears to be the only possible move between the four groups.

Natural history data relating to ovarian cancer incidence and survival were taken from the SEER program. Data relating to unobservable time to disease progression were based on mean responses from 39 gynaecological and medical oncologists who responded to a survey (total number approached: 80). Separate times were estimated based on stage (I to IV), grade (low or high) and histology (serous, mucinous, endometriod, clear cell, or adenomcarcinoma not otherwise specified). The sensitivity and specificity of TVS were taken from the PCLO study, and were 63% and 97% respectively. The specificity of CA 125 was assumed to be equal to 95%. Sensitivity of CA 125 was obtained by analysing data from the CARET study, and varies with time prior to clinical diagnosis. A

numerical values is only reported for one year prior to clinical diagnosis (67%), other values are graphically displayed and are approximately 22% (2 years prior), 20% (3 years prior) and 10% (4 or more years prior).

The impact of screening was included in the model as a stage-shift from late (stage III or IV) to early (stage I or II) ovarian cancer. If a stage-shift occurs, the time to mortality is resampled, although this is only used if it is longer than the time to mortality originally sampled for late-stage. It is unclear if this method differs from that reported in the previous version, which sampled the same percentile point from early and late stage survival distributions, to ensure that earlier diagnosis did not lead to earlier mortality. It is assumed that all positive screens (including false positives) receive surgery. The impact of surgery (if any) on mortality and morbidity are not stated.

The primary model results for cost-effectiveness as presented by the authors are represented in the top-half of Table 4.3 ('original results'). The incremental cost-effectiveness ratios (ICER; defined for the Drescher *et al* study as incremental costs divided by incremental life-years[12]) for each of the four screening options presented by the authors were calculated by comparing each of the four screening options with the no screen option. Whilst there is some merit in presenting cost-effectiveness results in this manner, it is not the most appropriate approach to use when using health economic evaluation to inform decision making using mutually exclusive health technologies (that is, we would only want to decide on funding one of the four presented screening strategies, or the strategy of no screening). A fully incremental cost-effectiveness analysis would have been more appropriate. This is achieved by ordering all of the five options in order of total discounted costs (from lowest to highest), and then calculate the ICERs for each option by comparing with the next cheapest option, after first removing any options that are both more expensive and more effective when compared with a single other option (are dominated) or when compared with a combination of two other options (are extendedly dominated). For this report, fully-incremental ICERs have been calculated and are presented in the bottom half of Table 4.3 ('derived results'). These derived calculations required knowledge of the total discounted life years as estimated by the health-economic model. This key information was not presented in the main paper and but were available in Table 2 of the supplementary data.

Table 4.3: Cost-effectiveness results are originally presented in Drescher *et al*[12], and as derived for this report.

Original results	No Screen	Option 1	Option 2	Option 3	Option 4
Mortality Reduction		13%	23%	25%	30%
Life years gained per screen-detected case		1.68	1.61	2.09	2.21
Total costs	$865m	$1,741m	$2,397m	$5,401m	$6,068
ICER compared with no screening		$88,993	$124,376	$205,248	$191,441
Derived results	**No Screen**	**Option 1**	**Option 2**	**Option 3**	**Option 4**
Total Life years (estimated)		9,835	12,312	22,097	27,174
Incremental ICER (including dominated strategies)		$89,000 vs no screening	$265,000 vs Option 1	$307,000 vs Option 2	$131,500 vs Option 3
Incremental ICER (removing dominated strategies)		$89,000 vs no screening	Extendedly dominated	Extendedly dominated	$249,500 vs Option 1

ICER: incremental cost-effectiveness ratio (incremental costs / incremental life years).

Using fully-incremental calculations, both option 2 and option 3 are both extendedly dominated by combinations of screening options 1 and 4, and as such were excluded from further cost-effectiveness calculations. Consequently, the ICER for option 4 (hypothetical biomarker and hypothetical imaging) compared with option 1, has been estimated to be $249,500 per life year (as compare to $191,441 when compared with no screening, as originally presented). The authors' original conclusion that option 4 may be cost-effective if costs can be reduced and, or test performances improved, may be true, but this conclusion is considerably less likely using the correct incremental analysis.

Overall, this paper is not as well presented as the study be Havrilesky *et al*[58]. Although the clinical element of the model appears good, the authors' assessment and presentation of cost-effectiveness results has scope for improvement. Further, the authors' original 1997 paper confuses marginal and average cost per life-year gained when presenting the main results, and does not indicate that at least one of the screening options assessed could be excluded due to the existence of extended dominance. The authors appear to have made similar errors in their more recent paper as previously described. The presentation of the total discounted costs for each option, including no screening, is however both helpful and appropriate. The 2012 analysis does not include the appropriate incremental cost-effectiveness analysis (as described) and as such the authors' results and conclusions with regards to screening options 2, 3 and 4 appear to be incorrect. The authors failed to

discuss the fact that options 2 and 3 should be excluded from consideration using the base-case ICERs due to extended dominance by options 1 and 4. In addition, they have significantly under-estimated the ICER for option 4 in comparing it to the no screening option. The presented ICER for option 4 is under-estimated by approximately $60,000 per life year gained, as option 4 should be compared with option 1 (ICER = $249,500 per life year gained). In addition, estimation of unobservable time to disease progression was based on physician's estimates. By not using a numerical calibration method, potentially more relevant estimates may have been missed. There is also no way to quantify the fits obtained. Finally, the economic evaluation has a US healthcare perspective, with both natural history data and costs based on US data. Hence there may be limited generalisability of the model inputs and results to a UK setting.

4.3. Discussion

This rapid review has considered published economic evaluations on screening for ovarian cancer amongst the general population. The quality of the latest study[55] is not considered to be study of acceptable quality for the reasons previously discussed. The two remaining economic evaluations are both from the USA. Both evaluations used years of life saved as the primary measure of effectiveness for their cost-effectiveness calculations. Another USA based economic evaluation of potential interest, was not in publication at the time of writing, but was available as a conference abstract.[52] As such this Ding *et al* economic evaluation was excluded from this review, however, given that it was the only study available which used quality-adjusted life-years (QALYs) as the main outcome measure, the results of the abstract are summarised below.

The stated aim of the Ding *et al*[52] analysis was to evaluate the cost-effectiveness of multimodal screening (annual CA-125 screening with TVS follow-up where deemed appropriate based on CA-125 level) compared to a no screening policy. A secondary analysis evaluated the cost effectiveness of the above screening method with that of annual TVS screening alone. The abstract states that screening was considered for USA post-menopausal women aged 65 to 69. Although not explicitly stated, the implication is that this age range is the start age for screening. No finish age was specified. The analysis purports to take a societal perspective although this cannot be verified from the information available. The abstract does not give the test sensitivities and specificities used in their model, but does indicate that they are taken from the UKCTOCS (trial NCT00058032), presumably using evidence from the prevalence screen[6]. The economic evaluation was conducted using a 'backward induction' (it is unclear what this method entails) model based on 5-year time periods over the patients'

lifetime. A 3% discount rate was used for both costs and QALYs and costs were reported in 2009 US$. The primary base case result quoted is an ICER of $221,662 per QALY for multimodal screening versus no screen, based on incremental costs and QALYs of $820 and 0.0037 respectively. The authors indicate that the cost-effectiveness of multimodal screening improves when compared with TVS alone (although quantitative results are not presented). The authors indicate that the ICERs are sensitive to disease incidence, and to cost of screening, but the sensitivity analysis model inputs or outputs are not reported, nor is clear whether the sensitivity analysis was probabilistic or deterministic in nature. The authors conclude that their primary base-case ICER is above currently acceptable US willingness-to-pay thresholds for oncology ($120,000 to $150,000 per QALY), and as such, screening for ovarian cancer using their model appears not to be cost-effective. Sensitivity analyses did however indicate that targeting screening at higher risk women and/or reducing the costs of screening could achieve an ICER below $120,000 per QALY.

The two main evaluation identified both appear to be built on robust and valid clinical pathway models. However, a significant weakness of both models include possible selection bias in that most of the key input parameters are based on single study sources and/or expert opinion assumptions. In addition, the majority of the screening options considered in the two economic evaluations are concerned with hypothetical rather than existing options. In addition, the cost-effectiveness results presented by Drescher *et al*[12] are not fully incremental and as such have limited relevance for decision making. Also, the sensitivity analyses presented by both papers are limited in that they are both concerned with univariate deterministic sensitivity analysis. The sensitivity analysis could have been made stronger by consideration of a multivariate probabilistic approach, which would also have facilitated value of information analyses.

In addition to the above methodological limitations, the two presented economic evaluations (along with the abstract of Ding *et al*) have limited usefulness from the perspective of the UK and NHS decision making bodies such as the National Institute for Health and Care Excellence (NICE). Both models are based on USA costings, have a US healthcare perspective (neither of which is likely to generalise to a UK setting) and neither of the two published evaluations consider QALYs as an outcome measure.

In conclusion, there is need for a good quality UK-based model, capable of probabilistic analyses (including value of information analyses). For the context of the UK NHS such a model should ideally consider QALYs as the primary measure of effectiveness, so that the results of the evaluation can be compared with other healthcare and screening interventions and to make it relevant for consideration by NICE. At the time of writing no such model appears to have been published.

5. Review of health-related quality of life studies.

The objective of this section is to present a systematic review of the evidence regarding the health-related quality of life (HRQoL) of women who either have ovarian cancer (including the impact of treatment) or receive screening for ovarian cancer. The methods and results of searching and sifting are summarised and the included economic evaluations reviewed and compared. The included, and one excluded, economic evaluations are discussed in terms of their quality, and their usefulness for informing UK decision making regarding NHS resources.

5.1. Methods and search results

The search for evidence pertaining to health-related quality of life was carried out at the same time as the search for cost-effectiveness studies. The methods for this are described in Section four, with further details provided in Appendix 1. The searches of electronic databases yielded 2,089 articles after removing duplicates. Sifting based on titles, abstracts, and keywords resulted in 78 articles of potential interest. Articles for which it did not appear possible to identify quality of life effects specifically for ovarian cancer (for example, if they only considered gynaecological cancers as a whole) were excluded, as were conference abstracts.

The remaining 78 articles consisted of 13 reviews and 65 articles. Of the 13 review papers, one was the meta-analysis of Reade *et al*[5], which is described in detail in section 3.2. The remaining 12 potentially relevant reviews were screened to assess their relevance to this study. Of these, 5 were excluded as they only considered methodological aspects of the measurement of HRQoL[65-69], 3 were excluded as their reported content was superseded by more recent reviews[70-72], and 3 reviews were excluded as they focused on patient populations that was not of direct relevance to this study[73-75]. One review article remained[76]. This systematic review of the HRQoL of women with ovarian cancer is described in detail below.

5.2. Evidence from systematic reviews.

5.2.1. Hess et al[76]

This study, published in 2012, is a systematic review and meta-analysis of HRQoL in Ovarian cancer studies. The justification given for undertaking this study was due to the importance of HRQoL in

women with ovarian cancer, given that ovarian cancer has the highest mortality rate of all cancers of the reproductive system, and the fourth highest cancer-related mortality amongst women. In addition, the relatively poor advances in survival in this cancer compared to other forms of cancer in recent times indicated the need for a review. This study was undertaken to update a previous systematic review undertaken in 1996. Although results are presented for HRQoL overall and HRQoL sub-scales in the original paper, given the primary aim of the current study to inform populating a cost-effectiveness model with health utility values, the sub-scale results are not considered here.

Studies were identified using searches with relevant ovarian cancer and HRQoL search terms for English language studies in Ovid or Medline up to the end of December 2011. Studies covering multiple disease sites were only included if results for ovarian cancer were reported separately. In total, 40% of the searched papers were repeat-reviewed by two other reviewers. As there was almost 100% agreement for these, further repeat-reviewing was not undertaken.

From 844 papers identified by the search, 170 were included, representing 139 individual studies. These included 48 RCTs, 45 cross-sectional observational studies, and 36 non-randomised clinical trials. The populations included were primarily white non-Hispanic (88.5%). Of the identified studies 27% were concerned with primary disease alone, 25% with recurrent disease alone, and the remainder with both. It was not possible to differentiate between stages at diagnosis in the pooled data, as only 1.2% and 11.8% of studies focused on patients with early or advanced disease respectively.

More than 90 different validated instruments were used in the included studies, the most common instrument being the EORTC-QOL (37.1%), followed by the FACT-O, which was reported in 23.5% of studies. The SF-36 or the SF-12, were reported in only 7.6% of studies.

Response rates appeared to be dependent on the type of study design, with RCTs having the lowest response rate for HRQoL questionnaires. Only around half of the studies provided data which could be pooled in meta-analysis. Of these studies, more than half did not produce a statistically significant result. There were six studies which identified HRQoL as a significant prognostic factor for improved survival.

Only 11 of the 38 identified RCTs reported significant differences in HRQoL outcomes between treatment arms. However, 9 of these 11 did not demonstrate corresponding significant improvements in either survival or tumour response. The remaining two studies reported inverse relationships between HRQoL and survival or response rate. The treatments considered in these trials were intraperitoneal therapy and concurrent docetaxel therapy.

No differentiation could be made between mean overall HRQoL for different therapeutic regimes, because of a lack of available data to pool. This in turn was due to heterogeneity in the different interventions, instruments and populations used in the studies.

In newly diagnosed disease, there was evidence of improved HRQoL over time using FACT-G, FACT-O and QLC-30, through treatment cycles and particularly when the follow-up period is included in the analysis. In recurrent disease, there was only evidence for improved HRQoL over time using FACT-G. Statistical significance was not achieved using FACT-O, and there was insufficient data for QLC-C30.

This systematic review indicated that the number of publications reporting HRQoL in respect of ovarian cancer has increased from only 12 prior to 1996 to over 800 in the current study. Despite the large number of studies now available, the limited meta-analysis in the current study demonstrated how challenging this is in the context of use of a large number of HRQoL instruments in included studies. As such most of the studies found could not be included any meta-analysis due to the disparity of instruments used and time points of data collection. Nevertheless, some statistical significance was achieved.

The results of the study suggest that HRQoL can improve or decline during the treatment phase and this may vary by instrument used. Cancer specific HRQoL instruments FACT-0, FACT-G, and QLQ-C30 demonstrate significant improvements in HRQoL after completion of primary therapy. However, the study authors indicated that there is only limited longitudinal data beyond the initial treatment and follow-up period.

Differences in HRQoL were seen in only a few RCTs comparing treatments, and the reviewers noted that any differences found are likely to reflect short term effects of toxicity rather than longer term HRQoL. Completion rates can also be affected in RCTs. Interpretations of toxicity have been shown to differ between physicians and patients, and also between patients and the general public.

Although this study indicated that the number of studies reporting HRQoL in ovarian cancer has increased since 1996, because of the differences in timings of data recording and use of a large number of HRQoL instruments, it remains challenging to pool data for meta-analysis. The study authors called for greater standardisation in the measurement and the collation of HRQoL data in ovarian cancer studies.

5.2.2. Conclusion

To conclude, the systematic review by Hess *et al*[76] provides an overview of the effects of treatment on HRQoL, but it does not provide evidence about the potential impacts on HRQoL of either screening for ovarian cancer, or the development of symptomatic ovarian cancer. To identify this evidence, the remaining 65 individual articles were screened to see if they provided any information on either the effects of screening or the development of symptoms on HRQoL. Articles whose results were unlikely to generalise to the English healthcare setting were excluded, as were those that only considered the HRQoL of a subset of women with ovarian cancer.

To ensure that no other potentially relevant articles were excluded, the citations of all the review articles were checked, and all of the articles that had referenced these review articles (using Google Scholar, searches performed on the 14th November 2014) were checked. This resulted in the identification of two extra articles[77,78]. In addition, the authors were also aware of one relevant paper published by the UKCTOCS team[35].

5.3. Evidence about the impact of developing symptomatic ovarian cancer on HRQoL.

Three studies were identified as being of potential relevance[58,77,78]. However, two of these studies[77,78] were restricted to surveys of ovarian cancer survivors, and so deemed to not be of relevance to this project. Details about the remaining study are provided below.

5.3.1. Havrilesky 2009[58]

The study by Havrilesky *et al* [58] was designed to generate a set of validated utility measurements for women diagnosed with ovarian cancer. A set of relevant health states was drafted by the study group, and then refined following consultation with a focus group, which included four gynaecological oncologists (covering three different clinical roles) and a clinical service worker. The resulting 25 health states covered ovarian cancer screening, diagnosis and treatment. These health states were valued by a sample of 37 female members of the public who did not have a history of ovarian cancer (known as volunteers) and 13 women with a history of ovarian cancer (patients). Both volunteers and patients valued health-states relating to chemotherapy treatment (which focused on its side-effects), whilst only the volunteers were used to value health states relating to screening for and diagnosing ovarian cancer. Valuations were performed using both time-trade-off (TTO) methods and the visual analogue scale (VAS).

In general, TTO-derived utilities were higher than those derived using the VAS. It has been noted that the former method is usually more accurate than the latter[10]. Hence, the subsequent discussion shall only consider TTO utilities.

For screening, two tests were considered: the use of a blood test or the use of TVS. Utilities for screening with these were elicited, as were utilities for screening leading to a false-positive result (giving four screening health states). All four health states had the same median utility of 0.97. Mean values were always below the median, ranging from 0.90 (both screening with a blood test and screening with TVS with a false-positive result) to 0.83 for screening with TVS. There was wide variation in valuations within a health state, with standard deviations ranging from 0.14 (screening with TVS with a false-positive result) to 0.27 (screening with TVS). In addition the results appeared to lack some face validity, as the mean value for screening with TVS with a false-positive result (0.90) was higher than screening with TVS with any result (0.83).

There were 10 different health states relating to diagnosed (including recurrent) ovarian cancer. The highest mean value was for ovarian cancer in clinical remission (0.83), whilst the lowest was for end-stage ovarian cancer (0.16). The mean value for newly diagnosed early ovarian cancer was 0.81, whilst for newly diagnosed advanced ovarian cancer it was 0.55. The remaining six health states related to the grade of toxicity experienced (either 1 to 2 or 3 to 4), along with whether the cancer was newly diagnosed, recurrent-responding, or recurrent-progressing. Mean values for these six health states ranged from 0.61 (recurrent ovarian cancer – responding to chemotherapy/grades 3 to 4 toxicity) to 0.40 (recurrent ovarian cancer – progressive/grades 1 to 2 toxicity). Again there was wide variation, with all standard deviations greater than 0.24.

Eleven health states were included which related to chemotherapy side-effects. Valuations were provided by both the patient and volunteer groups. Median values ranged from 0.97 for both grade 2 alopecia and grade 1-2 alopecia, to 0.33 for grade 3-4 myalgia/pain. Patients valued six out of the 11 side effects more highly than did the volunteers, although this difference was only statistically significant for grades 1-2 peripheral neuropathy and grade 3-4 Myalgia.

A limitation of the study by Havrilesky *et al*[58] is the small sample size, which manifests itself in the wide variation about the utility estimates. In addition, there were demographic differences between the volunteer and patient population, which may affect the validity of the results. However, the health state utility values presented are the most relevant that have been identified in the literature. In addition, the results are likely to be generalizable to an English setting (as they focus on the

impact of screening, diagnosis and treatment instead of being specific to any healthcare setting or levels of resource use).

5.4. Evidence about the impact of screening for ovarian cancer on HRQoL.

In addition to the study by Havrilesky *et al*[53], two studies were identified as being of potential relevance [35,79]. However, one of these studies[79] only reported baseline results at recruitment to a screening study and so was deemed to not be of relevance to this project. Details about the remaining study are provided below.

5.4.1. Barrett 2014[35].

Barrett *et al*[35] used longitudinal data from the psychosocial sub-study of the UKCTOCS (comprising 91.6% of the trial participants, n = 185,693) to investigate the impact of screening on HRQoL. Questionnaires were completed by all women prior to randomisation, with further questionnaires completed by women receiving a positive screen result (after the screen and annually). Two of the main health-related outcome measures were the State/Trait Anxiety Inventory and the General Health Questionnaire 12.

The authors noted that screening on its own was not associated with an increase in anxiety. In addition, anxiety was not increased by either repeat testing or more invasive testing.

Strengths of the study are the large sample size and the longitudinal collection of outcome measures. However, the authors noted that women who left the trial were excluded from the psychosocial sub-study, and that this may have an impact on the results. A further limitation is that HRQoL was not measured with the EQ-5D.

The findings reported by Barrett *et al*[35] are in agreement with those reported in the meta-analysis of Reade *et al*[5] – which, as discussed in Section 3.5.3, found no impact of screening on HRQoL. However, as with the UKCTOCS psychosocial sub-study, EQ-5D was not used as an outcome measure.

5.5. Summary of the available HRQoL evidence.

There is a lack of evidence that the HRQoL of women with ovarian cancer varies with the type of treatment received. Whilst a previous review[76] identified a large number of studies (170), less than half could be pooled in a meta-analysis due to heterogeneity in the utility measures used, treatments received and populations considered. Women with screen-detected cancers may receive different treatment to women who present symptomatically (for example, due to earlier detection). However, there is currently insufficient evidence as to the potential impact of any different treatments on HRQoL. However, there was consistent evidence to suggest that HRQoL following treatment was improved compared to HRQoL prior to treatment.

One study was identified that provided useful evidence on the impact of stage at diagnosis on HRQoL[58]. As expected, women diagnosed with more advanced disease are likely to have a lower HRQoL than women with early-stage disease.

A systematic review of screening trials concluded that screening does not impact on HRQoL[5]. This was based on three trials, with subsequent results from the psychosocial sub-study of the UKCTOCS trial confirming this finding[35].

A limitation of the studies identified is that none used the EQ-5D to measure utility values. The EQ-5D is the measure recommended in the NICE methods guide when conducting health economic evaluations[80]. The lack of EQ-5D studies along with the limited HRQoL evidence base suggests that there is considerable uncertainty over the HRQoL profile of women with ovarian cancer or receiving screening for ovarian cancer.

6. Costs and resource use.

This chapter provides evidence relating to care pathways for ovarian cancer in England. This covers screening, diagnosis and treatment. Five key evidence sources helped to inform the results presented here. They are: NHS reference costs[81], NICE clinical guidance 122 on the recognition and initial management of ovarian cancer[10], cancer registry data (supplied by Public Health England) on treatment use[14], the INCISIVE Health report[82], and expert clinical opinion.

6.1. Costs and resource use of ovarian cancer screening.

There are two main types of ovarian cancer screening (which may be used in isolation or combination): TVS or screening using the CA-125 blood test, which may be interpreted with or without the ROCA.

The cost of gynaecological ultrasound may be obtained from 2012/13 NHS reference costs[81] (2013/14 reference costs do not distinguish between different types of ultrasound), under the HRG code RA24Z "Ultrasound Scan, 20 minutes and over". There were 65,102 gynaecological-related examinations, of which all but 3 were amongst outpatient appointments. Of the outpatient examinations, 98% were classified as 'gynaecological' with a mean value of £55 (£43 to £65), the remaining 2% were classified as 'gynaecological oncology' with a mean value of £73 (£59 to £77).

It is anticipated that if screening for ovarian cancer is implemented in England, then the resource use required is likely to be similar to that observed for the UKCTOCS trial[6].

Within the UKCTOCS, first-level TVS were performed by 'type 1' sonographers, whilst second-level TVS were performed by 'type 2' sonographers. As type 2 sonographers are more experienced than type 1, the cost of a second-level TVS is likely to be greater than that of a first-level TVS. This cost difference may be estimated by assuming that type 1 sonographers are employed at Agenda for Change Band 7, and type 2 sonographers at Band 8a (these bands were frequently observed in published job adverts for sonographers and superintendent sonographers). Comparing pay at mid-points of the bands[83], type 2 sonographers cost 21% more.

There are no routine estimates of the cost of the CA-125 blood test. Instead, the cost estimates developed for the European Cancer Detection Consortium (ECDC, personal communication; there is an overlap in authors for this study and the ECDC) were used. The ECDC costing exercise estimates

separate costs for taking the blood sample (including the cost of consumables and transporting to a laboratory for processing), and for the CA-125 biomarker test. For the blood test it was assumed that 10 minutes of staff time would be required. The member of staff involved was assumed to be either a phlebotomist or a practice nurse, with hourly costs of £25 and £44 respectively, based on national routine costs[84]. It was assumed that the blood test was taken at a GP surgery and may or may not have been as part of the current NHS Health Checks; blood tests as part of the NHS Health Checks were assumed to cost £2, whilst those performed outside the NHS Health Checks were assumed to cost £25 (these costs were in addition to the staff costs). Two different cost estimates for the serum CA-125 biomarker test were available: one based on a cost-utility analysis performed to support NICE clinical guidance[10], which used a cost of £23, based on the consensus of the guideline development group. These costs were for 2007/08, which can be inflated to a 2013/14 price of £25.02[84]. An alternative cost for CA-125 was derived from a large district hospital, resulting in a price of £35. In addition, for ovarian cancer, the economic evaluation reported by Havrilesky *et al* used a cost of $29.07 (2007 prices) based on Medicare reimbursement data[53]. This value was also used in the economic evaluation reported by Drescher *et al* (inflated to 2010 prices)[12]. Assuming that $1 = £0.66, the 2007 US$ cost may be inflated to a 2013/14 price of £21.69. As this price is closer to the NICE estimate than the hospital-based estimate, the later was used in the base-case. Hence the base-case analysis assumed that the blood test was performed by a phlebotomist outside the NHS health checks, with the NICE estimate of CA-125 testing, resulting in a mean test cost of £54.19. Two sensitivity analyses considered alternative 'high' and 'low' costs. The high cost used a practice nurse for taking the blood test, and the hospital-based estimate of CA-125 testing, resulting in a mean test cost of £67.33. The low cost altered the base-case by assuming that the blood test occurred as part of the NHS Health Checks, resulting in a mean test cost of £31.19.

The cost (if any) of implementing ROCA is currently unknown, for the health economic modelling it was assumed that implementing ROCA would not incur any additional costs beyond that of CA-125 testing.

It was assumed that the costs involved in inviting people to the screening programme, communicating test results, and maintaining a helpline would be the same as that for colorectal cancer; this has previously estimated to be £2.09 per invitation[85].

6.2. Diagnosis and treatment pathways.

The current diagnosis pathway for suspected ovarian cancer, as recommended by NICE guidance[10], may be summarised as follows;

- Measure CA-125 levels and perform an ultrasound of both the abdomen and pelvis (if these have not already been measured and performed).
- Use the results of the ultrasound, CA-125 measurements and menopausal status to calculate the risk of ovarian cancer (known as the risk of malignancy index), referring women with a risk score of 250 or greater.
- For referred women, perform a computerised tomography (CT) scan, in preference to a Magnetic Resonance Imaging (MRI) scan.

Following the diagnosis pathway, women who are still suspected of having ovarian cancer undergo surgery to confirm (or refute) the suspicion. There is little explicit guidance about the types of surgery that should be performed, save that retroperitoneal lymphadenectomy (lymph node dissection) should not be included for women who appear to have Stage I cancer, with lymph node sampling performed instead. This recommendation is consistent with SIGN (Scottish) guidance[9]. In addition, it has been suggested that surgery should include hysterectomy, bilateral salpingo-oophorectomy, tumour debulking, and omentectomy[21].

The risk of malignancy index is a function of CA-125 level, menopausal status, and a classification of TVS. Referrals based on this index would be different to referrals that use a fixed CA-125 score (and so do not account for menopausal status or any characteristics identified via TVS. It would also be different to referrals based on ROCA, which takes into account changes over time in CA-125.

For ovarian cancer treatment, chemotherapy may be offered either before surgery (known as neo-adjuvant chemotherapy) or following surgery (known as adjuvant chemotherapy). The offer (or not) of chemotherapy, along with when it is delivered depends upon the diagnosed stage of ovarian cancer. For women with Stage I cancer it is recommended that chemotherapy is only offered if they received optimal surgical staging and are deemed to have high risk (high grade) disease. In these instances chemotherapy should consist of six cycles of adjuvant therapy with carboplatin. For women with ovarian cancer that is more advanced than Stage I, it is recommended that they have a choice between two different types of chemotherapy; either a platinum based-therapy (cisplatin or carboplatin) on its own or in combination with paclitaxel. Treatment is typically for six cycles, at 3-weekly intervals.

6.3. Costs and resource use for diagnosis and treatment.

A key document used as the starting point for estimates of resource use and costs was the INCISIVE report[82]. This document is briefly described, before presenting estimates of resource use and costs for diagnosis and treatment.

6.3.1. The INCISIVE report.

The INCISIVE report was commissioned by Cancer Research UK to examine the financial implications of achieving earlier diagnosis of colorectal, lung and ovarian cancers. To achieve this, patient pathways for each disease were created, populated and costed based on a mixture of national data sources, clinical guidance, clinical audit and input from clinical experts. For ovarian cancer these data sources included: the Map of Medicine; NICE clinical guidance, technology appraisals and interventional procedures guidance; the Systematic Anti-Cancer Therapy dataset; and the International Cancer Benchmarking Partnership. Where possible, costs were taken from standard national sources (reference details are not explicitly provided, but where possible these were traced back to their source by this study team to check their authenticity).

Resource use and costs were displayed for each stage of ovarian cancer, and separated by the different parts of the diagnosis and treatment pathway: diagnosis, surgery, chemotherapy, and other (follow-up and palliative care). All estimates of resource use were based on expert opinion. No evidence was presented about the likely resource use and costs for women with borderline ovarian cancer or for women without ovarian cancer.

6.3.2. Diagnosis of ovarian cancer.

Estimates for the diagnosis of ovarian cancer may be derived directly from the INCISIVE report. Alternative estimates were also obtained from clinicians based in Sheffield. These two estimates for the cost of diagnosing ovarian cancer are displayed in Table 6.1 by stage. For the INCISIVE estimates, the costs for women with borderline or no ovarian cancer were assumed to be the same as those for women with Stage I cancer. A detailed breakdown of the resource components, costs and resource use are provided in Appendix 2.

Table 6.1: Costs of diagnosing ovarian cancer, by stage.

	Stage I	Stage I	Stage I	Stage I	Borderline*
INCISIVE report	£462	£505	£548	£361	£462
Sheffield estimate	£110	£116	£126	£126	£112

* Includes no ovarian cancer, not explicitly stated in the INCISIVE report so assumed equal to Stage I.

The Sheffield estimates are much lower than those from the INCISIVE report, this is due to a reduced use of MRIs (as is consistent with NICE guidance) and no use of PET-CT scans.

6.3.3. Treatment for ovarian cancer.

There are two main types of treatment for ovarian cancer: surgery and chemotherapy.

The INCISIVE health report includes a breakdown of the different surgical procedures received, such as total hysterectomy, bilateral salpingo-oophorectomy (BSO) and infracolic-omentectomy, as well as the different chemotherapy regimens received. For each of these procedures, an estimate (based on expert opinion) of the proportion of all women with ovarian cancer who will receive this procedure is provided. However, for this study an estimate of the procedures received as a proportion of all women who receive surgery (which will be a subset of all women with ovarian cancer) was required. It was felt that these estimates could not be reliably obtained from the INCISIVE health data. Instead estimates of the proportions of each surgical procedure received, and each chemotherapy regimen received, by stage (amongst women who receive surgery) were directly obtained from Sheffield clinicians.

The estimates of resource use amongst women who receive surgery and/or chemotherapy were combined with data from the English Cancer registries[14] about how many women with ovarian cancer receive surgery and/or chemotherapy, to generate costs of treatment by stage. These costs are displayed in Table 6.2; a detailed breakdown of their derivation is available in Appendix 2.

Table 6.2: Cost estimates for the treatment of ovarian cancer, by stage.

	Proportion receiving[1]		Treatment cost[2]		Total Cost[3]
Stage	Surgery	Chemo	Surgery	Chemo	
I	89%	53%	£6,243	£2,655	£6,961
II	81%	71%	£6,243	£3,219	£7,325
III	58%	71%	£6,334	£7,465	£9,016
IV	23%	50%	£6,334	£8,763	£5,892

[1]Based on cancer registry data. [2]Based on Sheffield estimates. [3]Includes an additional cost of diagnostic laparoscopy (£1,184) amongst 5% of women with stage III and 5% of women with stage IV ovarian cancer. Chemo: chemotherapy.

Of note, treating stage IV ovarian cancer is estimated to be cheaper than treating earlier stages. Thus identifying ovarian cancer at an earlier stage may increase costs. However, it is anticipated that the lower costs for stage IV may be offset by the reduced life expectancy of being diagnosed at this stage.

6.4. End of life care.

Estimates of the cost of end of life care were taken from Guest *et al*[86], which was the same study as used by the INCISIVE report. Guest *et al* reported a cost of £4,789 for ovarian cancer at 2000/2001 prices. This value was inflated to a 2013/14 price of £7,080[84]

7. Conceptual modelling: ovarian cancer natural history, screening, diagnosis and treatment.

Conceptual models are an important component of the modelling process. The conceptual models are designed to be a visualisation of current knowledge and understanding about the disease processes. They encapsulate both patient health states and healthcare events, along with their potential interactions. These conceptual models shall then be used to help decide what detail shall go into the mathematical (simulation) model.

Conceptual modelling was carried out based on the framework described in Tappenden *et al*[87]. Three conceptual models were developed, one relating to the natural history of ovarian cancer in the absence of screening, and two relating to services for ovarian cancer. The two services were screening and treatment (including diagnosis). The disease-level (natural history) conceptual model was developed based on an understanding of the available ovarian cancer literature – including existing economic evaluations, as detailed in section 4 – and feedback from clinical advisors. The resulting conceptual model is displayed in Figure 7.1. The conceptual model for ovarian cancer screening is based on the screening process as used in the UKCTOCS trial[6] and displayed in Figure 7.2, whilst the treatment conceptual model is based on the treatment pathways developed by Incisive Health (in collaboration with their own clinical experts)[82] and is displayed in Figure 7.3.

7.1.1. Natural history conceptual model.

The scope of this evaluation considers a patient population of postmenopausal women who initially do not have ovarian cancer and are not at high risk of developing ovarian cancer. It is assumed that, if an ovarian cancer develops, it will be either:

a. FIGO stage I epithelial ovarian cancer, or
b. Ovarian cancer that is either borderline or non-epithelial. This includes ovarian neoplasms of uncertain behaviour.

Epithelial cancers are the most common type of ovarian cancer, constituting about 90% of all cases, as described in Section 2.1. Borderline cancers are of low-malignant potential with little impact on mortality (Section 2.3), their inclusion in the conceptual model is to represent concerns that any increased diagnosis of borderline cancers due to screening may reflect an over-diagnosis[6]. Over-diagnosis occurs if a lesion is diagnosed, and treated as if malignant when it is actually benign (or

very slow-growing) and so unlikely to lead to symptoms or other adverse health outcomes in a subject's lifetime[18].

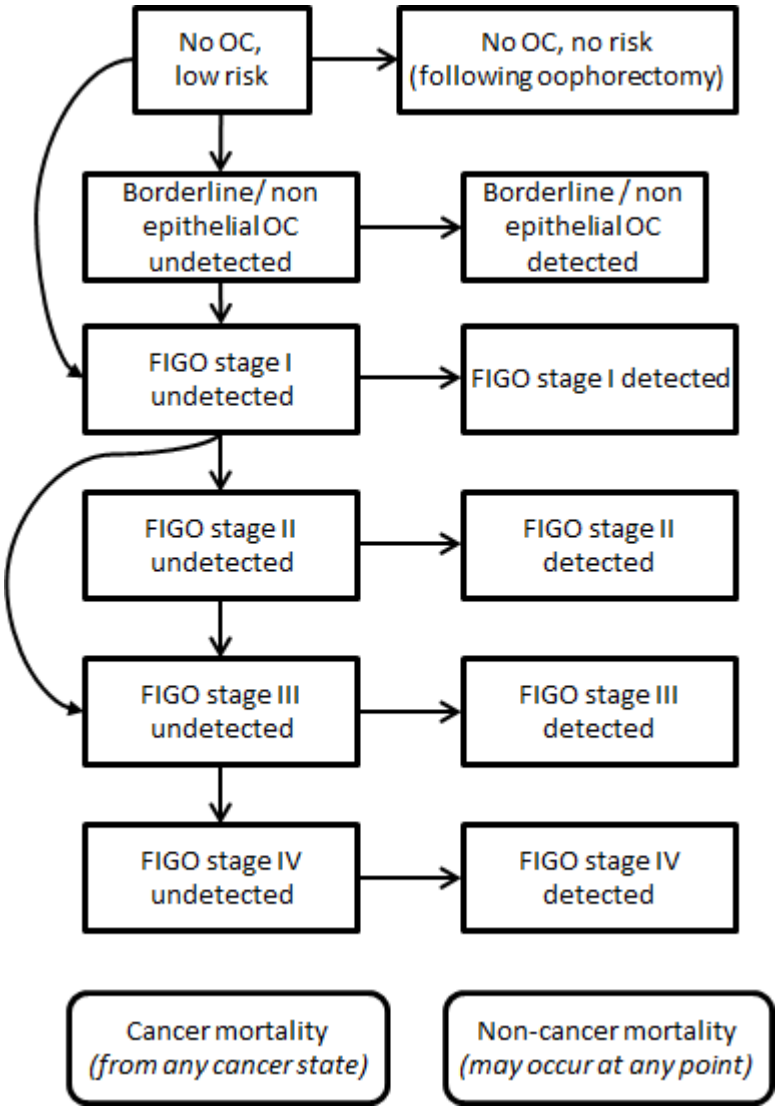
Following the initial development of cancer, undetected cancer progresses sequentially down the FIGO stages. The one exception is that from stage I a proportion of cancers may progress directly to stage III. This transition was originally used in a previous ovarian cancer screening economic evaluation[53], based on an argument of clinical plausibility (that cancers can progress directly from stage I to stage III), and is used in the conceptual model for this project following agreement with the clinical advisors for this project.

Women with undetected ovarian cancer may present with symptoms, and so have their cancer detected (diagnosed). Following diagnosis, no further change in health status was considered in the conceptual model. This was because cancer survival is measured as time since diagnosis (and any stage breakdowns presented are based on stage at diagnosis), and so any available data will implicitly include any changes in health status due to treatment, progression or recurrence.

Some women will present with symptoms suggestive of ovarian cancer, without having ovarian cancer. It was assumed that some of these women will undergo oophorectomy (surgical removal of the ovaries) and so will no longer be at risk of developing ovarian cancer. This is consistent with the assumptions of previous ovarian cancer screening economic evaluations[12,53].

Women with diagnosed ovarian cancer may die as a result of their cancer, it was assumed that of the women with undiagnosed ovarian cancer, only those with stage IV cancer could die from their cancer. However, it should be noted that (by definition) rates of cancer-specific mortality for women with undetected ovarian cancer are unknown. Women may die from causes unrelated to ovarian cancer at any point. For the conceptual model a simplifying assumption was made that transition rates between cancer stages would vary depending on the woman's existing stage of cancer, but would be independent of other characteristics (such as age, or prior time spent with cancer).

Figure 7.1: Disease-level conceptual model.



OC: Ovarian cancer.

7.1.2. Screening conceptual model.

It is important to consider both uptake and compliance (as defined and discussed in Section 3.6). Uptake refers to the probability that a woman who is invited to attend for a screen will actually attend. Compliance refers to the probability that a woman who has begun the screening process will finish it. Women who do not take-up a screening test will not be screened. Women who are initially screened but do not complete their screening test are said to receive a partial screen, whilst women who take-up their test and complete it receive a full screen.

For simplicity, women may be categorised as either having ovarian cancer or being ovarian cancer free. Similarly, the results of a screening test may either indicate that a woman does or does not have ovarian cancer (known as 'abnormal' and 'normal' results, respectively). Hence, the possible combinations of a woman's true health state and the results of a screening test may be summarised in a two-by-two (contingency) table, as shown in Table 7.1.

Table 7.1 Contingency table for the performance of an ovarian cancer screening test.

Screening result → **True disease status ↓**	**Normal**	**Abnormal**
No ovarian cancer	True negative	False positive
Ovarian cancer	False negative	True positive

These potential outcomes may be used to define a number of summary measures of the performance of a screening test. Three measures that are of particular interest are defined in Box 7.2.

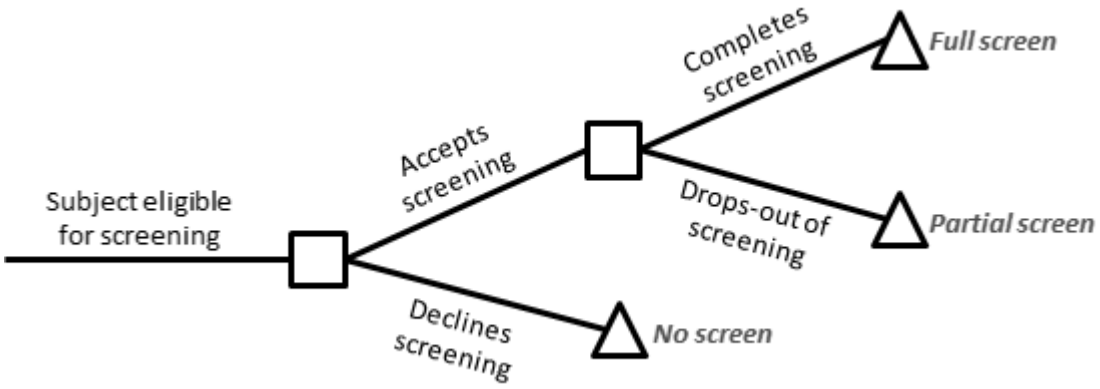
Box 7.2: Definitions of screening test performance measures

***False positive rate**:* The proportion of all positive screens that occur amongst women without ovarian cancer = Number of false positives / (Number of false positives + Number of true positives).

***Sensitivity**:* The proportion of all positive screen results that occur amongst women with ovarian cancer = Number of true positives / (Number of false positives + Number of true positives).

***Specificity**:* The proportion of all negative screen results that occur amongst women without ovarian cancer = Number of true negatives / (Number of true negatives + Number of false negatives).

Figure 7.2: Screening conceptual model.

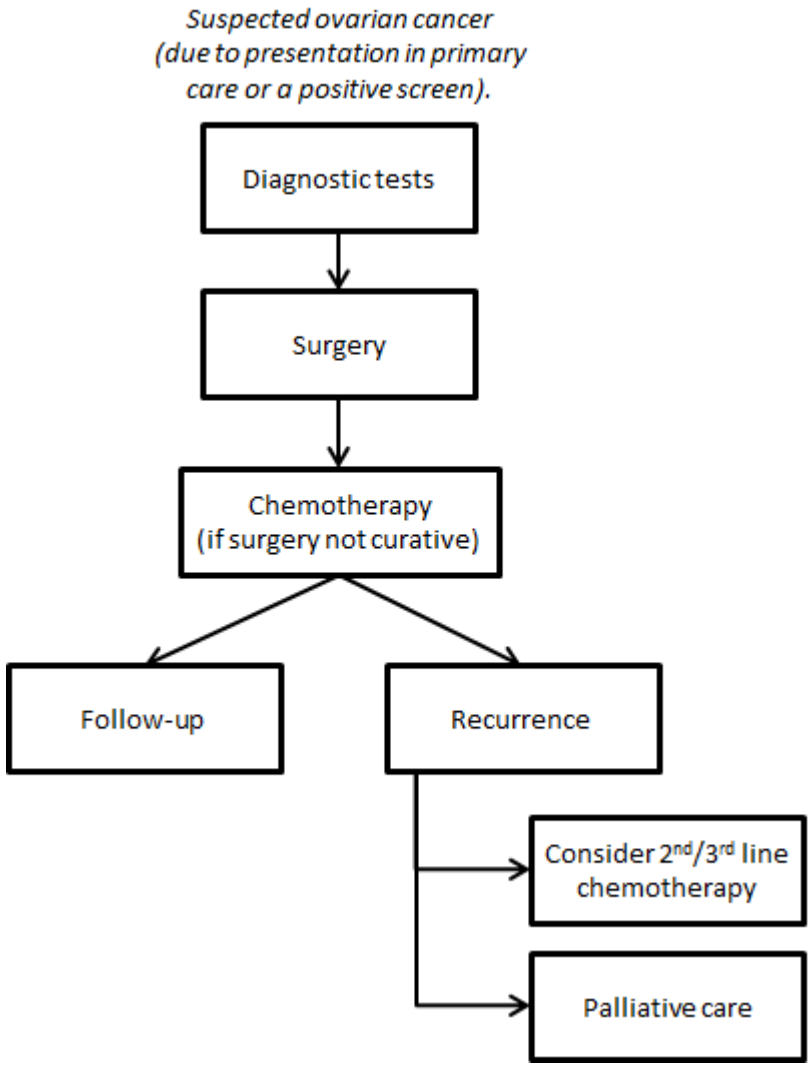


Screening has a number of potential effects on the natural history of ovarian cancer. There is likely to be an increase in both the proportion of ovarian cancers that are detected (true positives) and the proportion of patients who move from being at low-risk of developing ovarian cancer to not being at risk (false positives who receive an oophorectomy). The increased detection of ovarian cancers may lead to a decrease in ovarian cancer mortality, either due to a stage-shift or due to earlier detection within a stage. Stage shift is defined to occur when screening detects a patient's cancer at a stage that is earlier than it would have been detected in the absence of screening. This earlier diagnosis may be associated with an improved prognosis. Earlier detection within a stage may mean that the cancer has grown less, and so is more amenable to treatment. However, the surgery required to confirm (or refute) a positive screen has an associated probability of both mortality and morbidity.

7.1.3. Treatment conceptual model.

Before undertaking surgery, a diagnosis of ovarian cancer usually requires additional diagnostic tests to rule out other possible conditions that may lead to a suspicion of ovarian cancer (due to either a positive screen result, or the presence of symptoms suggestive of ovarian cancer). These are included in the initial stages of the treatment conceptual model. Treatment for ovarian cancer typically consists of surgery and / or chemotherapy. There is uncertainty about the respective roles of surgery and chemotherapy in treating ovarian cancer, with regards to if both are required and if so which should be performed first[88,89]. The optimal treatment sequence is likely to be dependent upon a number of factors including patient age and the stage of ovarian cancer at diagnosis, as detailed in section 6.2.

Figure 7.3: Treatment conceptual model.



8. Health economic evaluation of screening strategies for ovarian cancer.

This chapter provides details of the development of a model-based health economic evaluation of screening strategies for ovarian cancer.

8.1. Economic analysis scope

The main research question addressed by the economic evaluation is *"What is the cost-effectiveness of different ovarian cancer screening strategies, including no screening, amongst post-menopausal women"* The scope of the health economic analysis is summarised in Box 8.1. The patient population, comparator and interventions were all chosen to match those used in the UKCTOCS study[6].

Box 8.1: Economic analysis scope

***Population**:* Post-menopausal women aged 50 to 74 with none of the following: history of bilateral oophorectomy, active malignancy, previous history of ovarian cancer, participation in other ovarian cancer screening trials, or increased risk of familial ovarian cancer.

***Comparator and interventions**:* The comparator was no screening. Two interventions were considered: (1) annual multimodal screening with first-line CA-125 screening (interpreted using ROCA) and transvaginal ultrasound scan as the second-line test; (2) annual screening with transvaginal ultrasound as both the first and second-line test.

***Primary outcome**:* Incremental cost per quality adjusted life year (QALY) gained

Time horizon: Patients' remaining lifetime

Perspective: NHS and Personal Social Services (PSS)

Discount rate: 3.5% per year for both costs and QALYs.

Price year: 2013-2014

The population included in the health economic analysis relates to post-menopausal women aged 50 to 74 who do not have a high risk of ovarian cancer. Three potential screening strategies were evaluated: (1) no screening; (2) annual multimodal screening; (3) annual ultrasound screening. The economic evaluation takes the form of a cost-effectiveness analysis. The measure of effectiveness used is the QALY, all costs were valued at 2013-2014 prices (with values from previous years inflated using the HCHS inflation indices published by the Personal and Social Services Research Unit[84]. The

primary outcome was the incremental cost per incremental QALY gained. As screening may have an impact on mortality, a lifetime horizon was modelled to capture all the relevant differences in costs and health outcomes, which were evaluated from the perspective of the NHS and PSS over the patients' remaining lifetime. In line with current recommendations[80], all costs and effects were discounted at a rate of 3.5% per year. Uncertainty was captured via probabilistic sensitivity analyses (PSA) and scenario analyses.

8.2. Health economic model.

8.2.1. Implemented model structure

A cohort-level Markov state transition model was constructed using Microsoft Excel. This mathematical model simulates the life experience of a cohort of women from the general English population who initially do not have ovarian cancer, and who may or may not develop ovarian cancer during their lifetime.

State transition models represent a natural framework for incorporating repeated events (such as annual screens), and can represent the natural history of ovarian cancer – as described in Figure 7.1 – by simulating transitions between different health states. A state transition model is also able to reflect the on-going risk of developing ovarian cancer over time. The development of the model structure was informed by both a review of existing economic models for ovarian cancer screening (as described in Section 4) and by the prior experience of the study team[48,90].

The health economic model originally built consisted of three interlinking sub-models: the first sub-model simulated the natural history of ovarian cancer in the absence of screening. The second sub-model captured the impact of screening on this natural history. The third sub-model simulated mortality, which varied with age and had three main components; mortality may have been due to having ovarian cancer, may have resulted from the surgery involved in diagnosing and treating ovarian cancer, or may have been due to other causes (unrelated to ovarian cancer). It was envisaged that detailed effectiveness data from the UKCTOCS trial[7] would form a key evidence source for this model. Data from the control arm of the trial would have informed that natural history sub-model, whilst data from the two screening arms would have informed the screening sub-model. Bayesian calibration methods previously used by this study team[91] for constructing a natural history model for colorectal cancer (along with the history of screening on this natural history) had been set-up and tested using data provided by the English cancer registries[14]. However, it was not possible to obtain the required data within the timescales of this study. Instead, summary UKCTOCS

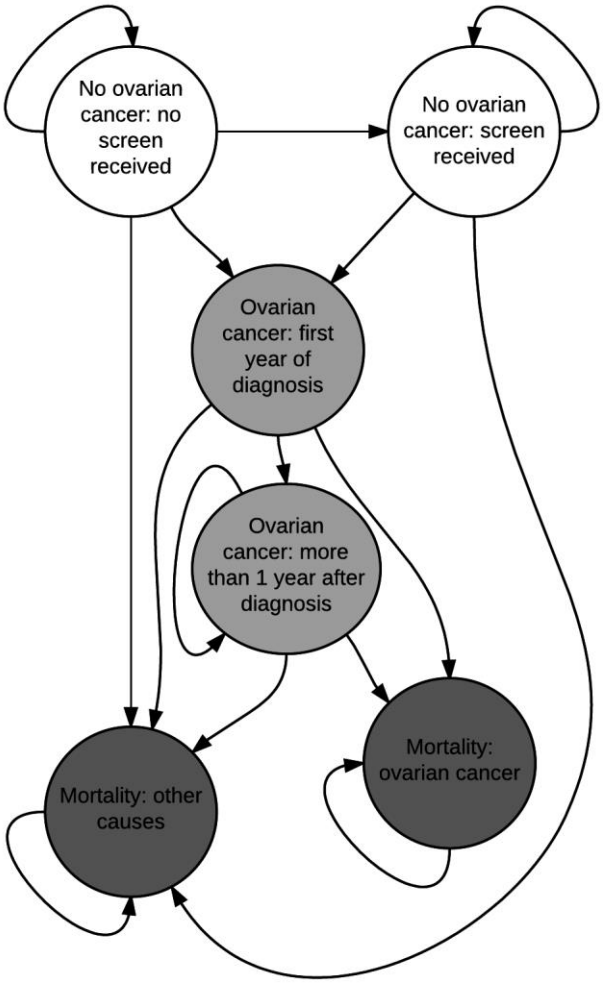
effectiveness data were obtained from the most recent publication[7]. The Kaplan-Meier curves (for both incidence of ovarian cancer and mortality) presented in this publication were digitised using EnGauge software. The digitised data, along with summary statistics from the 2016 publication were used to replicate individual patient data (IPD) in R using the method described by Guyot *et al*[92]. These IPD were used for generating estimates of in-trial effectiveness and for extrapolating these effectiveness estimates across participants' remaining lifetime. Due to the limited effectiveness data, the natural history of ovarian cancer was not modelled. Instead outcomes for the three UKCTOCS treatment arms were modelled. Whilst this approach allows for estimates of the cost-effectiveness of the two screening strategies evaluated in the UKCTOCS trial, it does not allow for further analyses examining different implementation strategies for screening. For example, screening strategies with different age-ranges or different age-intervals could be considered using a natural history modelling approach. This approach could also assess the potential impact on cost-effectiveness of improvements in screening test characteristics (such as sensitivity and specificity).

The implemented Markov model had the following six health states:

1. No ovarian cancer, no screen received.
2. No ovarian cancer, screen received.
3. Ovarian cancer: first year of diagnosis.
4. Ovarian cancer: subsequent years of diagnosis.
5. Death due to ovarian cancer.
6. Death due to non-ovarian cancer causes

An annual Markov cycle was used. It was assumed that all women initially started in the first health state (no ovarian cancer, no screen received). The last two health states (corresponding to two different causes of death) are absorbing states (once entered, women cannot leave these states). The health state 'ovarian cancer: first year of diagnosis' is a tunnel state as women can only stay in it for one model cycle. It should be noted that the first two health states implicitly include women with undiagnosed cancer. A model schematic is provided in Figure 8.1.

Figure 8.1 Schematic of the implemented health economic model.



8.3. Evidence used to inform the model parameters

8.3.1. Summary of evidence used to inform the model parameters

The evidence sources used to inform the inputs to the health economic model are summarised in Table 8.1. The implementation of this evidence within the health economic model is described in subsequent sections.

Table 8.1: Evidence used to inform the model parameters

Model parameters	Source
Cost parameters	
Resource use for treating ovarian cancer (by stage)	Cancer registry data[14], expert clinical opinion.
Cost of treating ovarian cancer (by stage)	Incisive Health Report[82], expert clinical opinion.
Costs of screening	UKCTOCS data[6], NHS reference costs[81], expert opinion.
Health-related quality of life parameters	
Health utility values: ovarian cancer and general population	Havrilesky *et al*[58]
Impact of screening on health-related quality of life.	Barret *et al*[35], Reade *et al*[5].
Clinical effectiveness	
Incidence of ovarian cancer for three UKCTOCS screening arms	Reconstructed individual patient data from UKCTOCS[7,92]
Incidence of ovarian cancer for three UKCTOCS screening arms	Reconstructed individual patient data from UKCTOCS[7,92]
Mortality due to ovarian cancer	UKCTOCS[7] data
Other cause mortality	UKCTOCS[7], Life table.
Uptake and compliance with screening	UKCTOCS[7] data

8.3.2. Costs

There were four main types of cost that were included in the health economic model. These were the costs of screening, diagnosis, treatment and end of life care. The derivations of costs for these are described in Section 6. However, these estimates did not contain any indication of uncertainty. In the absence of more relevant data, variations in cost from NHS reference costs were used as an indication of uncertainty. These reference costs contain the codes MA06 "Major, Open or Laparoscopic, Upper or Lower Genital Tract Procedures for Malignancy" and MA26 "Complex, Open or Laparoscopic, Upper or Lower Genital Tract Procedures for Malignancy". The average costs for these two procedures are £3,742 and £5,386 with estimated standard errors of £1,074 and £1,172 respectively. It was assumed that the higher of these two standard errors (£1,172) could be used as the standard error for treatment costs. For diagnostics and end of life care, a standard error of 10% of the mean was assumed. The treatment cost data used within the PSA are summarised in Table 8.2.

Table 8.2: Ovarian cancer diagnosis and treatment costs.

Treatment type, by stage	Mean	Standard error	95% Confidence interval	Gamma parameters* (alpha, beta)	
Diagnostics					
Borderline / No OC	£110	£11	£132 to £88	100.00	1.12
Stage 1	£116	£11	£138 to £94	100.00	1.10
Stage 2	£126	£12	£150 to £102	100.00	1.16
Stage 3	£126	£13	£151 to £101	100.00	1.26
Stage 4	£112	£13	£137 to £87	100.00	1.26
Treatment					
Borderline / No OC	£3,000	£1,172	£5,297 to £703	6.55	457.87
Stage 1	£6,961	£1,172	£9,258 to £4,664	35.27	197.33
Stage 2	£7,325	£1,172	£9,622 to £5,028	39.07	187.50
Stage 3	£9,016	£1,172	£11,313 to £6,719	59.18	152.34
Stage 4	£5,892	£1,172	£8,189 to £3,595	25.28	233.10
End of life care	£7,080	£708	£8,468 to £5,692	100.00	70.80

OC: ovarian cancer *Used in probabilistic sensitivity analysis

Two different screening regimens were considered in the economic evaluation. These were the multimodal screening (MMS) and transvaginal ultrasound screening (USS) regimens employed in the UKCTOCS trial. Each regimen had two levels of screening, for MMS level 1 screens involved one or more CA-125 blood tests, interpreted using the patented ROCA algorithm. The MMS level 2 screen involved a transvaginal ultrasound performed by a type 2 sonographer. For USS level 1 screens involved a transvaginal ultrasound performed by a type 1 sonographer, whilst level 2 screens involved a type 2 sonographer. The derivations of costs for these (CA-125 blood test and transvaginal ultrasound performed by either a type 1 or type 2 sonographer) are detailed in section 6. Further calculations were required to adjust for the numbers requiring level 1 and level 2 screens (including repeat screens) and non-compliance. Data for these were taken from the prevalence screen of the UKCTOCS. Full details of the calculations are provided in Appendix 2. There are currently no indications of a cost arising from use of the ROCA that was developed with support from public and charitable monies, it was assumed that using the ROCA would not lead to any increase in the cost of a screen within the UK setting. Uncertainty in the cost of an ultrasound was taken from NHS reference costs. Uncertainty in the cost of CA-125 was represented by using beta distributions and assuming an arbitrarily sample size of five. For CA-125 a range of £25 to £75 was assumed (as this range includes the high and low CA-125 estimates derived in Section 6). These details are presented in Table 8.3.

Table 8.3: Screening costs and distributions used in PSA.

Ultrasound	Mean	Standard error	Gamma (alpha, beta)	
Level one	£55.82	£16.36	11.64	4.80
Level two	£67.55	£16.36	17.04	3.96
Other	**Mean (Range)**	**Sample size**	**Beta (alpha, beta)**	
CA-125 test	£54.19 (£25 to £75)	5	2.919	2.081

8.3.3. Health-related quality of life

The review of health-related quality of life in ovarian cancer is described in detail in Section 5. This review only identified three articles of potential relevance to the health economic model. These three articles and their findings are briefly re-summarised.

Hess *et al*[76] reported the results of a systematic review of the health related quality of life of women with ovarian cancer. The review was conducted up to the end of December 2011. A total of 139 studies were identified, which included both randomised and non-randomised trials, along with observational studies. Over 90 different HRQoL (or patient-reported outcomes) instruments were used. There was no evidence that HRQoL varied by the treatment strategy received. This lack of evidence was due to both a paucity of available data, and heterogeneity amongst the data that was available, with regards to the interventions, instruments and populations considered. Pooled analyses of changes in HRQoL during treatment were possible for three instruments (FACT-G, FACT-O and QLC-30). For newly diagnosed disease, results from the QLC-30 showed a statistically significant ($P < 0.001$) improvement during treatment, but no such association was identified by either of the FACT instruments. However, a statistically significant improvement in HRQoL by the completion of treatment was identified for all three instruments ($p < 0.001$ for all).

The study by Havrilesky *et al*[58] was the only identified study that both reported HRQoL values by severity of cancer and was deemed to have enough validity (external and internal) to be of relevance to the study. The authors also presented HRQoL for selected screening outcomes. Estimates were derived using both the time trade-off method and visual analogue scores. However, results for the latter have been noted to have inherent biases, so only the time trade off values are reported here. The estimates reported here were based on a sample of 37 women without ovarian cancer. The

estimated time trade-off utilities for screening health states are presented in Table 8.4, along with those for selected diagnosis-related health states. The authors also reported estimates for recurrent ovarian cancer and for chemotherapy-related side effects. These are not considered here as there is no evidence to suggest that screen-detection of cancers would lead to either a differential rate of recurrence or to different treatment strategies.

Table 8.4: Time-trade off utility values for ovarian cancer screening and diagnosis related health states from Havrilesky *et al*[58].

Health state	N	Median	Range	Mean	SD
Screening					
Blood test	15	0.97	0.33 to 0.97	0.90	0.18
Transvaginal ultrasound	15	0.97	0.03 to 0.97	0.83	0.27
False-positive blood test, followed by a negative transvaginal ultrasound	16	0.97	0.03 to 0.97	0.88	0.26
False-positive transvaginal ultrasound test, followed by removal of an ovary by laparoscopy.	15	0.97	0.50 to 0.97	0.90	0.14
Diagnosis					
Ovarian cancer – clinical remission	16	0.95	0.03 to 0.97	0.83	0.25
Early ovarian cancer – newly diagnosed	16	0.93	0.03 to 0.97	0.81	0.26
Advanced ovarian cancer – newly diagnosed	14	0.50	0.03 to 0.93	0.55	0.29
Newly diagnosed ovarian cancer – chemotherapy grades 1 to 2 toxicity	16	0.67	0.03 to 0.97	0.60	0.25
Newly diagnosed ovarian cancer – chemotherapy grades 3 to 4 toxicity	15	0.50	0.03 to 0.97	0.49	0.36
End stage ovarian cancer	15	0.03	0.03 to 0.83	0.16	0.25

SD: Standard deviation

In the author's health state descriptions, early ovarian cancer was defined as having not spread out of the ovary, whilst advanced ovarian cancer was defined as having spread to other organs in the abdomen. Newly diagnosed ovarian cancer was not defined by its location. With the exception of end stage ovarian cancer, all of these health states include the impact of surgery (laparotomy) to remove the uterus, tubes, ovaries and lymph nodes. A limitation of this study is that no values are presented for the general population.

Barrett *et al*[35] used longitudinal data from the UKCTOCS to investigate the impact of screening on HRQoL. The authors noted that screening on its own was not associated with an increase in anxiety.

These results were consistent with those of a previous meta-analysis of screening trials which also found no evidence for an impact of screening on HRQoL[5].

Based on the results of these three studies, the following assumptions regarding HRQoL were used in the health economic model:

1. There is no impact on HRQoL due to receiving a screen for ovarian cancer.
2. There is no impact on HRQoL due to receiving a positive screen result (although see 7).
3. There is no impact on HRQoL due to having ovarian cancer prior to treatment.
4. There is an impact of treatment on HRQoL, which varies by stage at diagnosis.
5. The impact of treatment on HRQoL only lasts for the duration of treatment (which is assumed to be one year).
6. Following treatment women have the same HRQoL as the general population, irrespective of initial stage at diagnosis.
7. Women receiving surgery for a false-positive result experience the same disutility as women diagnosed with stage I ovarian cancer.

The impact of these assumptions on the model results was tested in sensitivity analyses.

To derive the utility values used in the health economic model, the mean time trade-off values reported by Havrilesky *et al*[58] were used, along with the following assumptions/steps:

1. The value reported for blood tests (0.90) represents the utility of women without ovarian cancer.
2. The value for early ovarian cancer (0.81) represents the utility of women diagnosed with stage 1 ovarian cancer who do not receive chemotherapy, but may or may not receive surgery.
3. The value for advanced ovarian cancer (0.55) represents the utility of women diagnosed with stage 4 ovarian cancer who do not receive chemotherapy, but may or may not receive surgery.
4. The values for women diagnosed with stages 2 or 3 ovarian cancer can be linearly interpolated from the values for stages 1 and 4. This gives values of (0.81 - 1/3 * [0.81 – 0.55] =) 0.723 for stage 2 and (0.81 - 2/3 * [0.81 – 0.55] =) 0.6373 for stage 3.

5. The impact of chemotherapy on HRQoL is as an additive disutility, and may be estimated as the difference between the utility for early ovarian cancer (0.81) and newly diagnosed ovarian cancer – chemotherapy grades 1 to 2 toxicity (0.60), giving a disutility of 0.21

To derive stage-specific treatment utilities, the values estimated at steps 1 to 4 were down-weighted by the chemotherapy disutility multiplied by the proportion of women who receive chemotherapy (based on data supplied by the Cancer Registries, as displayed in Table 8.5).

Table 8.5 Utility values used in the economic model for women with ovarian cancer undergoing treatment.

Stage at diagnosis	Utility, no chemotherapy	% receiving chemotherapy	Utility
Stage I	0.810	53%	0.700
Stage II	0.723	71%	0.575
Stage III	0.637	71%	0.487
Stage IV	0.550	50%	0.445
Disutility to chemotherapy:		0.210	

Uncertainty in these values was modelled using beta distributions (restricted to the range 0 to 1). Within the PSA, logical constraints were used so that more increasing stage at diagnosis was never associated with an improved HRQoL. To incorporate this, the utilities reported in Table 8.5 were converted into disutilities relative to the previous stage (for Stage 1 this was relative to women with no ovarian cancer). As an example, the disutility for Stage 1 relative to no ovarian cancer was (0.9 – 0.7 =) 0.2. The uncertainty around this disutility was modelled as a beta distribution. For women with no ovarian cancer, the observed mean (0.9) and standard deviation (0.18) were used to derive alpha and beta parameters. For the disutilities, it was assumed that the evidence from Havrilesky *et al* was equal to evidence obtained from a sample of size 14 (as this is the smallest reported sample size in Table 8.4). This was used to calculate the alpha parameter by multiplying 14 by the disutility value (so for stage 1 the alpha parameter = 0.200*14 = 2.807). The beta parameter was derived as 14 minus the beta parameter (so for stage 1 the beta parameter = 14-2.807 = 11.193).These values are summarised in Table 8.6.

Table 8.6 Utility values used in the economic model for women with ovarian cancer undergoing treatment.

Health state	Utility	Disutility*	Beta parameters (alpha, beta)
No ovarian cancer	0.900		1.600, 0.178
Stage I		0.200	2.807, 11.193
Stage II		0.124	1.741, 12.259
Stage III		0.088	1.228, 12.772
Stage IV		0.042	0.591, 13.409

*Disutility is relative to the previous health state.

8.4. Clinical effectiveness

8.4.1. Estimating within-trial clinical effectiveness data.

In the absence of detailed individual patient data (IPD) from the UKCTOCS trial, estimates of clinical effectiveness were derived from summary statistics and graphics made publicly available in the UKCTOCS publication[7]. These data included the incidence of ovarian cancer by trial screening arm (including no screening) and mortality due to ovarian cancer by trial screening arm. These measures of effectiveness were converted into IPD by first digitising the presented Kaplan-Meier curves (Web figure 9 and Figure 2B respectively) using EnGauge software. The digitised data, along with summary statistics from the 2016 publication[7] were used to replicate IPD in R using the method described by Guyot *et al*[92]. These IPD were used for generating estimates of in-trial effectiveness and for extrapolating these effectiveness estimates across participants' remaining lifetime. The UKCTOCS publication included an analysis which removed prevalent cancers when estimating effectiveness. However, it is noted that such an analysis is unlikely to reflect the real-world effectiveness of ovarian cancer screening, and so is not considered for this study.

Other-cause mortality was taken from the UKCTOCS publication (Appendix Web Table 6)[7], which presented the number of deaths and women-years by screening arm. As there was no evidence that other-cause mortality varied with screening arm, the evidence was pooled to obtain an overall annual value of 0.61% (13,296 deaths from 2,194,447 women years). The arm-specific annual values were 0.61% for no screening (6,658 deaths from 1,097,089 women years), 0.62% for MMS (3,376 deaths from 548,533 women years), and 0.59% for USS (3,262 deaths from 548,825 women years).

The recreated IPD cover the time-horizon of the UKCTOCS trial (median follow-up 11.1 years). However, for the economic evaluation a lifetime horizon was required. Hence to generate estimates

of lifetime effectiveness extrapolation was required. Separate extrapolation methods were employed for the incidence and mortality data, which are discussed in turn.

8.4.2. Modelling and extrapolation of incidence data.

Follow-up in the UKCTOCS trial was until l 31st December 2014, three years after screening finished. In addition, published results showed no evidence of over-diagnosis due to active screening[7]. Because of this, it was not anticipated that there would be any difference in overall rates of ovarian cancer incidence during the extrapolation period (there may have been differences in the age-specific incidence rates due to earlier diagnosis amongst the two active screening arms, but there was insufficient evidence available to model this). Hence to model and extrapolate the incidence data, the following approach was taken:

- Fit arm-specific parametric models to the observed trial data.
- Use the arm-specific fitted parametric models to generate estimates of incidence for the first eleven years of the model.
- For subsequent years use the parametric model for 'no screening' to generate estimates of incidence for all three screening arms.

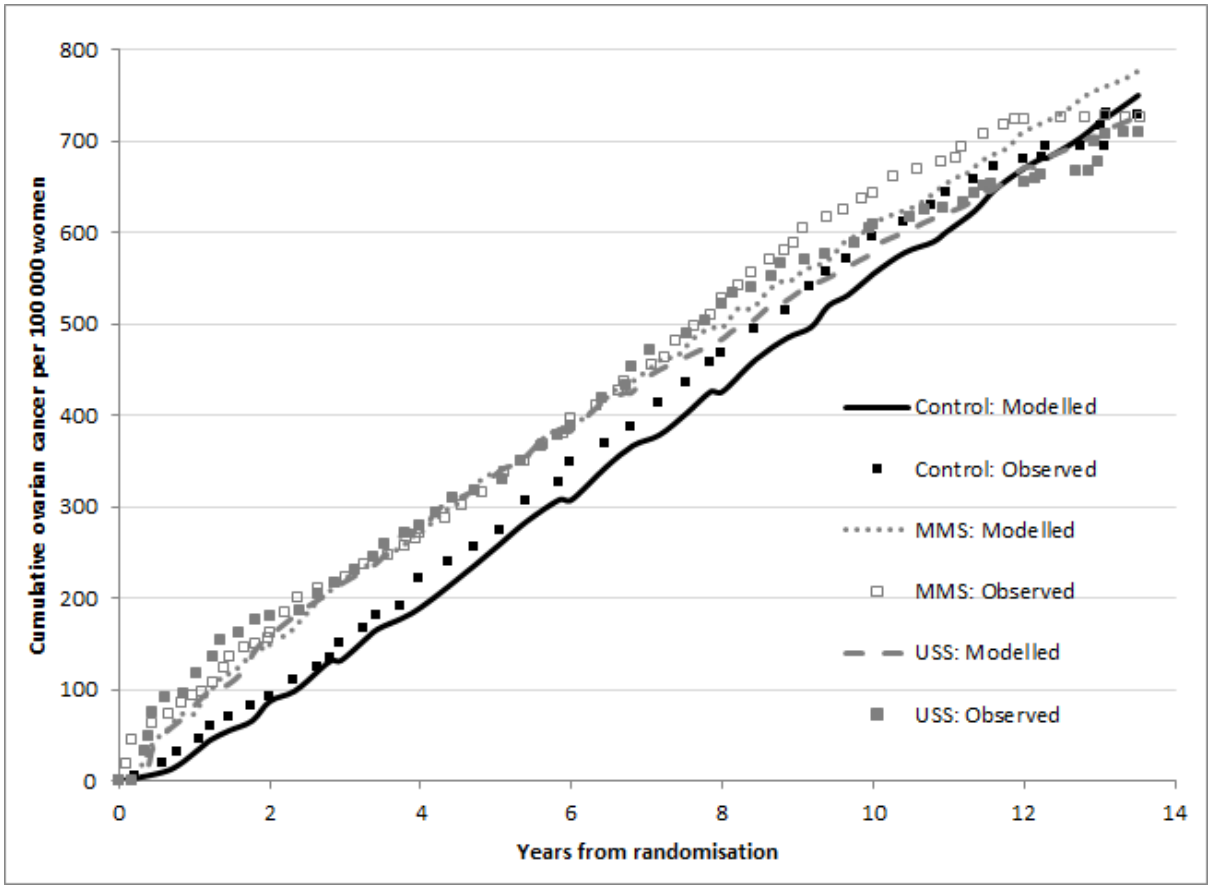
Five potential standard parametric model types were considered (exponential, Weibull, Gompertz, log-logistic and log-Normal). The choice between these was based on minimising the Bayesian Information Criteria (BIC), although differences in the Akaike's information criteria (AIC) were also noted. The information criteria are displayed in Table 8.7 (with minimum values in bold), and show that the best fitting survival models to the observed data were the Log-Normal for no screening and USS, and log-logistic for MMS. Use of either the AIC or the BIC led to the same conclusions. The fit to the observed data for these models is displayed in Figure 8.2.

Table 8.7 Information criteria (Bayesian and Akaike's) for incidence across the three screening strategies.

Incidence	No screening		Multimodal screening		Ultrasound screening	
	BIC	AIC	BIC	AIC	BIC	AIC
Exponential	8,584.35	8,574.82	4,805.55	4,796.72	4,590.44	4,581.61
Weibull	8,588.06	8,569.01	4,805.05	4,787.39	4,586.89	4,569.22
Gompertz	8,595.80	8,576.75	4,806.45	4,788.79	4,583.40	4,565.74
Log-logistic	8,587.83	8,568.78	**4,805.03**	**4,787.37**	4,586.82	4,569.16
Log-Normal	**8,573.91**	**8,554.86**	4,805.52	4,787.86	**4,580.64**	**4,562.97**

AIC: Akaike's information criteria. BIC: Bayesian information criteria.

Figure 8.2: Ovarian cancer incidence: comparison of model estimates and observed data.



8.4.3. Modelling and extrapolation of mortality data.

In the UKCTOCS publication, the effects of active screening (with either MMS or USS) on ovarian cancer mortality were modelled as hazard ratios relative to ovarian cancer mortality under no screening[7]. The authors noted that the hazard ratios varied over time (so were not proportional),

and modelled these using Royston-Parmar (R-P) models[93]. These models fit restricted cubic splines to the observed data, resulting in flexible parametric survival models. For consistency with the UKCTOCS mortality analysis, R-P models were used to model the within-trial probability of mortality for each of the three screening arms. An R-P model was used to estimate the hazard function for no screening; this was used to derive annual time-varying transition probabilities of ovarian cancer mortality for the first 11 years for no screening. Separate R-P models were then fitted for MMS and USS, and used to estimate time-varying hazard ratios. These were applied to the hazard function for no screening, and used to estimate transition probabilities for use in the economic model.

Time-series methods[94,95] were used to extrapolate the mortality effectiveness data. Time-series methods have the explicit aim of producing extrapolations (predictions of the future), and are designed to incorporate the additional uncertainty that arises due to predicting future values. This forecasting uncertainty is in addition to parametric uncertainty that arises when parametric models are used to characterise observed data. Further details of the extrapolation methods are provided in Appendix 3.

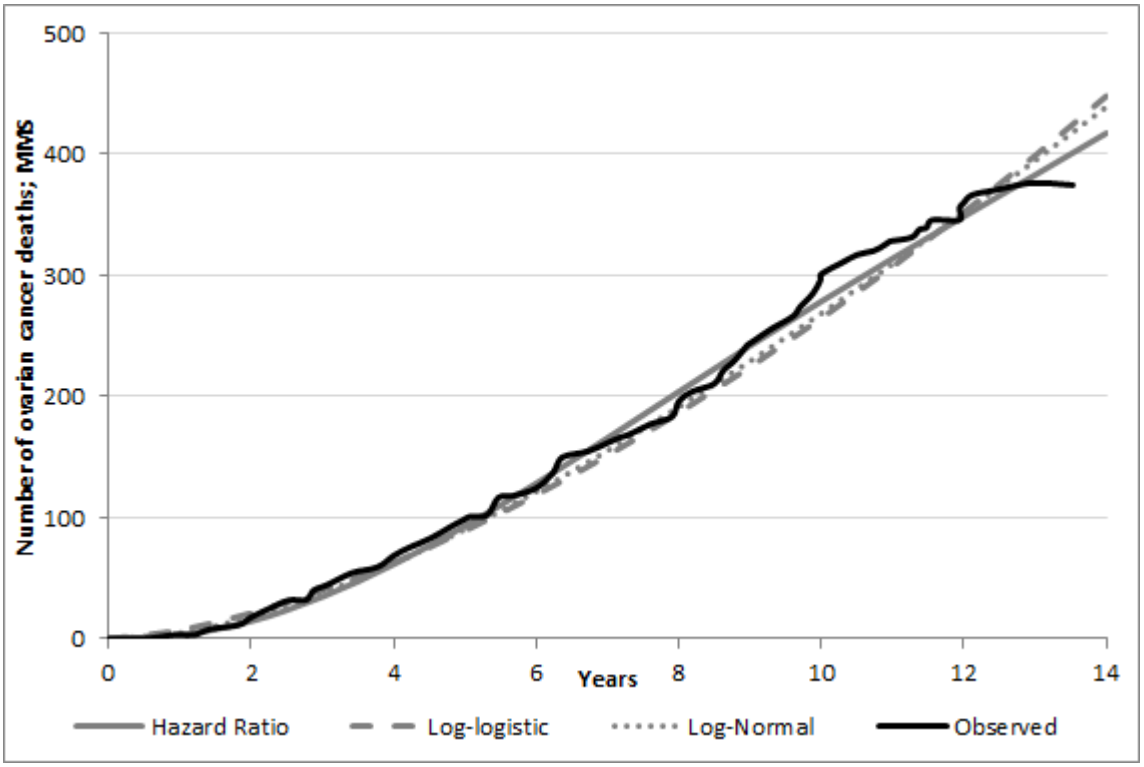
In addition to the use of hazard ratios, two other potential approaches were identified for the modelling and extrapolation of mortality data. These approaches both involve the use of standard parametric models. One approach is to fit arm-specific parametric models. This is the same approach as described for the incidence data in Section 8.4.2, with the same five potential models considered, and the choice based on minimising an information criteria (either AIC or BIC). The fitted models may then be used directly for extrapolation. Whilst standard parametric models have some flexibility to model time-varying effects, they are not as flexible as R-P models. Hence, when used for extrapolation, estimates of long-term effectiveness may be influenced by the hazard function imposed by the model. When comparing different screening arms, the use of arm-specific models may result in different estimates of long-term effectiveness due to different model structures. Because of this, it may be preferable to fit the same model structure to all three screening arms. These considerations led to two approaches: one using separate standard parametric models for each screening arm; and one using the same parametric model. In the latter case the choice of model type was based on the combined information criteria across all three arms. It should also be noted that when used for extrapolation estimates of uncertainty from standard parametric models do not include forecasting uncertainty.

To summarise, three different extrapolation methods were considered. These are discussed in more detail in Appendix 3, and were:

1) Use of R-P models to estimate time-varying hazard ratios for each of the two active screening arms compared to the no screening arm (for which a time-varying hazard was estimated). Use of time-series methods to extrapolate these hazard ratios.
2) Use of standard parametric models, with the same model structure used for all three screening arms.
3) Use of standard parametric models, with separate model structures allowed for the three screening arms.

The first of these three methods was used for the base-case. However, there is no guidance on which approach is more suitable for extrapolation and each of the three approaches has merits and limitations. Hence the other three methods were considered in sensitivity analyses. The three methods generally produced similar fits to the observed data, but differed with respect to their long-term estimates of effectiveness. For illustrative purposes, for MMS fits to the observed data are displayed in Figure 8.3, with the resulting extrapolations displayed in Figure 8.4. Graphs for USS and no screening are displayed in Appendix 3 (for MMS the best-fitting parametric model is the log-Normal, which is also the model used for method 3. Hence for illustrative purposes the second-best fitting model – the log-logistic - is also displayed. The difference in AIC and BIC between the log-Normal and log-logistic is 0.16%). Information criteria values for both the standard parametric models and the R-P models are displayed in Table 8.8, with minimum values in bold. Two R-P models were fitted: one to MMS and no screening and one to USS and no screening: information criteria are displayed for both; for comparison combined information criteria for the standard parametric models are also displayed.

Figures 8.3 and 8.4: Comparison of model estimates and observed data for mortality: for the trial period (Figure 8.3, top), and for a lifetime horizon (Figure 8.4, bottom); multimodal screening.



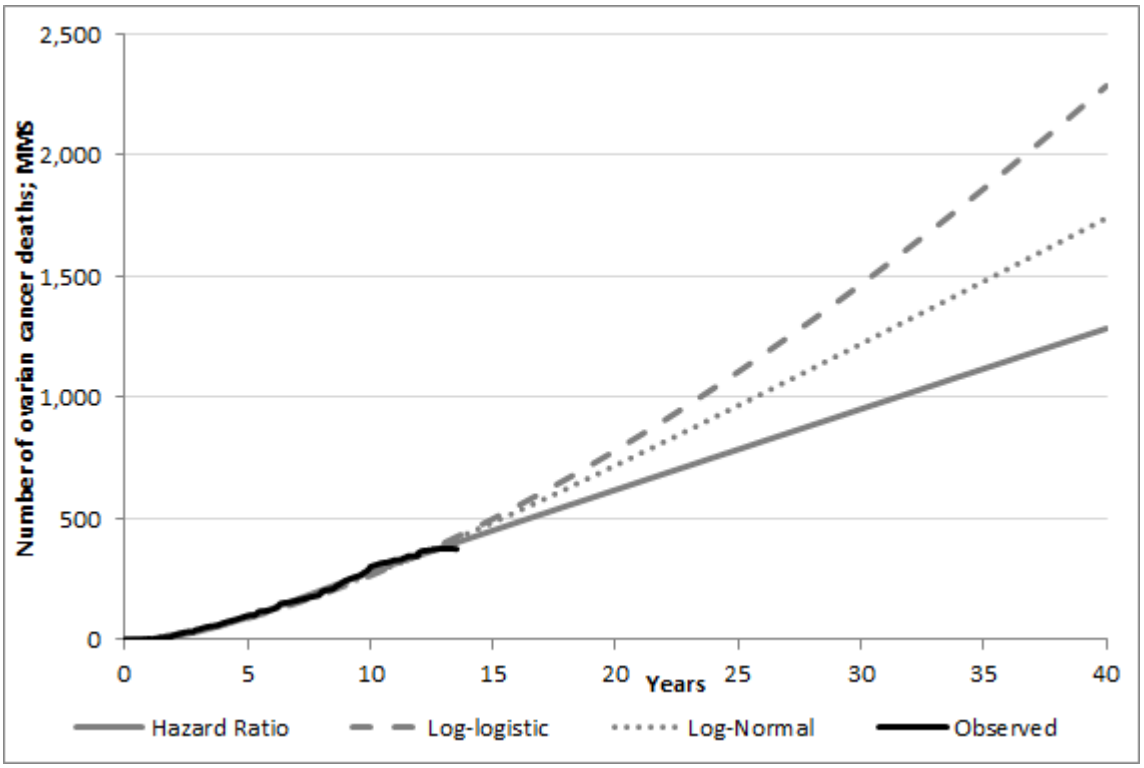


Table 8.8 **Information criteria (Bayesian and Akaike's) for mortality across the three screening strategies.**

Mortality	**No screening**		**Multimodal screening**		**Ultrasound screening**	
	BIC	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**
Exponential	5,314.01	5,304.49	2,326.05	2,317.22	2,375.51	2,366.68
Weibull	**5,187.21**	**5,168.16**	2,307.81	2,290.15	2,355.30	2,337.63
Gompertz	5,197.81	5,178.76	2,320.63	2,302.96	2,365.09	2,347.43
Log-logistic	5,187.26	5,168.21	2,307.78	2,290.11	2,355.28	2,337.61
Log-Normal	5,190.95	5,171.90	**2,304.02**	**2,286.35**	**2,354.03**	**2,336.36**
Combined values	**Multimodal and no screening***		**Ultrasound and no screening***		**All three combined**	
	BIC	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**
Exponential	7,640.06	7,621.70	7,689.52	7,671.16	10,015.57	9,988.38
Weibull	7,495.03	7,458.31	**7,542.51**	**7,505.79**	9,850.32	9,795.94
Gompertz	7,518.43	7,481.72	7,562.90	7,526.18	9,883.52	9,829.14
Log-logistic	7,495.04	7,458.32	7,542.54	7,505.82	9,850.32	9,795.94
Log-Normal	**7,494.97**	**7,458.25**	7,544.98	7,508.26	**9,848.99**	**9,794.61**
Royston-Parmar models	**Multimodal and no screening**		**Ultrasound and no screening**			
	BIC	**AIC**	**BIC**	**AIC**		
	7,532.19	7,452.74	7583.32	7503.87		

AIC: Akaike's information criteria. BIC: Bayesian information criteria. *For comparison with Royston-Parmaer models.

Based on minimising the information criteria, the best fitting survival models to the observed data were the Weibull for no screening and the log-Normal for MMS and USS. When using the same model structure for all three screening arms, the log-Normal had the best fit. Use of either the AIC or the BIC led to the same conclusions. A comparison between the R-P models and the standard parametric models showed that both resulted in similar fits to the observed data. The R-P models had lower AIC values but higher BIC values than the standard parametric models. Both information criteria measures penalise goodness of fit by model complexity (to avoid over-fitting); the BIC applies a stronger penalty than the AIC. R-P models require the estimation of more parameters than standard parametric models (and hence are more complex), which explains the differences in interpretation between AIC and BIC values. As it is not possible to assess differences in information criteria for statistical significance, it is likely that the differing interpretations indicate no difference in the goodness of fit of the different modelling approaches to the observed data.

Other cause mortality was extrapolated using National Life Tables[25], and assuming that women would be on average 60 when they enter the model (so that after the within-trial period has ended, women would be on average 72 years).

8.5. Model evaluation methods

Central estimates of cost-effectiveness were obtained by running the health economic model probabilistically over 5,000 Monte Carlo samples. For each screening strategy estimates of lifetime cost incurred and lifetime QALYs gained were obtained. The incremental cost-effectiveness of alternative screening strategies was evaluated using standard cost-effectiveness decision rules[96]. Confidence intervals around the ICER were percentile-based. Uncertainty surrounding the incremental costs and health outcomes was represented using cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs). In addition, univariate sensitivity analysis was undertaken to examine the impact of individual parameters on the results of the economic analysis.

Alongside the cost-effectiveness analysis, value of information methods were used to estimate the value of further research. The population expected value of perfect information (EVPI), was estimated using standard methods[96]. The EVPI provides an estimate of how much should be spent to eliminate uncertainty in the cost-effectiveness results (also known as decision uncertainty). The EVPI for individual parameters (known as the expected value of perfect parameter information; EVPPI) was estimated using recent methods reported by Strong *et al*[97], available online using the online Sheffield Accelerated Value of Information application (http://savi.shef.ac.uk/SAVI/). When calculating population EVPI, it was assumed that the annual number of women affected by the decision about whether or not to fund screening was 7,121,000. This value is the Office for National Statistics Mid-2010 Population Estimates for women aged 50 to 74[98]. This value was then adjusted down by the uptake to UKCTOCS and subsequent compliance with screening (see Section 3.6) to give a value of £1,045,914. A time horizon for the decision of one year was used, as the population value implicitly covers all of the relevant years.

8.6. One-way sensitivity analyses

To establish which model inputs and assumptions had the greatest impact on cost-effectiveness results a number of sensitivity analyses were performed. These are described below, categorised by the type of uncertainty that they are assessing.

8.6.1. Structural uncertainties.

- Use of alternative extrapolation methods.

As discussed in Section 8.4.3, three different extrapolation methods were considered. These were:

1) Use of R-P models to estimate time-varying hazard ratios for each of the two active screening arms compared to the no screening arm (for which a time-varying hazard was estimated). Use of time-series methods to extrapolate these hazard ratios.
2) Use of standard parametric models, with the same model structure used for all three screening arms.
3) Use of standard parametric models, with separate model structures allowed for the three screening arms.

The first of these three methods was used for the base-case. However, there is no guidance on which approach is more suitable for extrapolation and each of the three approaches has merits and limitations. In addition, the model discrepancy approach was also considered. Under the model discrepancy approach a bias is introduced to shrink the estimated extrapolated hazard ratios towards one (which indicates no effect of screening). The bias is cumulative, with effects shrunk by an average of amount for the first year of extrapolation, and an additional amount for each subsequent year. The model discrepancies explored ranged from 1% to 5% (with respective average shrinkages of 27% and 235% by year 40).

8.6.2. Costs

- Treatment costs monotonically increase with stage at diagnosis.

In the base-case analysis, costs monotonically increase with age, with the exception of women diagnosed with stage IV ovarian cancers, which has the lowest treatment costs. This reduction is primarily due to less women receiving treatment when diagnosed at this stage. This can result in early diagnosis leading to increased treatment costs, if a lower proportion of stage IV cancers are

diagnosed. To test the impact of this, costs were changed to always monotonically increase, by 10% of the previous stage's cost (the costs for Stage I were the same as the base-case).

- Use Birmingham diagnosis and treatment cost estimates.

These are alternative cost estimates to the Sheffield estimates used in the base-case.

- Alternative estimates of MMS

As discussed in Section 6.1, a mean test cost of £54.19 was used for MMS (which covered the cost of both the blood test and processing CA-125). Two sensitivity analyses considered alternative 'high' and 'low' costs of £67.33 and £31.19.

8.6.3. Utilities

- Include an end-of-life disutility (of 0.4 over a year) for women who die from ovarian cancer.

Women who die from ovarian cancer may experience a worsening of health-related quality of life in the year before they die. However, there is no evidence in the literature to suggest if this occurs, and if so what magnitude of disutility is experienced. Hence an arbitrary disutility of 0.4 for one year was modelled.

- Earlier treatment leads to reduced utility.

Compared to a strategy of no screening, women identified via screening are more likely to be diagnosed whilst asymptomatic. It is unclear if treatment of asymptomatic women results in different HRQoL compared with treatment of symptomatic women. This question has not been addressed in existing screening trials, and in the base-case it was assumed that there was no utility loss due to the treatment of screen-detected cancers. However, a UK-based randomised trial (MRC OV05/EORTC 55955) compared early versus delayed treatment of relapsed ovarian cancer[99]. Early treatment was on the basis of increasing CA-125 levels, whilst delayed treatment occurred when the woman presented with symptoms. HRQoL was measured using the EORTC QLQ-C30 questionnaire[100]. The authors found no difference in overall survival, but noted that women randomised to early treatment experienced a shorter time from randomisation to the first deterioration in global health score or death (median 3.2 months versus 5.8 months), with a similar deterioration for most of the QLQ-C30 sub-scales. Whilst this finding, from women with relapsed ovarian cancer, may not be generalisable to women diagnosed via screening, the effect of a disutility of 0.2 for the first year following diagnosis was evaluated.

8.7. Model verification and validation

Best-practice was followed in order to validate the model and verify that it was credible[101]. The developed conceptual models were checked to ensure that they were sufficient to meet the objectives of the health economic evaluation. The structure and implementation of the health economic model were checked for any errors or lack of credibility. This included checking that any logical relationships between model inputs (such as correlations or monotonic relationships) were preserved, that sampled input values did not violate boundary constraints (such as sampling negative survival times), and assessing the performance of the model under extreme input values. Any errors or omissions identified as a result of the validation checks were rectified, and the validations checks repeated, leading to a process of iterative model improvement.

9. Health economic results

9.1. Base-case cost-effectiveness results.

Base-case lifetime cost-effectiveness results for the three screening strategies are presented in Table 9.1, sorted by ascending cost. Both probabilistic and deterministic results are presented, although probabilistic results are preferred, as they incorporate any non-linear dependencies between model inputs and outputs. Table 9.1 also includes deterministic cost-effectiveness results for a time-horizon of 11 years. This allowed for an estimate of the cost-effectiveness of ovarian cancer screening based on the observed time horizon of the UKCTOCS trial (which had a median of 11.1 years follow-up). A breakdown of costs (based on deterministic results) is presented in Table 9.2

Table 9.1 Base-case cost-effectiveness results: lifetime average costs and QALYs per woman.

Lifetime horizon	Costs	QALYs	Life years	Incremental cost-effectiveness ratio
Probabilistic results				
No screening	£179	14.290	24.660	Reference
Multimodal screening	£598	14.357	24.803	£8,864 vs no screening
Ultrasound screening	£824	14.297	24.729	Dominated by multimodal screening
Deterministic results				
No screening	£176	14.281	24.665	Reference
Multimodal screening	£593	14.330	24.813	£8,459 vs no screening
Ultrasound screening	£811	14.291	24.741	Dominated by multimodal screening
11-year horizon				
Deterministic results				
No screening	£58	8.243	11.093	Reference
Multimodal screening	£510	8.240	11.093	Dominated by no screening
Ultrasound screening	£611	8.232	11.093	Dominated by no screening

QALYs: quality-adjusted life years. Costs and QALYs are discounted, life years are not.

The results displayed in Table 9.1 show that, based on lifetime results, both of the screening strategies are more effective than no screening (resulting in higher average lifetime QALYs), but both are also more expensive. Compared to no screening, the probabilistic ICER (per QALY) for MMS is £8,864 (95% CI £2,600 to £51,576), for USS it is £88,282 (95% CI dominated by no screening to £494,926). However, USS is dominated by MMS as it is estimated to be both more expensive and less effective. Compared to a strategy of no screening, both MMS and USS result in slight increases in life expectancy of 0.58% and 0.28% respectively – equating to an extra 7.39 weeks (1.71 months) for

MMS and 3.58 (0.83 months) weeks for USS. The estimated ICER for MMS (£8,864) is lower than the traditional willingness to pay levels for the NHS[102] of £20,000 per QALY. Recent NIHR-funded research has suggested that this willingness to pay threshold is too high, and a value of £13,000 per QALY may be more appropriate[103]. Based on this stricter criterion MMS would still be below the threshold for cost-effectiveness.

Results based on a time horizon of 11 years estimate that both of the active screening strategies are more costly and produce less QALYs than no screening, and hence are both dominated by it (95% confidence interval for MMS: dominated by no screening to £1,187,704; for USS the 95% confidence interval is always dominated by no screening). These mid-term cost-effectiveness results are because both active screening strategies are associated with both earlier diagnosis (which results in earlier treatment dis-benefit) and false-positive results (which lead to increased costs and reduced QALYs), whilst the late-effect of mortality is not yet sufficient to impact on cost-effectiveness. An exploratory analysis removed false positives, in order to quantify the effect of these on the 11-year results. The total QALYs for no screening, MMS and USS were then 8.2428, 8.2427 and 8.2426 respectively, emphasising that false positive results were driving the majority of the QALY loss, but that at an average of 11 years the dis-benefits of earlier treatment are still likely to outweigh the delayed mortality benefit.

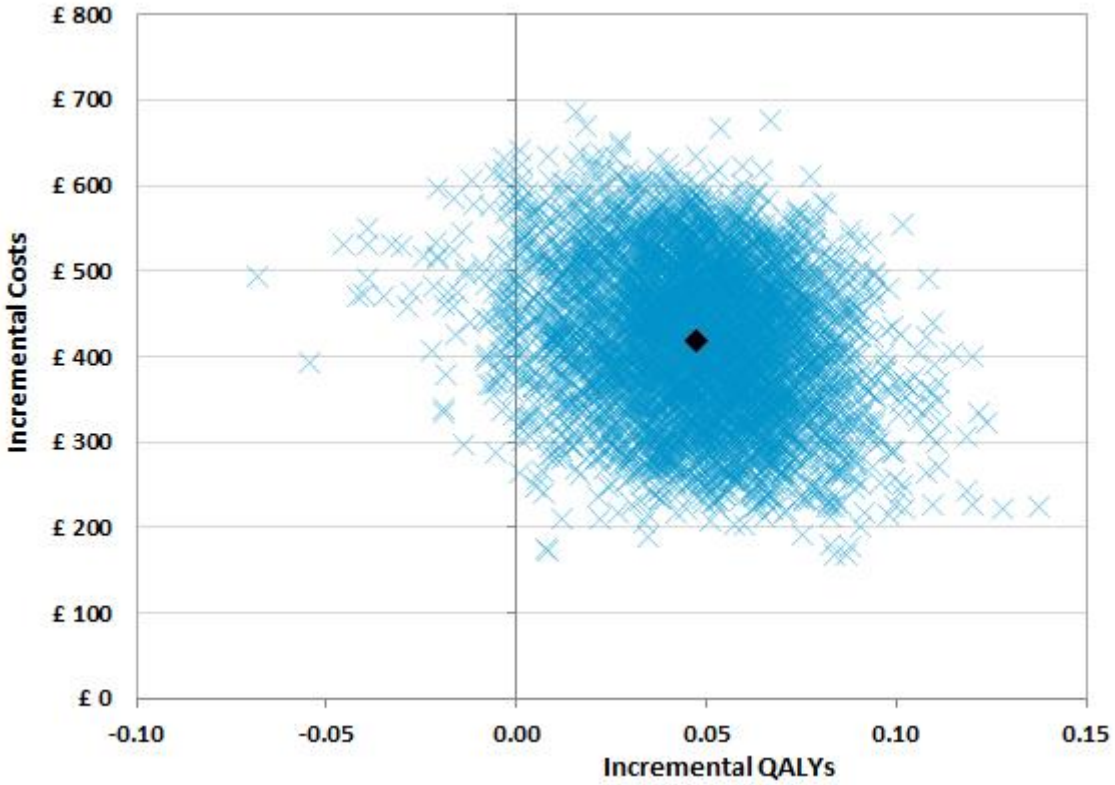
Table 9.2 Cost breakdown (deterministic results).

	Screening costs	Treatment costs	End of life costs
No screening	£0.00	£69.10	£106.67
Multimodal screening	£411.26	£134.54	£47.00
Ultrasound screening	£394.01	£341.05	£75.75

The results from Table 9.2 show that the main cost driver between the different screening strategies is the cost of screening, with both active screening strategies incurring similar screen costs. Each of the two active screening strategies is also associated with an increase in treatment costs. This increase is mainly driven by the additional cost of treating false-positives (for every screen-detected woman with ovarian cancer it is estimated that there will be an additional two surgeries for MMS and 10 for USS). The two active screening strategies are also associated with a decrease in end of life costs, due to a reduced number of women dying from ovarian cancer (for MMS the decrease in end of life costs relative to no screening almost offsets the increase in treatment costs).

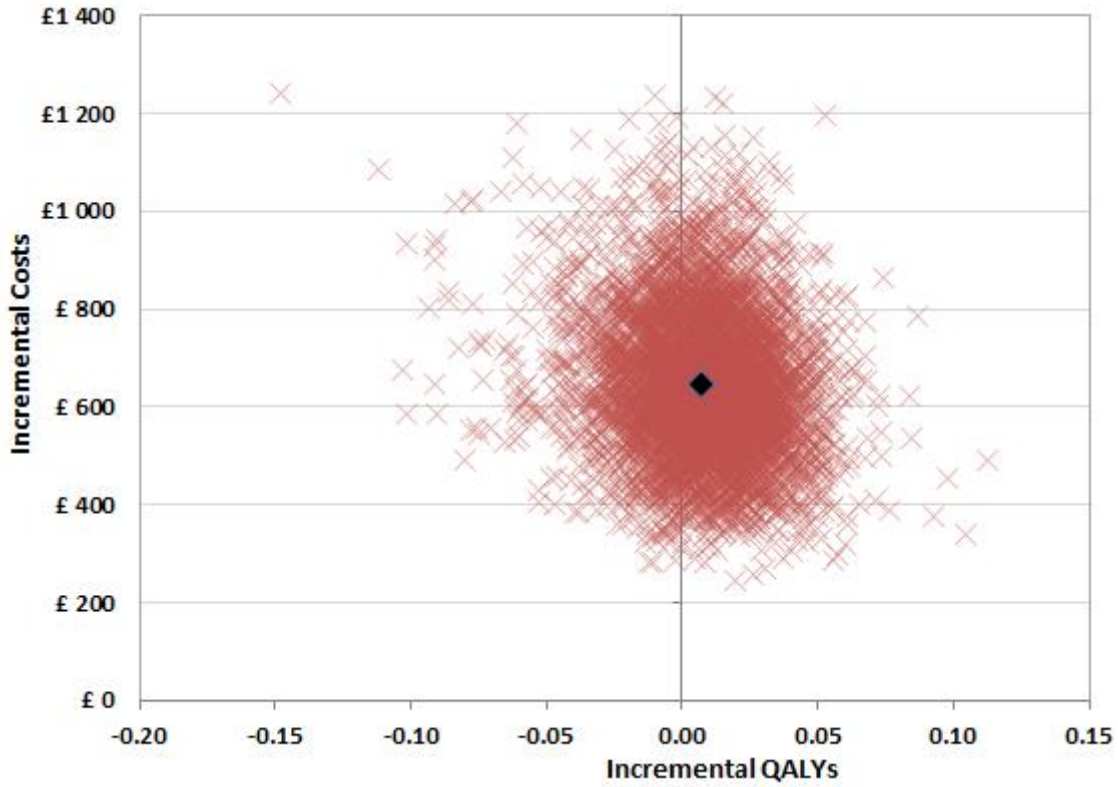
Scatterplots of the incremental costs and QALYs for both MMS and USS when compared with no screening are presented in Figures 9.1 and 9.2 respectively. Both active screening strategies are always more expensive than one of no screening. For MMS the 95% confidence interval for the incremental costs is £255 to £578 (99% confidence interval £222 to £617). For USS the 95% confidence interval for the incremental costs is £386 to £973 (99% confidence interval £328 to £1,094). For MMS the 95% confidence interval for the incremental QALYs was 0.002 to 0.088 (99% confidence interval -0.017 to 0.106). For USS the 95% confidence interval for the incremental QALYs was -0.042 to 0.049 (99% confidence interval -0.073 to 0.062). In 2.1% of the PSA results MMS was estimated to be less effective than no screening, whilst USS was estimated to be less effective in 35.1% of the PSA results. As both screening strategies were always more expensive they were dominated by no screening for these results.

Figure 9.1: Scatterplot of the incremental costs and QALYs for MMS compared to no screening.



QALYs: Quality-adjusted life years. QALYs (costs) greater than zero indicate that MMS is more effective (expensive) than no screening. The average of the PSA runs is indicated by a black diamond.

Figure 9.2: Scatterplot of the incremental costs and QALYs for USS compared to no screening.



QALYs: Quality-adjusted life years. QALYs (costs) greater than zero indicate that USS is more effective (expensive) than no screening. The average of the PSA runs is indicated by a black diamond.

9.2. Uncertainty in the effect of screening on long-term mortality.

Within the UKCTOCS publication[7] a potential late-effect of screening on mortality was identified. Within this study a variety of different approaches to extrapolating mortality were identified, as discussed in Section 8.6 and Appendix 3. The approach taken for the base-case may be summarised as follows:

- Royston-Parmar models were used to estimate the time-varying hazard for the no screening group, and time-varying hazard ratios for MMS and USS (both relative to no screening). These estimates covered the observed trial period.
- Extrapolations were obtained using time-series (exponential smoothing) models. Based on grounds of plausibility, the extrapolated hazard ratios were damped, but the extrapolated hazard for no screening was not damped. The impact of dampening was that the extrapolated hazard ratios approached a constant value (instead of continuously decreasing over time).

The resulting estimated hazard for no screening and hazard ratio for MMS are displayed in Figures 9.3 and 9.4, whilst the estimated proportion of the cohort dying from ovarian cancer over time is displayed in Figures 9.5 and 9.6 for no screening and MMS. Similar graphs are provided for USS in Appendix 4; the focus for this Section is comparing MMS with no screening as these are the two screening strategies that are potentially cost-effective.

Figure 9.3 **Estimated hazard for no screening over-time: top-pane within trial estimates, bottom-pane over lifetime.**

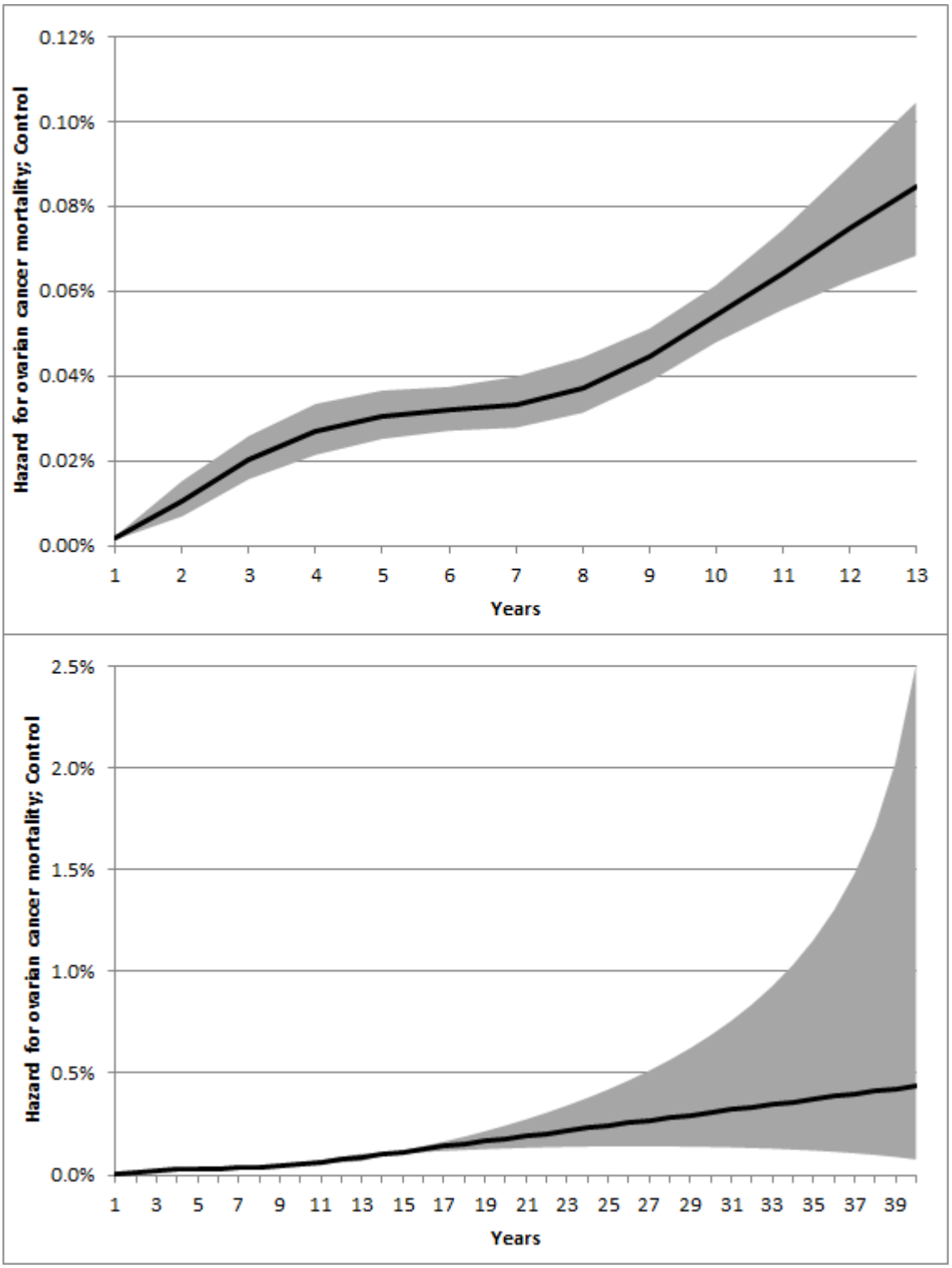


Figure 9.4 **Estimated hazard ratio for multimodal screening over-time: top-pane within trial estimates, bottom-pane over lifetime.**

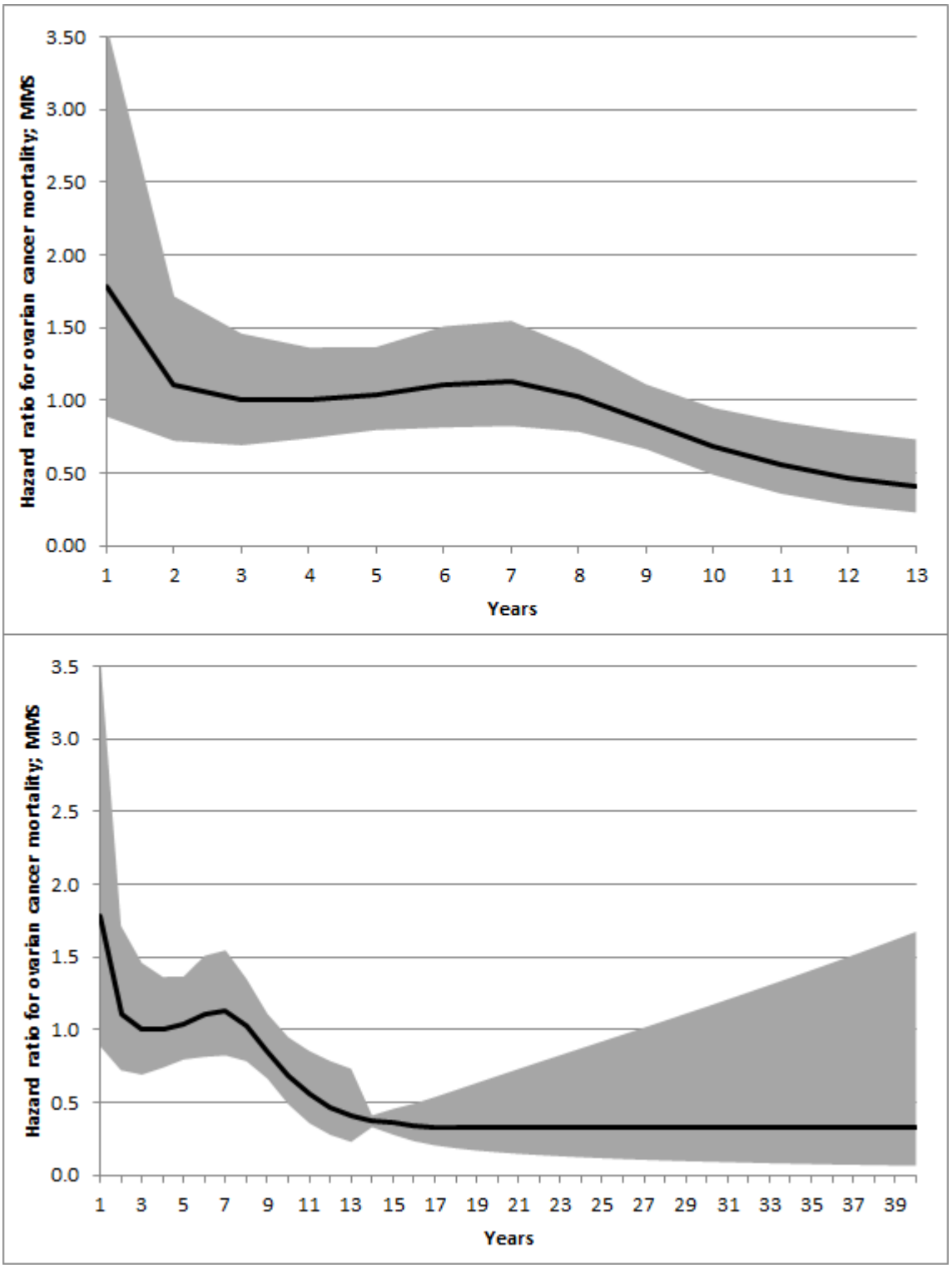


Figure 9.5: **Cumulative ovarian cancer mortality for no screening**

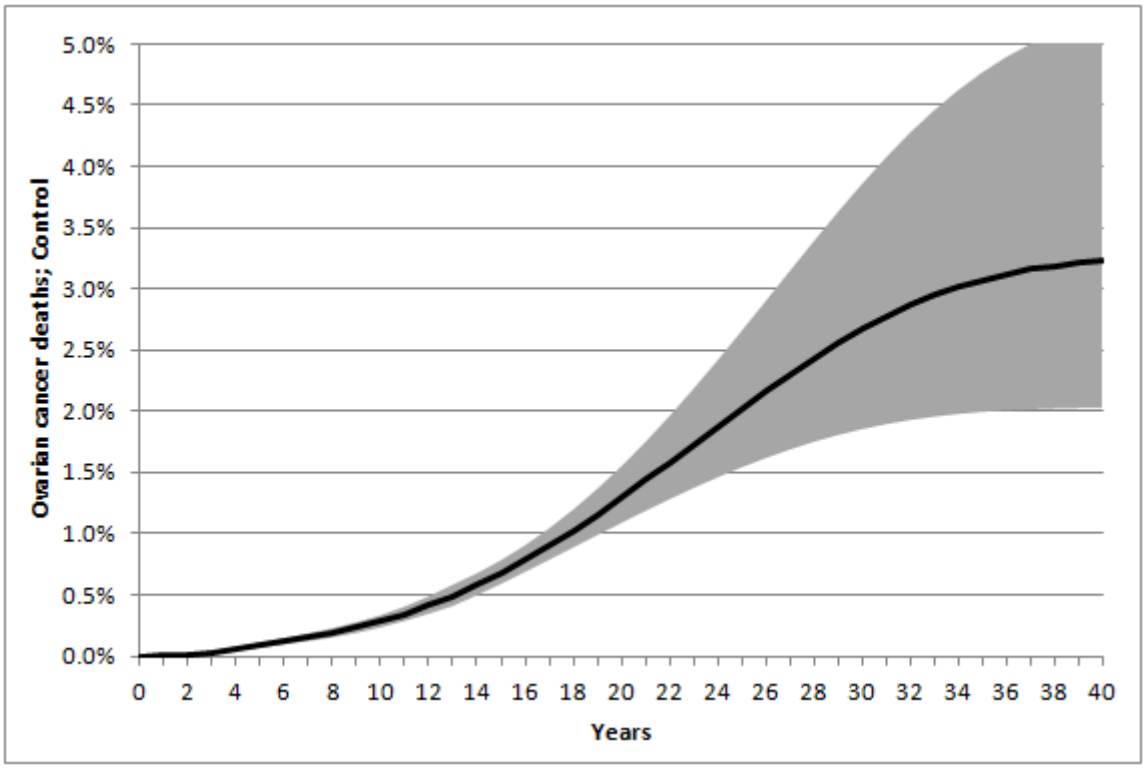
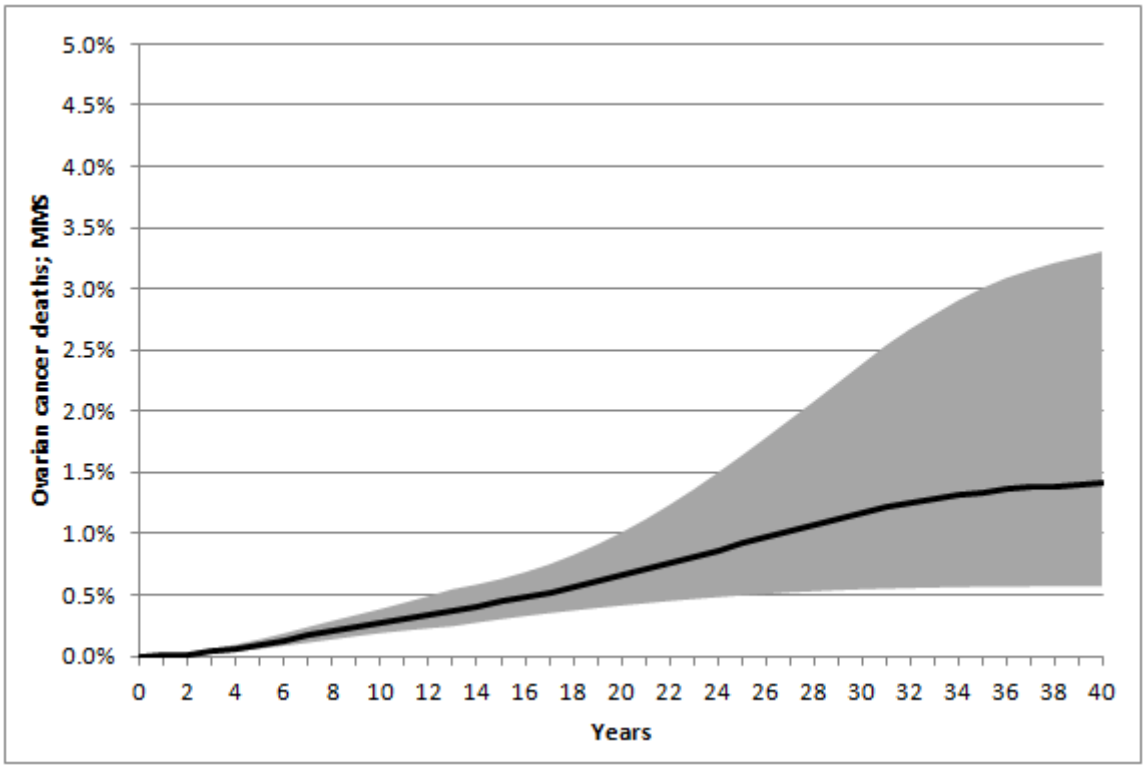


Figure 9.6: **Cumulative ovarian cancer mortality for multimodal screening**



Figures 9.3 to 9.6 highlight that there is uncertainty in both the estimates of mortality for no screening and the effect of screening on mortality. The majority of this uncertainty relates to the extrapolated period, with increasing uncertainty as extrapolations move further into the future. Figure 9.4 also suggests that the time-series model is under-predicting the uncertainty in the hazard ratio for MMS, at least for the initial extrapolation years.

As discussed in Section 8.6, alternative extrapolation methods on estimates of cumulative ovarian cancer mortality are possible. These include the use of model discrepancy and the use of parametric models, with or without a restriction on if the same parametric model is used for each trial arm. Estimates of cumulative ovarian cancer mortality for these approaches are displayed in Appendix 4. The impact on both health outcomes and cost-effectiveness are displayed for no screening and both MMS and USS in Tables 9.3 to 9.4.

Table 9.3 Estimates of life years and ovarian cancer mortality for different extrapolation assumptions.

Base-case results	**No screening**	**MMS**	**USS**
Life years	24.660	24.803	24.729
Ovarian cancer deaths	3.19%	1.41%	2.35%
Separate parametric models	**No screening**	**MMS**	**USS**
Life years	24.743	24.820	24.818
Ovarian cancer deaths	1.99%	1.04%	1.05%
Same parametric models	**No screening**	**MMS**	**USS**
Life years	24.778	24.821	24.818
Ovarian cancer deaths	1.49%	1.03%	1.05%
Model discrepancy; 1% per year	**No screening**	**MMS**	**USS**
Life years	24.659	24.796	24.714
Ovarian cancer deaths	3.21%	1.54%	2.62%
Model discrepancy; 2% per year	**No screening**	**MMS**	**USS**
Life years	24.660	24.787	24.702
Ovarian cancer deaths	3.20%	1.69%	2.84%
Model discrepancy; 3% per year	**No screening**	**MMS**	**USS**
Life years	24.658	24.782	24.688
Ovarian cancer deaths	3.22%	1.79%	3.07%
Model discrepancy; 4% per year	**No screening**	**MMS**	**USS**
Life years	24.659	24.770	24.668
Ovarian cancer deaths	3.24%	1.92%	3.27%
Model discrepancy; 5% per year	**No screening**	**MMS**	**USS**
Life years	24.659	24.800	24.722
Ovarian cancer deaths	3.22%	1.99%	3.42%

Table 9.4 Estimates of lifetime costs, effects, and cost-effectiveness for different extrapolation assumptions.

Base-case results	**No screening**	**MMS**	**USS**
Costs	£179	£598	£824
QALYs	14.290	14.357	14.297
ICER	-	£8,864 vs no screening	Dominated by MMS
Separate parametric models	**No screening**	**MMS**	**USS**
Costs	£143	£588	£782
QALYs	14.352	14.376	14.361
ICER	-	£18,372 vs no screening	Dominated by MMS
Same parametric models	**No screening**	**MMS**	**USS**
Costs	£128	£587	£780
QALYs	14.343	14.356	14.341
ICER	-	£36,769 vs no screening	Dominated by MMS
Model discrepancy; 1% per year	**No screening**	**MMS**	**USS**
Costs	£180	£602	£829
QALYs	14.321	14.367	14.325
ICER	-	£9,257 vs no screening	Dominated by MMS
Model discrepancy; 2% per year	**No screening**	**MMS**	**USS**
Costs	£179	£606	£834
QALYs	14.263	14.305	14.262
ICER	-	£10,004 vs no screening	Dominated by MMS
Model discrepancy; 3% per year	**No screening**	**MMS**	**USS**
Costs	£180	£609	£842
QALYs	14.280	14.322	14.275
ICER	-	£10,378 vs no screening	Dominated by MMS
Model discrepancy; 4% per year	**No screening**	**MMS**	**USS**
Costs	£180	£611	£848
QALYs	14.274	14.314	14.266
ICER	-	£10,889 vs no screening	Dominated by MMS
Model discrepancy; 5% per year	**No screening**	**MMS**	**USS**
Costs	£179	£614	£851
QALYs	14.209	14.244	14.194
ICER	-	£12,643 vs no screening	Dominated by MMS

The results presented in Tables 9.3 and 9.4 show that the use of parametric curves for extrapolation results in a much lower estimated lifetime number of ovarian cancer deaths across all three screening arms, with the largest reductions observed for the no screening group (reduced from

3.19% to 1.99% or 1.49% depending on the parametric function used). This in turn affects the ICER for MMS, which relative to no screening increases from £8,864 in the base-case to either £18,372 or £36,769 depending on the parametric function used. As anticipated, the cost-effectiveness of MMS relative to no screening worsened with increasing levels of bias. However, for the levels tested the results are less affected by model discrepancy than by different extrapolation assumptions. The most extreme bias tested; of a cumulative 5% per year (which results in the hazard ratio being multiplied by 2.35 at the end of the time-horizon), resulted in an increase in the ICER of 43% to £12,643, which remains both of the willingness to pay thresholds considered for this study.

9.3. Sensitivity analyses and model validation.

Results of the sensitivity analyses described in Section 8.6 are displayed in Table 9.5 over-page. For all of the sensitivity analyses USS remained dominated by MMS, so the discussion of this section focuses on comparing MMS with no screening. Of the eight sensitivity analyses considered, results were robust to different assumptions about treatment costs, with relative changes in the ICER for MMS of less than 0.5%. Results were moderately robust to different assumptions about the utility of women with ovarian cancer (both at the end of life, and the impact of earlier treatment on utility), with relative changes in the ICER for MMS of about 5%.

Using the UKCTOCS results, active screening reduces ovarian cancer mortality. Under the base-case this leads to reduced end of life costs, which are only modelled for ovarian cancer deaths. Removing this end-of life cost increased the ICER for MMS by 14.3% to about £10,000 per QALY, suggesting some sensitivity to this assumption. However, this ICER is still below both estimates of national willingness to pay levels for the NHS (of £13,000 and £20,000 per QALY).

For the base-case, false-positive screen results are modelled for MMS and USS, but it is assumed that under no screening, when ovarian cancers are likely to be diagnosed in primary care, there would not be any false-positive results. Instead assuming that diagnosis in primary care has the same false-positive rate as MMS leads to a reduced ICER for MMS of £6,691 per QALY (a 21% relative reduction).

Cost-effectiveness results were sensitive to the assumed cost of MMS; using the high cost estimate led to a relative increase in the ICER for MMS of 23%, whilst using the low estimate led to a relative reduction of 40%. The corresponding ICERs are £10,394 and £6,691 per QALY with the high ICER remaining below both the £13,000 and £20,000 willingness to pay thresholds.

Table 9.5 Results from deterministic one-way sensitivity analyses.

Base-case results	**No screening**	**MMS**	**USS**
QALYs	14.281	14.330	14.291
Costs	£176	£593	£811
ICER	-	£8,459 vs no screening	Dominated by MMS
False-positives for no screening same as for MMS	**No screening**	**MMS**	**USS**
QALYs	14.277	14.330	14.291
Costs	£237	£593	£811
ICER		£6,691 vs no screening	Dominated by MMS
Treatment costs monotonically increase by 10%	**No screening**	**MMS**	**USS**
QALYs	14.281	14.330	14.291
Costs	£179	£594	£814
ICER		£8,423 vs no screening	Dominated by MMS
Birmingham treatment cost estimates	**No screening**	**MMS**	**USS**
QALYs	14.281	14.330	14.291
Costs	£180	£596	£815
ICER		£8,454 vs no screening	Dominated by MMS
No end of life costs	**No screening**	**MMS**	**USS**
QALYs	14.281	14.330	14.291
Costs	£69	£546	£735
ICER		£9,670 vs no screening	Dominated by MMS
Low-cost estimate for MMS	**No screening**	**MMS**	**USS**
QALYs	14.281	14.330	14.291
Costs	£176	£426	£811
ICER		£5,071 vs no screening	Dominated by MMS
High-cost estimate for MMS	**No screening**	**MMS**	**USS**
QALYs	14.281	14.330	14.291
Costs	£176	£688	£811
ICER		£10,394 vs no screening	Dominated by MMS
End-of-life disutility for ovarian cancer deaths	**No screening**	**MMS**	**USS**
QALYs	14.275	14.327	14.286
Costs	£176	£593	£811
ICER	-	£8,020 vs no screening	Dominated by MMS
Earlier treatment leads to reduced utility	**No screening**	**MMS**	**USS**
QALYs	14.281	14.328	14.289
Costs	£176	£593	£811
ICER	-	£8,773 vs no screening	Dominated by MMS

To validate the model results, probabilistic results for a time-horizon of 11 years were estimated, and compared to those from the UKCTOCS trial (which had a median of 11.1 years follow-up). This comparison is presented in Table 9.6. Estimates of uncertainty in the results are represented by 95% confidence intervals (percentile-based for model estimates and using the Wilson method[104] for observed trial data). The results show close agreement for both the proportion of cancers diagnosed and the proportion of deaths due to ovarian cancer.

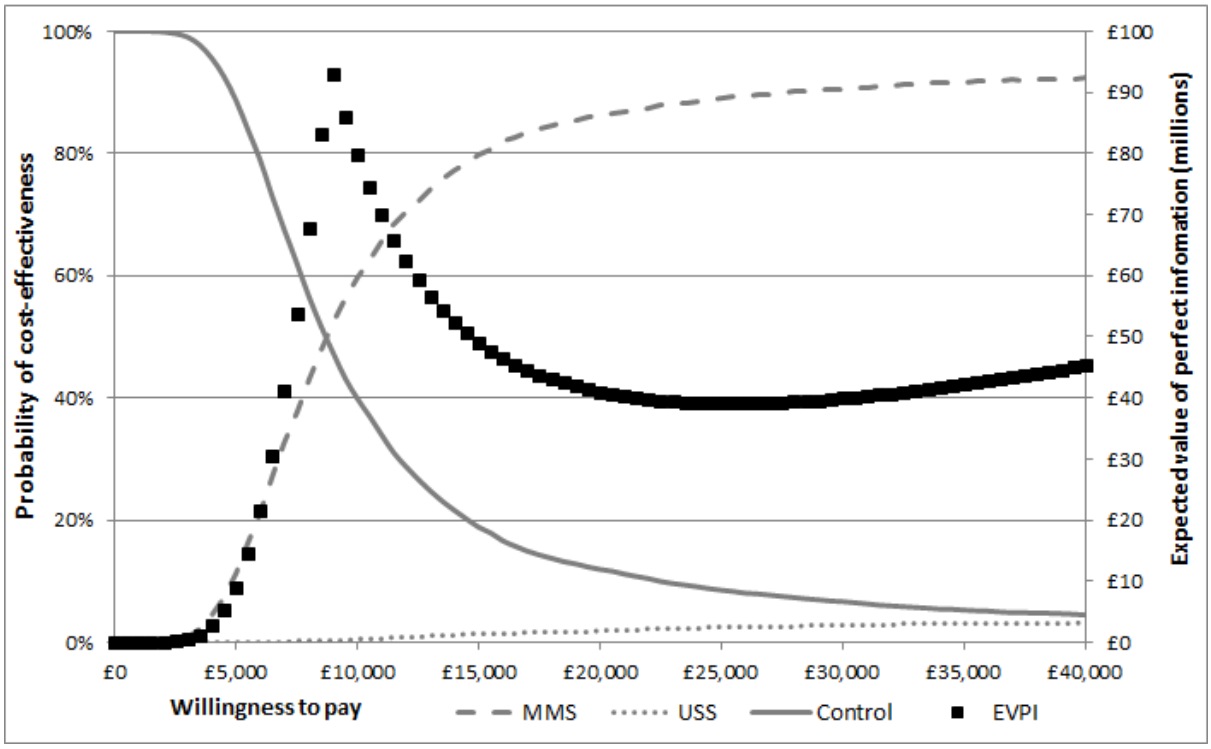
Table 9.6 Comparison of UKCTOCS results with model-estimates (time-horizon 11 years).

Model 11 year results	**Cancers diagnosed**	**95% confidence interval**	**Deaths due to ovarian cancer**	**95% confidence interval**
No screening	0.59%	0.55% to 0.64%	0.35%	0.32% to 0.39%
Multimodal screening	0.64%	0.58% to 0.72%	0.30%	0.26% to 0.36%
Ultrasound screening	0.61%	0.55% to 0.69%	0.31%	0.26% to 0.37%
Observed trial results	**Cancers diagnosed**	**95% confidence interval**	**Deaths due to ovarian cancer**	**95% confidence interval**
No screening	0.62%	0.58% to 0.67%	0.34%	0.31% to 0.38%
Multimodal screening	0.67%	0.60% to 0.74%	0.29%	0.25% to 0.34%
Ultrasound screening	0.62%	0.56% to 0.69%	0.30%	0.26% to 0.36%

9.4. The value of further research.

The probability of each of the screening strategies being cost-effective is displayed in Figure 9.7 for willingness to pay values between £0 and £40,000. The population expected value of perfect information is also displayed on a separate axis in Figure 9.7. Willingness to pay values indicate the relative weight that is given to gain a QALY. For example, a willingness to pay of £20,000 indicates that a decision-maker is willing to pay £20,000 to gain an additional QALY.

Figure 9.7: The probability that each screening strategy is cost-effective for different willingness to pay values (left-axis) and the expected value of perfect information.



EVPI: Expected value of perfect information. MMS: Multimodal screening. USS: Ultrasound screening.

At low willingness to pay thresholds cost is given more weight than effectiveness, and so the strategy of no screening is more likely to be cost-effective. As more weight is given to effectiveness, MMS becomes more likely to be cost-effective. The probability of the USS strategy, which is dominated by MMS in the base-case, being cost-effective increases monotonically with increasing willingness to pay levels; at a level of £40,000 it has a probability of 3.2% of being cost-effective. The probabilities of no screening and MMS being cost-effective for certain willingness to pay values is displayed in Table 9.7

Table 9.7 Probability of being-cost effective.

Willingness to pay	Probability no screening cost-effective	Probability multimodal screening cost-effective
£13,000	24.8%	74.2%
£20,000	11.9%	86.2%
£30,000	6.7%	90.6%
£40,000	4.5%	92.3%

Population EVPI reaches a maximum at a willingness to pay threshold of £9,000, with a value of approximately 90 million pounds.

Estimates of EVPPI, taken at a willingness to pay threshold of £20,000, show that the key drivers of decision uncertainty are the parameters used to generate estimates of mortality for the MMS screening group and the HRQoL for women with ovarian cancer. These contribute to 53% and 22% of the decision uncertainty, respectively. It is estimated that it would be worth spending approximately £20 million to eliminate the uncertainty in the long-term effectiveness of screening. Results were relatively robust to the inclusion of model discrepancy, although as the size of the bias increased, the cost of screening with MMS contributed more to decision uncertainty.

It should be noted that whilst there is uncertainty in the cost of ultrasound screening, and the long-term effects of USS on mortality, these have a negligible impact on decision uncertainty as USS has a very low probability of being cost-effective. Limitations of the EVPI and EVPPI estimates are that they do not incorporate structural uncertainties; only parameter uncertainties are quantified under the assumption that there is no uncertainty in the model structure. The model discrepancy approach does address this limitation to a degree, although as the results from Table 9.4 show, alternative structural extrapolation assumptions have a larger impact on cost-effectiveness results than the model discrepancy approach, for the range of bias explored.

10. Discussion

The modelled results suggest that both screening strategies are likely to result in health benefits when compared to no screening, but at increased costs. Screening using MMS is estimated to be both more effective and cheaper than USS. The base-case lifetime ICER comparing MMS against no screening was estimated to be £8,864 per QALY (95% confidence interval £2,600 to £51,576). Based on an 11-year time horizon both MMS and USS are dominated by a strategy of no screening, emphasising that the benefits of active screening require a long time-horizon before they are observed.

Results of the sensitivity analyses suggest that the cost-effectiveness results are robust to a range of different assumptions relating to the cost of both treatment and end of life care, along with certain assumptions about the utility of women with ovarian cancer. However, they were sensitive to the estimated cost of MMS. Model results were also sensitive to the approach taken to extrapolation. The base-case approach extrapolated both the hazard function for no screening and the hazard ratios for MMS and USS. Alternative approaches include the use of parametric survival models for extrapolation, with or without a constraint on using the same parametric model. These approaches led to ICERs for MMS relative to no screening of £18,372 and £36,769, which represent a doubling and quadrupling of the base-case ICER, respectively.

Value of information analyses carried-out for the economic evaluation suggest that of these uncertainties, uncertainties in the long-term mortality benefit due to screening with MMS had the largest impact on decision uncertainty. Uncertainty in the utility values for women with ovarian cancer also led to decision uncertainty. Other uncertainties (in the cost of USS screening, long term estimates of incidence, and the cost of diagnosis, treatment and palliative care) had a very small impact on decision uncertainty. There is the potential to extend follow-up of the UKCTOCS trial. In the context of the cost-effectiveness results presented in this report, whether or not extended follow-up would represent a valuable use of resources depends on decision makers' willingness to pay to eliminate this uncertainty. At a threshold of £20,000, these results suggest that it would be worth spending approximately £20 million to eliminate all of the uncertainty in the long-term effectiveness of MMS.

There are currently no published economic evaluations of screening for ovarian cancer from a UK perspective. Of the three economic evaluations provided, one[55] was deemed to not have any

useable results. The two remaining economic evaluations[12,53] both used a US healthcare perspective and neither considered HRQoL as a health economic outcome, using instead cost per life years. Havrilesky *et al*[53] considered hypothetical screening tests and concluded that annual screening had the potential to be cost-effective, with a relative ovarian cancer mortality reduction of 43%. Drescher *et al*[12] considered annual testing with CA-125 followed by TVS, along with hypothetical tests. For the former they estimated a mortality reduction due to earlier diagnosis of 9%, and suggested that a more effective screening test, which would reduce mortality by at least 25%, was required to be cost-effective. The results presented here estimated a lifetime reduction in ovarian cancer deaths of 56% with MMS (absolute reduction of 1.8%) for the base-case. Alternative extrapolation assumptions led to relative reductions between 31% and 48% (absolute reductions 0.46% and 0.95%).

To inform this work three systematic reviews of the literature were performed to identify evidence on clinical effectiveness, existing economic evaluations, and utility values pertinent to ovarian cancer screening, diagnosis and treatment. A combination of national reference sources, expert opinion and English Cancer registry evidence were used to derive estimates of resource use and cost. A number of evidence gaps were identified: there is uncertainty over the long-term impact of ovarian cancer screening on mortality, there is a paucity of high-quality evidence quantifying the health-related quality of life for women with ovarian cancer, and there are uncertainties in the costs of ovarian cancer screening; in particular the cost of a CA-125 is based on expert opinion, whilst it was assumed that implementing ROCA would not lead to an increase in screening costs.

The lack of age and stage breakdowns for both the incidence of, and mortality from, ovarian cancer led to a number of limitations. It was not possible to create a model of the natural history of cancer. This limited the analysis in that it was not possible to estimate the potential cost-effectiveness of alternative screening strategies (such as different screening intervals or different age-ranges), nor was it possible to evaluate the impact on cost-effectiveness of improvements in the screening test characteristics (such as sensitivity and specificity) or changes in compliance. A previous US-based economic evaluation of screening for ovarian cancer suggested that test frequency and specificity may be key drivers of cost-effectiveness[53]. The lack of a natural history model (which would have enabled calibration of unobservable parameters such as time spent with undiagnosed ovarian cancer[91]) also limits current understanding of the natural history of ovarian cancer. This includes

elucidation of the potential mechanisms by which screening is associated with a reduction in mortality, which may be due to a combination of diagnosis at an earlier stage, and earlier diagnosis within a given stage. Further, only average effects from the UKCOTCS could be captured. This included modelling a cohort of women with an average age of 60 when screening commenced. However, menopause may start earlier than this, and the UKCTOCS randomised women from the age of fifty. Hence the health economic results presented may not be entirely representative of the impact of a screening programme that began inviting women at the age of 50. In addition, 11 years of screening were modelled, reflecting the maximum duration of screening in the UKCTOCS. However, if screening were to be implemented, it is unclear how long annual screening would occur for. If women commenced screening at the age of 50, screening for 11 years would see them screened until the age of 61. There are also concerns about the feasibility of annual screening; results from the UKCTOCS[7] showed a year-on-year decrease in compliance with screening, with half of all eligible women attending 11 screens.

In conclusion, the base-case ICER comparing MMS with no screening is £8,864 per QALY. A key driver of uncertainty about the cost-effectiveness of this result is the long-term effectiveness of MMS for reducing ovarian cancer mortality. The other screening strategy considered by UKCTOCS – USS – is unlikely to be cost-effective.

Appendix 1: Ovarian Cancer Searches and Results

All searches were performed during September and October 2014.

The search results by database are provided in the Table.

Table Search results by database.

Database	RCTs($$rct)	Economic and Quality of Life ($$economicqol)
Medline	1114	324 and 345 = 669 (642 deduplicated)
Embase	3054	558 (Quality of Life) 890 (Economics)= 1448 (1400 deduplicated)
CINAHL	238	313 and 160 = 473 (455 deduplicated)
Web of Science	490	473 (183 records after deduplication)
Cochrane CENTRAL	195	Records come from Medline so not searched.
Cochrane NHS EED	Not searched	80
Cochrane CDSR	6	Not searched
Cochrane DARE	5	Not searched
Cochrane HTA	15	Not searched
Cochrane Methods	8	Not searched
Econlit	Not searched	11
Total	5125	2771
Total after deduplication*	2716	2089

Search One - Economic and Quality of Life ($$economicqol)

Canadian Ovarian Cancer terms AND Quality of Life terms.
OR
Canadian Ovarian Cancer terms AND Ruth screening terms AND Ruth economic filter.

Search Two - RCTs($$rct)

Canadian Ovarian Cancer terms AND Canadian screening terms (n.b we will be excluding the terms from the Canadian search relating to hereditary cancers).
2011-2014.
No filter.
Limit to Humans and English Language

What we are not searching for

We are not going to identify any evidence on treatments for ovarian cancer that do not involve screening. This information, which may inform the Quality of Life review, will be obtained through discussion with experts.

*Deduplication also involved removal of pre 2011 papers from the RCT search, removal of obvious non English studies and removal of obvious animal studies. However these were not searched for systematically so errors may remain.

Canadian Ovarian Cancer Terms (include tubal and peritoneal terms but exclude hereditary terms)

1. Exp Ovarian Neoplasms/
2. Fallopian Tube Neoplasms/
3. Peritoneal Neoplasms/
4. (fallopian tube adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp
5. (peritoneal adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp
6. (ovarian adj3 cancer*).mp
7. (ovarian adj3 neoplas*).mp
8. (ovarian adj3 tumo*).mp
9. (ovarian adj3 carcinoma*).mp
10. (ovarian adj3 adenocarcinoma*).mp

Ruth Screening Terms

1 Tumor Markers, Biological/
2. Biological Markers/
3. CA-125 Antigen/
4. Mass Screening/
5. (screen$ or test$ or imag$ or predict$ or surveillance).tw.
6. "Early Detection of Cancer"/
7. (earl$ adj (diagnos$ or detect$)).ti,ab.

Canadian screening terms

1. Mass Screening/
2. "Early Detection of Cancer"/
3. screen*.mp

Ruth RCT filter

1. Randomized controlled trials as Topic/
2. Randomized controlled trial/
3. Random allocation/
4. randomized controlled trial.pt.
5. Double blind method/
6. Single blind method/
7. Clinical trial/
8. exp Clinical Trials as Topic/
9. controlled clinical trial.pt.
10. clinical trial$.pt.
11. multicenter study.pt.
12. or/1-12
13. (clinic$ adj25 trial$).ti,ab.
14. ((singl$ or doubl$ or treb$ or tripl$) adj (blind$ or mask$)).tw.
15. Placebos/
16. Placebo$.tw.
17. (allocated adj2 random).tw.
18. or/13-18
19. 12 or 18
20. Case report.tw.
21. Letter/
22. Historical article/
23. 20 or 21 or 22
24. exp Animals/
25. Humans/
26. 24 not (24 and 25)
27. 23 or 26
28. 19 not 27

Ruth economic evaluation filter

1. exp "Costs and Cost Analysis"/
2. Economics/
3. exp Economics, Hospital/
4. exp Economics, Medical/
5. Economics, Nursing/
6. exp models, economic/
7. Economics, Pharmaceutical/
8. exp "Fees and Charges"/
9. exp Budgets/
10. budget$.tw.
11. ec.fs.
12. cost$.ti.
13. (cost$ adj2 (effective$ or utilit$ or benefit$ or minimi$)).ab.
14. (economic$ or pharmacoeconomic$ or pharmaco-economic$).ti.
15. (price$ or pricing$).tw.
16. (financial or finance or finances or financed).tw.
17. (fee or fees).tw.
18. (value adj2 (money or monetary)).tw.
19. quality-adjusted life years/
20. (qaly or qalys).af.
21. (quality adjusted life year or quality adjusted life years).af.
22 or/1-21

Quality of Life terms

1.(EORTC or QLQ-C30).mp.
2.FACT or functional assessment of cancer therapy or FLIC or (FACT adj3 cancer)
3. RSCL or SDS or Fact-G or CNQ-SF or CNQ or CARES-SF or CARES or ESAS
4.quality adjusted life year/ or quality adjusted life.tw. or (qaly$ or qald$ or qale$ or qtime$).tw.
4. disability adjusted life.tw. or daly$.tw.
5. health status indicators/
6. (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
7. (sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
8. (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
9. (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen).tw.
10. (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
11. (euroqol or euro qol or eq5d or eq 5d).tw.
12. (hql or hqol or h qol or hrqol or hr qol).tw.
13. (hye or hyes).tw. or (health$ year$ equivalent$).tw.
14. health utilit$.tw. or (hui or hui1 or hui2 or hui3).tw.
15. ("Quality of Life" or “Outcome Assessment (Health Care)"/ or quality of life.tw.) AND (preference based or utilit$ or generic preference).tw.

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

1. exp Ovarian Neoplasms/
2. Fallopian Tube Neoplasms/
3. Peritoneal Neoplasms/
4. (fallopian tube* adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp.
5. (peritoneal adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp.
6. ovarian cancer*.mp.
7. ovarian tumo*.mp.
8. ovarian carcinoma*.mp.
9. ovarian adenocarcinoma*.mp.
10. ovarian neoplas*.mp.
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. (EORTC or QLQ-C30).mp.
13. ((FACT or FLIC) adj3 cancer).mp.
14. ((RSCL or SDS or Fact-G or CNQ-SF or CNQ or CARES-SF or CARES or ESAS) adj3 cancer*).mp.
15. quality adjusted life year/ or quality adjusted life.mp. or (qaly* or qald* or qale* or qtime*).mp.
16. (disability adjusted life or daly$).mp.
17. health status indicators/
18. (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).mp.
19. (sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).mp.
20. (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).mp.
21. (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen).mp.
22. (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).mp.
23. (euroqol or euro qol or eq5d or eq 5d).mp.
24. (hql or hqol or h qol or hrqol or hr qol).mp.
25. (hye or hyes or health* year* equivalent*).mp.
26. (health utilit* or (hui or hui1 or hui2 or hui3)).mp.
27. (Quality of Life/ or Outcome Assessment/) and (preference based or utilit* or generic preference).mp.
28. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29. 11 and 28
30. limit 29 to (english language and humans)
31. Tumor Markers, Biological/
32. Biological Markers/
33. CA-125 Antigen/
34. Mass Screening/
35. "Early Detection of Cancer"/
36. (screen$ or test$ or imag$ or predict$ or surveillance).tw.
37. (earl$ adj (diagnos$ or detect$)).ti,ab.
38. Early Detection of Cancer/
39. 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38
40. exp "Costs and Cost Analysis"/
41. exp Economics, Hospital/
42. exp Economics, Medical/
43. Economics, Nursing/
44. exp models, economic/

45. Economics, Pharmaceutical/
46. exp "Fees and Charges"/
47. exp Budgets/
48. budget$.tw.
49. ec.fs.
50. cost$.ti.
51. (cost$ adj2 (effective$ or utilit$ or benefit$ or minimi$)).ab.
52. (economic$ or pharmacoeconomic$ or pharmaco-economic$).ti.
53. (price$ or pricing$).tw.
54. (financial or finance or finances or financed).tw.
55. (fee or fees).tw.
56. (value adj2 (money or monetary)).tw.
57. quality-adjusted life years/
58. (qaly or qalys).af.
59. (quality adjusted life year or quality adjusted life years).af.
60. 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59
61. 11 and 39 and 60
62. limit 61 to (english language and humans)
63. 30 or 62

Embase 1988

1. exp Ovarian Neoplasms/
2. Fallopian Tube Neoplasms/
3. Peritoneal Neoplasms/
4. (fallopian tube* adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp.
5. (peritoneal adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp.
6. ovarian cancer*.mp.
7. ovarian tumo*.mp.
8. ovarian carcinoma*.mp.
9. ovarian adenocarcinoma*.mp.
10. ovarian neoplas*.mp.
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. (EORTC or QLQ-C30).mp.
13. ((FACT or FLIC) adj3 cancer).mp.
14. ((RSCL or SDS or Fact-G or CNQ-SF or CNQ or CARES-SF or CARES or ESAS) adj3 cancer*).mp.
15. quality adjusted life year/ or quality adjusted life.mp. or (qaly* or qald* or qale* or qtime*).mp.
16. (disability adjusted life or daly$).mp.
17. health status indicators/
18. (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).mp.
19. (sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).mp.
20. (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).mp.
21. (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen).mp.
22. (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).mp.
23. (euroqol or euro qol or eq5d or eq 5d).mp.
24. (hql or hqol or h qol or hrqol or hr qol).mp.

25. (hye or hyes or health* year* equivalent*).mp.
26. (health utilit* or (hui or hui1 or hui2 or hui3)).mp.
27. (Quality of Life/ or Outcome Assessment/) and (preference based or utilit* or generic preference).mp.
28. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29. 11 and 28
30. limit 29 to (english language and humans)
31. Tumor Markers, Biological/
32. Biological Markers/
33. CA-125 Antigen/
34. Mass Screening/
35. "Early Detection of Cancer"/
36. (screen$ or tested or testing or test or tests or imag$ or predict$ or surveillance).tw.
37. (earl$ adj (diagnos$ or detect$)).ti,ab.
38. Early Detection of Cancer/
39. 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38
40. exp "Costs and Cost Analysis"/
41. exp Economics, Hospital/
42. exp Economics, Medical/
43. Economics, Nursing/
44. exp models, economic/
45. Economics, Pharmaceutical/
46. exp "Fees and Charges"/
47. exp Budgets/
48. 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47
49. 11 and 39 and 48
50. limit 49 to (human and english language)
51. 30 or 50

Econlit

Limited search to Ovarian Cancer terms only. The only ovarian cancer term that had any hits was ovarian cancer*.mp. so the search was limited to this.

1. Ovarian cancer.mp

CINAHL

S48 S28 OR S47
S47 S37 AND S46
S46 (S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45)
S45 TX (value N2 (money or monetary))
S44 TX fee or fees
S43 TX financial or finance or finances or financed
S42 TX price? or pricing?
S41 TX economic* or pharmacoeconomic* or pharmaco-economic*
S40 TX cost*
S39 TX budget*
S38 MH economics
S37 S4 AND S36
S36 S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35
S35 TX earl* N (diagnos* or detect*)

S34 TX screen* or test* or imag* or predict* or surveillance
S33 MH Early Detection of Cancer
S32 MH Mass Screening
S31 MH CA-125 Antigen
S30 MH Biological Markers
S29 MH Tumor Markers, Biological
S28 (S4 AND S27)
S27 S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S26
S26 (TX (preference based or utilit* or generic preference)) AND (S24 AND S25)
S25 TX preference based or utilit* or generic preference
S24 MH Outcome Assessment
S23 MH Quality of Life
S22 (health utilit* or (hui or hui1 or hui2 or hui3))
S21 hye or hyes or health* year* equivalent*
S20 TX hql or hqol or h qol or hrqol or hr qol
S19 TX euroqol or euro qol or eq5d or eq 5d
S18 TX sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty
S17 TX sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen
S16 TX sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve
S15 TX sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six
S14 TX (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six)
S13 MH health status indicators
S12 TX disability adjusted life or daly*
S11 TX qaly* or qald* or qale* or qtime*
S10 TX quality adjusted life
S9 MH qaly
S8 MH quality adjusted life year
S7 TX ((RSCL or SDS or Fact-G or CNQ-SF or CNQ or CARES-SF or CARES or ESAS) N3 cancer*)
S6 TX ((FACT or FLIC) N3 cancer)
S5 TX (EORTC or QLQ-C30)
S4 S1 OR S2 OR S3
S3 TX ((fallopian tube* or peritoneal or ovarian) N1 (neoplasm* or cancer* or tumo* or carcinoma* or adenocarcinoma))
S2 MH Peritoneal Neoplasms
S1 MH ovarian neoplasms

Cochrane Library

#1 MeSH descriptor: [Ovarian Neoplasms] explode all trees
#2 MeSH descriptor: [Peritoneal Neoplasms] explode all trees
#3 MeSH descriptor: [Fallopian Tube Neoplasms] explode all trees
#4 ((peritoneal or ovarian or fallopian) near/1 (neoplasm* or cancer* or tumo* or carcinoma* or adenocarcinoma*)):ti,ab,kw (Word variations have been searched)
#5 #1 or #2 or #3 or #4
#6 (qaly* or qald* or qale* or qtime* or daly):ti,ab,kw (Word variations have been searched)

#7 ((disability or quality) near (adjusted life year*)):ti,ab,kw
#8 sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six or sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve or sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen or sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty:ti,ab,kw
#9 euroqol or euro qol or eq5d or eq 5d or hql or hqol or h qol or hrqol or hr qol or hye or hyes or health* year* equivalent* or health utilit* or hui or hui1 or hui2 or hui3:ti,ab,kw
#10 MeSH descriptor: [Quality-Adjusted Life Years] explode all trees
#11 #5 or #6 or #7 or #8 or #9 or #10
#12 #5 and #11
#13 MeSH descriptor: [Tumor Markers, Biological] explode all trees
#14 MeSH descriptor: [Biological Markers] explode all trees
#15 MeSH descriptor: [Mass Screening] explode all trees
#16 MeSH descriptor: [Early Detection of Cancer] explode all trees
#17 screen$ or test$ or imag$ or predict$ or surveillance:ti,ab,kw
#18 (earl* near (diagnos* or detect*)):ti,ab,kw
#19 #13 or #14 or #15 or #16 or #17 or #18
#20 #5 and #19
#21 #12 or #20

Web of Science

8 #7 AND #6 AND #1
7 TITLE: (Cost* or economic* or budget* or price or pricing or financial or finance or finances or financed or fee or fees or money or monetary)
6 TITLE: (((screen$ or tested or testing or test or tests or imag$ or predict$ or surveillance) or (earl$ NEAR (diagnos$ or detect$))))
5 #4 AND #1
4 #3 OR #2
3 TITLE: ((qaly* or qald* or qale* or qtime or daly$ or sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six or sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve or sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen or sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty or euroqol or euro qol or eq5d or eq 5d or hql or hqol or h qol or hrqol or hr qol or hye or hyes or (health* year* equivalent*) or health utilit* or hui or hui1 or hui2 or hui3))
2 TITLE: ((EORTC or QLQ-C30 or ((FACT or FLIC or RSCL or SDS or Fact-G or CNQ-SF or CNQ or CARES-SF or CARES or ESAS) NEAR/2 (cancer*))))
1 TITLE: ((peritoneal or ovarian or fallopian) NEAR/1 (neoplasm* or cancer* or tumo* or carcinoma* or adenocarcinoma*))

Medline and Embase RCTs

1. exp Ovarian Neoplasms/
2. Fallopian Tube Neoplasms/
3. Peritoneal Neoplasms/

4. (fallopian tube adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp.
5. (peritoneal adj (neoplasm* or cancer* or tumo* or carcinoma*)).mp.
6. (ovarian adj3 cancer*).mp.
7. (ovarian adj3 neoplas*).mp.
8. (ovarian adj3 tumo*).mp.
9. (ovarian adj3 carcinoma*).mp.
10. (ovarian adj3 adenocarcinoma*).mp.
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. Mass Screening/
13. "Early Detection of Cancer"/
14. screen*.mp.
15. 12 or 13 or 14
16. 11 and 15
17. limit 16 to yr="2011 -Current"
18. limit 17 to (human and english language)

CINAHL

S9 (S4 OR S5 OR S6) AND (S7 AND S8)
S8 S4 OR S5 OR S6
S7 S1 OR S2 OR S3
S6 TX screen*
S5 MJ detection
S4 MJ screening
S3 TX ((fallopian tube* or peritoneal or ovarian) N3 (neoplasm* or cancer* or tumo* or carcinoma* or adenocarcinoma))
S2 MH peritoneal neoplasms
S1 MH ovarian neoplasms

Web of Science

5 #2 AND #1
Refined by: PUBLICATION YEARS: (2013 OR 2011 OR 2012 OR 2014) AND DOCUMENT TYPES: (ARTICLE)
4 #2 AND #1
3 #2 AND #1
2 TI=(screen or screening or detect or detecting or detection or detects)
1 TS=((peritoneal or ovarian or fallopian) NEAR/1 (neoplasm* or cancer* or tumo* or carcinoma* or adenocarcinoma*))

Cochrane

#1 MeSH descriptor: [Ovarian Neoplasms] explode all trees
#2 MeSH descriptor: [Peritoneal Neoplasms] explode all trees
#3 MeSH descriptor: [Fallopian Tube Neoplasms] explode all trees
#4 ((peritoneal or ovarian or fallopian) near/1 (neoplasm* or cancer* or tumo* or carcinoma* or adenocarcinoma*)):ti,ab,kw (Word variations have been searched)
#5 MeSH descriptor: [Mass Screening] explode all trees
#6 MeSH descriptor: [Early Detection of Cancer] explode all trees
#7 screen*:ti,ab,kw
#8 (1 or 2 or 3 or 4) and (5 or 6 or 7)

Appendix 2: Details about the resource components, costs and resource use for ovarian cancer diagnosis, treatment and screening.

A2.1 Diagnosis of ovarian cancer

There are three main estimates of the stage-specific costs for diagnosing ovarian cancer. These are all based on the resource components (and their costs) reported by the INCISIVE report[82]. The three different costs are derived from three different estimates of resource use. The first estimate is that reported by the INCISIVE report. These estimates of resource components, cost and use were shared with clinicians in both Sheffield and Birmingham who provided alternative estimates. These estimates, and the overall stage-specific costs for diagnosing ovarian cancer are summarised in Table A2.1.

Table A2.1: Resource components, use and cost for the diagnosis of ovarian cancer, by stage.

Diagnosis: resource components	Cost	Stage I	Stage II	Stage III	Stage IV	Borderline / No OC
INCISIVE report						
CA-125 serum	£23	100%	100%	100%	100%	Not provided
Alfa fetoprotein	£23	10%	10%	10%	10%	Not provided
Beta human chorionic gonadotrophin	£23	10%	10%	10%	10%	Not provided
MRI of pelvis	£284	85%	85%	60%	10%	Not provided
CT scan	£128	85%	85%	75%	40%	Not provided
PET-CT	£423	20%	30%	60%	60%	Not provided
Total cost:		**£462**	**£505**	**£548**	**£361**	Not provided
Sheffield clinicians						
CA-125 serum	£23	100%	100%	100%	100%	100%
Alfa fetoprotein	£23	2%	2%	1%	1%	2%
Beta human chorionic gonadotrophin	£23	2%	2%	1%	1%	2%
MRI of pelvis	£179	10%	0%	0%	0%	30%
CT scan	£91	75%	100%	100%	100%	30%
PET-CT	£423	0%	0%	0%	0%	0%
Other marker	£23	0%	5%	50%	50%	30%
Total cost:		**£110**	**£116**	**£126**	**£126**	**£112**
Birmingham clinicians						
CA-125 serum	£23	95%	95%	98%	98%	Not provided
Alfa fetoprotein	£23	2%	95%	95%	95%	Not provided
Beta human chorionic gonadotrophin	£23	2%	95%	95%	95%	Not provided
MRI of pelvis	£179	0%	0%	0%	0%	Not provided
CT scan	£91	75%	100%	100%	100%	Not provided
PET-CT	£423	0%	0%	0%	0%	Not provided
Other marker	£23	0%	0%	50%	50%	Not provided
Total cost:		**£91**	**£157**	**£169**	**£169**	Not provided

OC: Ovarian cancer.

The stage-specific diagnosis costs derived from the INCISIVE report are two to five times greater than those based on expert clinical input. There are two main drivers for this increase in costs. They are an increased use of MRIs (which is reported to never be used in Birmingham and only used for 10% of women with stage I cancer in Sheffield) and PET-CT scans (which are used in neither Birmingham nor Sheffield). The main difference between the Sheffield and Birmingham estimates are that the latter include a higher use of both alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (beta-hCG).

Overall, values reported by Sheffield clinicians appear to be more in keeping with NICE guidance, which recommends a CT scan in preference to an MRI, and only recommends the use of AFP and beta-hCG amongst women below the age of 40. For the base-case analysis, the stage-specific diagnosis costs were based on estimates provided by Sheffield clinicians. The two alternative cost estimates (from the INCISIVE report and Birmingham clinicians) were used in sensitivity analyses. When these were used it was assumed that values for women with either borderline ovarian cancer or no ovarian cancer were the same as values for women with stage 1 ovarian cancer.

A2.2 Treatment of ovarian cancer

The two main treatments for ovarian cancer are surgery and chemotherapy. Surgery can include a number of different procedures, whilst chemotherapy can include different regimens.

Estimates of the number of women who receive surgery and/or chemotherapy were available from the English Cancer Registries. Estimates were sought from Sheffield and Birmingham clinicians about the resource use (in terms of surgical procedures and chemotherapy regimens) amongst women who received surgery and/or chemotherapy. These estimates were then combined with data on resource use by stage at presentation from the English cancer registries[14] to obtain estimates of treatment cost by stage at diagnosis.

A2.2.1 Cancer registry data

Data on treatment resource use were provided by Public Health England[14]. Data are from the English cancer registries, and cover all women diagnosed between 2008 and 2010 inclusive. No treatment data were available for 103 individuals (85 with epithelial ovarian cancer and 18 with borderline ovarian cancer), so these were excluded from all subsequent analyses, resulting in a dataset of 16,972 women. Data were broken down by type of ovarian cancer (epithelial by stage, or borderline), with imputation of missing stage data as described in section 2.2. The available data distinguished between both the type of treatment received (chemotherapy or surgery) and the order in which the treatments were received (for women who received both chemotherapy and surgery). Differences in the order of treatment received were not considered for this study, as they were unlikely to result in a difference in costs. In other words, the cost of surgery (or chemotherapy) was assumed to be the same regardless of if it occurred before or after chemotherapy (or surgery). Rates of treatment by stage are presented in Table A2.2. Rates for borderline cancers are not

presented. The majority of women diagnosed with borderline ovarian cancer were treated with just surgery (83%), 7% received both surgery and chemotherapy, 1% received just chemotherapy and the remaining 9% did not receive treatment.

Table A2.2: Type of treatment received as a proportion, by stage. Epithelial ovarian cancers, diagnosed 2008 to 2010.

Stage	No Treatment	Chemo, no surgery	Surgery, no chemo	Chemo and surgery	Total
I	7%	4%	41%	48%	100%
II	12%	7%	17%	64%	100%
III	21%	22%	8%	50%	100%
IV	47%	30%	4%	19%	100%

Chemo: chemotherapy.

There is substantial variation in the type of treatment received. Combination treatment (both surgery and chemotherapy, in either order) is the most common treatment, with the exception of the most advanced stage (IV). For this stage the most frequent treatment option is no treatment (47%). The proportion of women receiving surgery (either on its own or with chemotherapy) is highest for the least advanced stage (89%). This proportions decreases monotonically with increasing stage, to 23% for stage IV. There is little clear pattern, in the proportion of women receiving chemotherapy (either on its own, or in combination with surgery).

A2.2.2 Estimates of resource use and cost

The types of surgical procedures and chemotherapy regimens considered, along with initial estimates of their costs were shared with clinical teams from both Sheffield and Birmingham. The subsequent estimates of resource use, and the corresponding stage-specific costs are provided in Table A2.3.

Table A2.3 Resource components, use and cost amongst women receiving treatment for ovarian cancer, by stage.

Diagnosis: resource components	Cost	Stage I	Stage II	Stage III	Stage IV	Borderline / No OC
Sheffield clinicians						
Surgery						
Pelvic and peritoneal washings and biopsies	£1,909	5%	5%	5%	5%	5%
Total hysterectomy	£2,861	95%	95%	95%	95%	10%
Bilateral salpingo-oophorectomy	£1,418	100%	100%	95%	95%	100%
Infracolic-omentectomy	£1,909	100%	100%	100%	100%	100%
Pelvic and para-aortic lymph node sampling	£2,059	5%	5%	5%	5%	2%
Ultra-radical (extensive) surgery	£4,100	0%	0%	2%	2%	0%
Retroperitoneal lymph node dissection	£3,975	0%	0%	2%	2%	0%
Total cost:		**£6,243**	**£6,243**	**£6,334**	**£6,334**	**£3,750**
Chemotherapy						
Carboplatin	£2,655	100%	86%	61%	63%	N/A
Paclitaxel with carboplatin	£6,600	0%	14%	28%	21%	N/A
Bevacizumab	£36,078	0%	0%	11%	16%	N/A
Total cost:		**£2,655**	**£3,219**	**£7,465**	**£8,763**	**£0**
Birmingham clinicians						
Surgery						
Pelvic and peritoneal washings and biopsies	£1,909	5%	5%	5%	5%	5%
Total hysterectomy	£2,861	95%	95%	95%	95%	10%
Bilateral salpingo-oophorectomy	£1,418	100%	100%	95%	95%	100%
Infracolic-omentectomy	£1,909	100%	100%	100%	100%	100%
Pelvic and para-aortic lymph node sampling	£2,059	5%	5%	5%	5%	2%
Ultra-radical (extensive) surgery	£4,100	0%	0%	2%	2%	0%
Retroperitoneal lymph node dissection	£3,975	0%	0%	2%	2%	0%
Total cost:		**£6,243**	**£6,243**	**£6,334**	**£6,334**	**£3,750**
Chemotherapy						
Carboplatin	£2,655	100%	50%	19%	19%	N/A
Paclitaxel with carboplatin	£6,600	0%	50%	76%	76%	N/A
Bevacizumab	£36,078	0%	0%	5%	5%	N/A

Total cost:		**£2,655**	**£4,628**	**£7,252**	**£7,252**	**£0**

N/A: Not applicable.

In addition, Sheffield clinicians commented that 5% of women with Stage III or IV ovarian cancer would receive diagnostic laparoscopy but no further treatment. Diagnostic laparoscopy was assumed to be represented in NHS reference costs by the HRG code MA27Z, giving a cost of £1,184. These estimates of total cost for surgery and for chemotherapy were combined with the stage-specific proportions of women who received surgery and/or chemotherapy (presented in Table A2.2), to obtain stage-specific costs. These stage-specific costs were combined with evidence from the UKCTOCS on the distributions of the stage of ovarian cancer at diagnosis for each screening arm[7], to calculate treatment costs for each screening arm.

A2.2.3 Alternative estimates from the INCISIVE report

The INCISIVE report produced, based on expert clinical opinion, estimates of surgical procedures and chemotherapy regimens used as a proportion of all women with ovarian cancer. However, as these resources were not mutually exclusive (for example, women could receive both a total hysterectomy and a BSO), it was unclear how these figures related to resource use as a proportion of women who actually received surgery or chemotherapy. Because of this it was decided that it was not possible to incorporate the INCISIVE report estimates into the model (it would not have been appropriate to directly use the estimates, which are reported as a proportion of all diagnosed women as it was anticipated that screening would have a differential impact on the numbers of women diagnosed and treated).

Data from the English cancer registries combined with estimates from Sheffield clinicians was used in the base-case analysis. The use of estimates from Birmingham clinicians was used in a sensitivity analysis.

A2.3 Screening for ovarian cancer

Resource components and use for ovarian cancer screening were based on those described for the UKCTOCS trial[6]. The resource components for MMS were CA-125 tests interpreted using the ROCA, and type 2 TVS. For USS type 1 and type 1 TVS were used.

No costings were available from the UKCTOCS, so these were taken from the literature. It was assumed that first-level TVS (undertaken by type 1 sonographers) cost the same as an outpatient gynaecological ultrasound scan lasting 20 minutes or longer. The most recent cost estimate for this was £55, taken from 2012/13 NHS reference costs[81]. This cost was inflated, using HCHS inflation indices[84] to give a 2013/14 value of £55.82. It was assumed, as described in section 6.1, that second-level TVS (undertaken by type 2 sonographers) would cost 21% more than a first-level TVS and so these were costed at £67.55. For CA-125 tests there are no standard costs. As described in section 6.1, an estimate derived from expert opinion used in NICE clinical guidance may be used and inflated to give a 2013/14 cost of £25.02 As the costs of implementing the ROCA are unknown it was assumed that this would equate to an additional £35 per screen.

Overall costs per completed screen and per partial screen (for those who drop-out mid screening) were required for the economic model. These resource use (total number of level 1 and level 2 scans, along with repeat scans) were based on data presented in Figure 2 and 3 of Menon *et al*[6]. For completed screens, resource use was adjusted to account for the reported levels of drop-out. For example, in Figure 2 of Menon *et al* it was reported that of the 50,078 original level 1 scans for MMS, 8.6% (4,315) required a repeat scan but only 8.2% (4,121) attended their scan and of these first repeat scans a further 24.5% (1,008/4,121) required a second repeat scan but only 23.8% (979/4,121) attended. The number of repeat scans after adjusting for drop-out were then first repeat: 4,315 (8.6% x 50,078) and second repeat 1,055 (24.5% x 4,315). Based on these calculations, the average number of level 1, level 2 and repeat scans required per completed screening episode for both types of screening are presented in Table A2.4.

Table A2.4 Average number of level 1, level 2 and repeat screens per completed screening episode.

MMS screening	Cost	N	%
Level 1 CA-125 screen	£60.02	50,078	100%
Level 1 - first repeat	£60.02	4,315	8.6%
Level 1 - second repeat	£60.02	1,055	2.1%
Level 2 screen	£67.55	430	0.9%
Level 2 - repeat	£67.55	164	0.3%
Overall cost	**£67.26**		
US screening	**Cost**	**N**	**%**
Level 1 scan	£55.82	48,230	96%
Level 1 - first repeat	£55.82	3,005	6.0%
Level 2 screen	£67.55	2,898	5.8%
Level 2 - first repeat	£67.55	47	0.1%
Overall cost	**£61.09**		

The majority of drop-outs (205/266 for MMS and 328/356 for USS) occurred after the initial level-one screen. Therefore only the cost of this was used for drop-outs, giving an average cost of £60.02 for MMS and £55.82 for USS.

To generate values for the PSA, a standard error of £16.36 was calculated for level 1 TVS, based on the difference between upper and lower NHS reference costs quartiles[81]. It was assumed that this standard error also applied to level 2 TVS, with values for the two sampled from Gamma distributions. PSA values for ROCA and CA-125 were sampled from beta-distributions with the following characteristics: ROCA (mean £35, minimum £5, maximum £100), CA-125 (mean £25, minimum £10, maximum £50).

Appendix 3: Methods for extrapolating survival data.

As discussed in Section 8.4.3, three different extrapolation methods were considered. They were:

1) Use of Royston-Parmar (R-P) models to estimate time-varying hazard ratios for each of the two active screening arms compared to the no screening arm (for which a time-varying hazard was estimated). Use of time-series methods to extrapolate these hazard ratios.
2) Use of standard parametric models, with separate model structures allowed for the three screening arms.
3) Use of standard parametric models, with the same model structure used for all three screening arms.

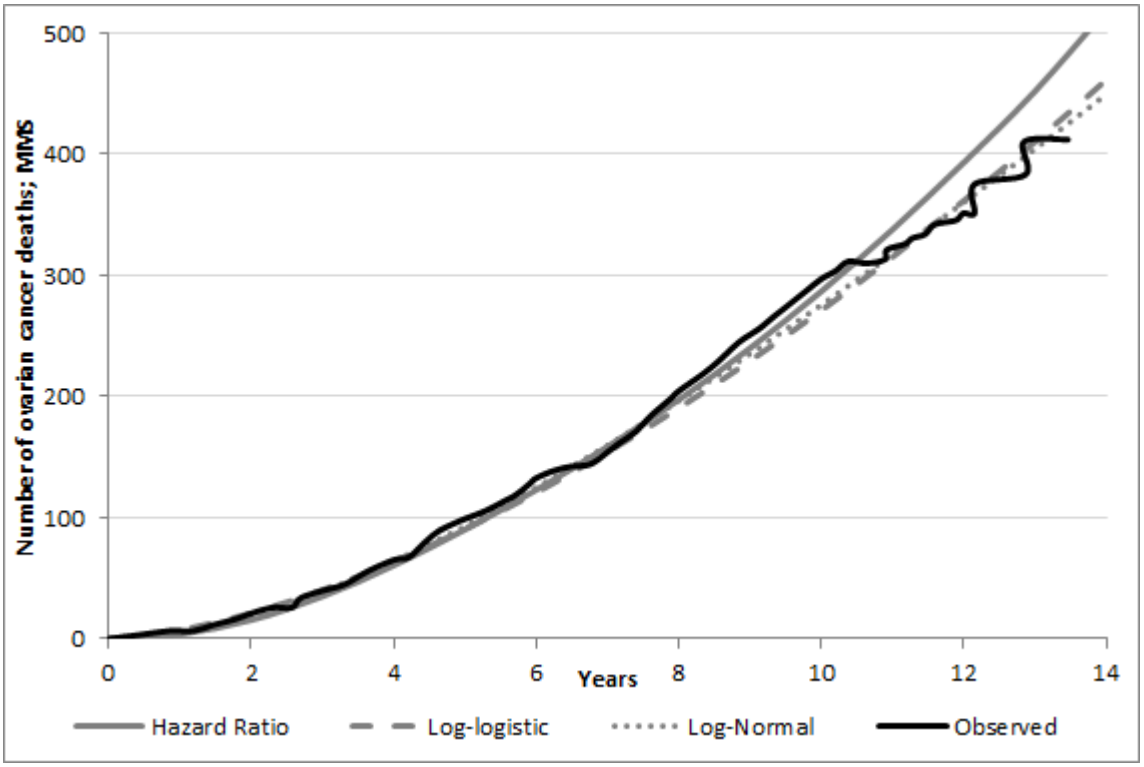
Five potential standard parametric models were considered (exponential, Weibull, Gompertz, log-logistic and log-Normal). The choice between these was based on minimising the Bayesian Information Criteria (BIC), although differences in the Akaike's information criteria (AIC) were also noted. For method 2), the model choice was based on minimising the combined BIC for all three screening arms. It should be noted that under this method all three screening arms had the same model type, but there was no constraint on the parameters for any given model. Models were fit in STATA MP version 14[105].

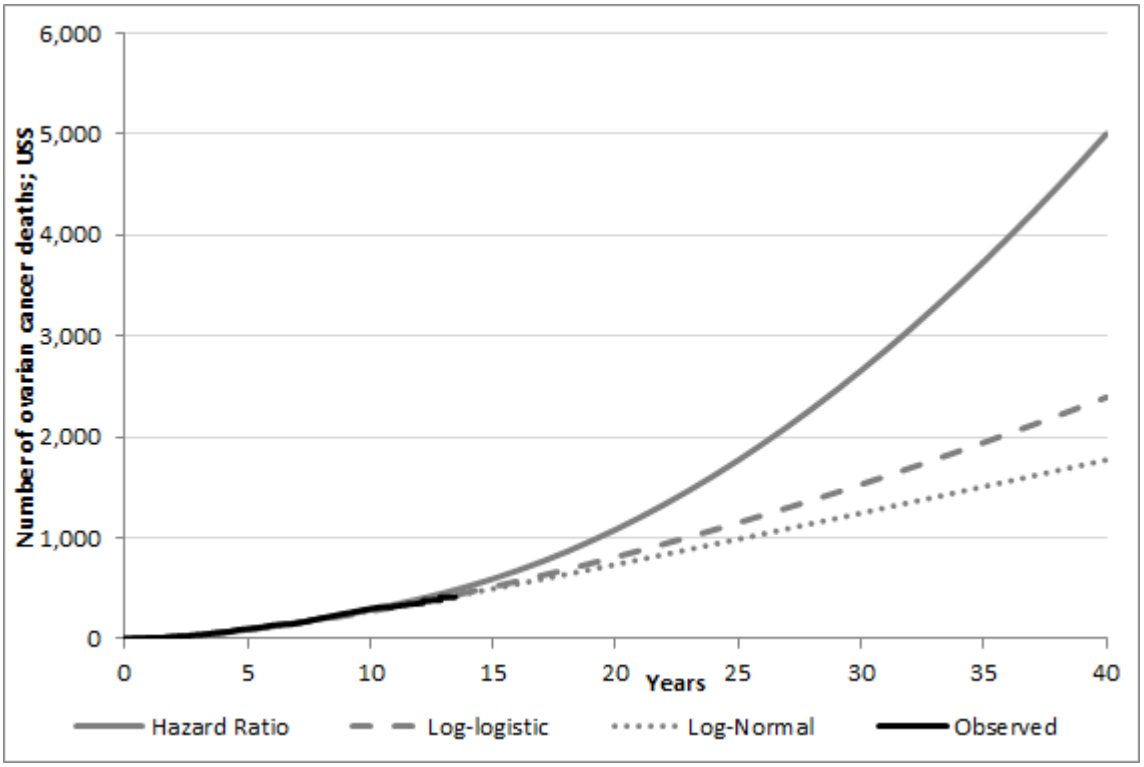
For the R-P models, choices are required about the number of knots to use for both the underlying hazard, and the time-varying hazard ratios. Three were used for both the underlying hazard and the hazard ratios. This number of knots was chosen as it resulted in similar model estimates to those reported in the UKTOCS publication[7]. Models were fit in STATA MP version 14[105], using the stpm2 function. The time-series method considered for this study was exponential smoothing (ES). Other methods, such as autoregressive integrated moving average models are available. However, ES models were used, as these are simpler, more robust and easier to interpret[94]. Forecasts were obtained using the forecast package in R[95]. This allows for the automatic fitting of ES models. These models can estimate both trend, and a 'dampening' parameter. The choice between ES models was based on the default option of minimising the AIC. Neither hazards nor hazard ratios can be negative. To accommodate this constraint, the natural logarithm of the data was taken prior to forecasting, with the exponential of the resulting forecasts taken. The resulting forecasts of the hazard for no screening suggested that it was constant for the majority of the extrapolated period.

This was judged to lack face validity (as the hazard is likely to increase due to an ageing cohort), and so when extrapolating the hazard for no screening a damped parameter was not estimated.

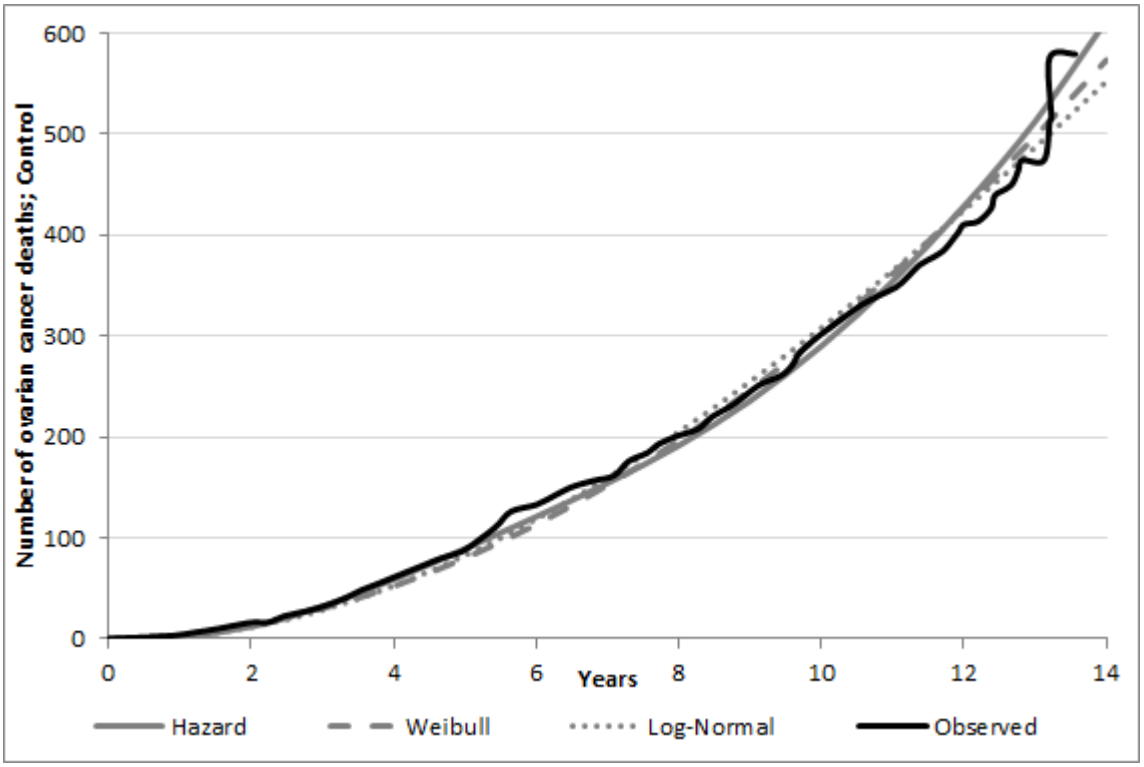
For the base-case the first method was used, with the other three methods considered in sensitivity analyses. A comparison of the within-trial estimates, and the resulting extrapolations, are displayed for both USS and no screening in Figure A3.1 to A3.4 (graphs for MMS are displayed in Section 8.4.3).

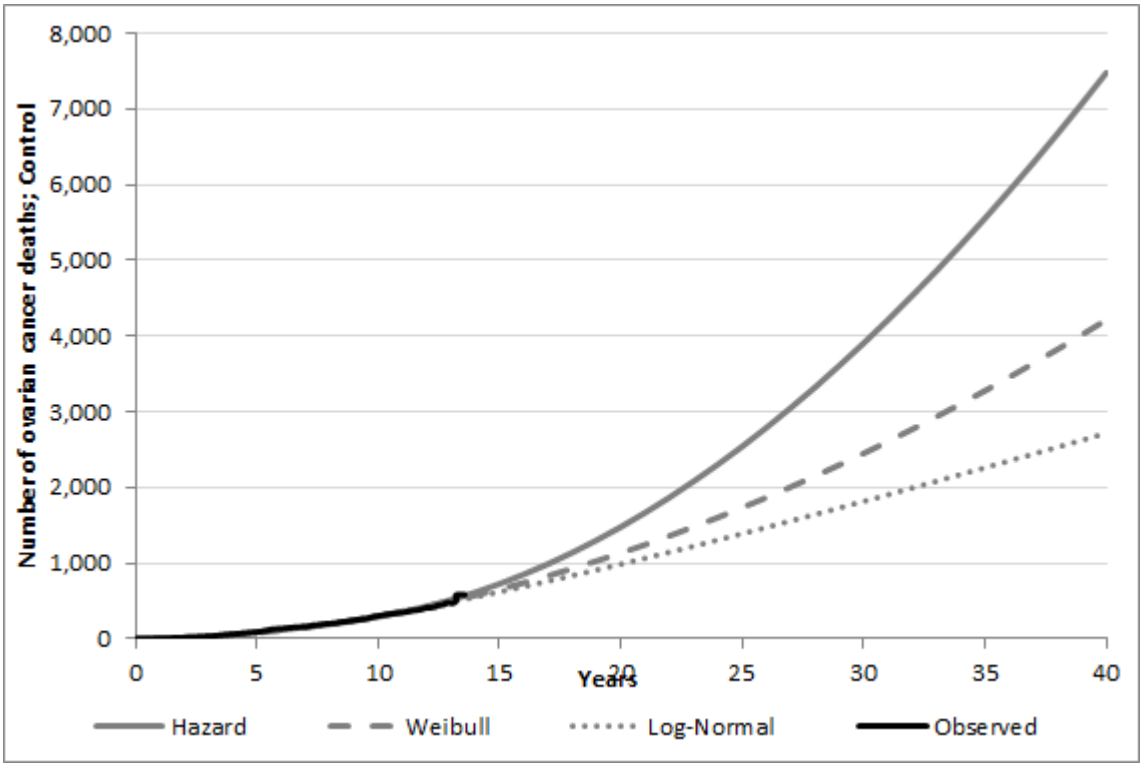
Figures A3.1 and A3.2: Ultrasound screening: comparison of model estimates and observed data for mortality: for the trial period (Figure A3.1, below), and for a lifetime horizon (Figure A3.2, over-page).





Figures A3.3 and A3.4: No screening: comparison of model estimates and observed data for mortality: for the trial period (Figure A3.3, below), and for a lifetime horizon (Figure A3.4, over-page).




Number of ovarian cancer deaths; Control
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
0
5
10
15
20
25
30
35
40
Years
Hazard
Weibull
Log-Normal
Observed

Appendix 4: Additional cost-effectiveness results.

Figure A4.1 **Estimated hazard ratio for ultrasound screening over-time: top-pane within trial estimates, bottom-pane over lifetime.**

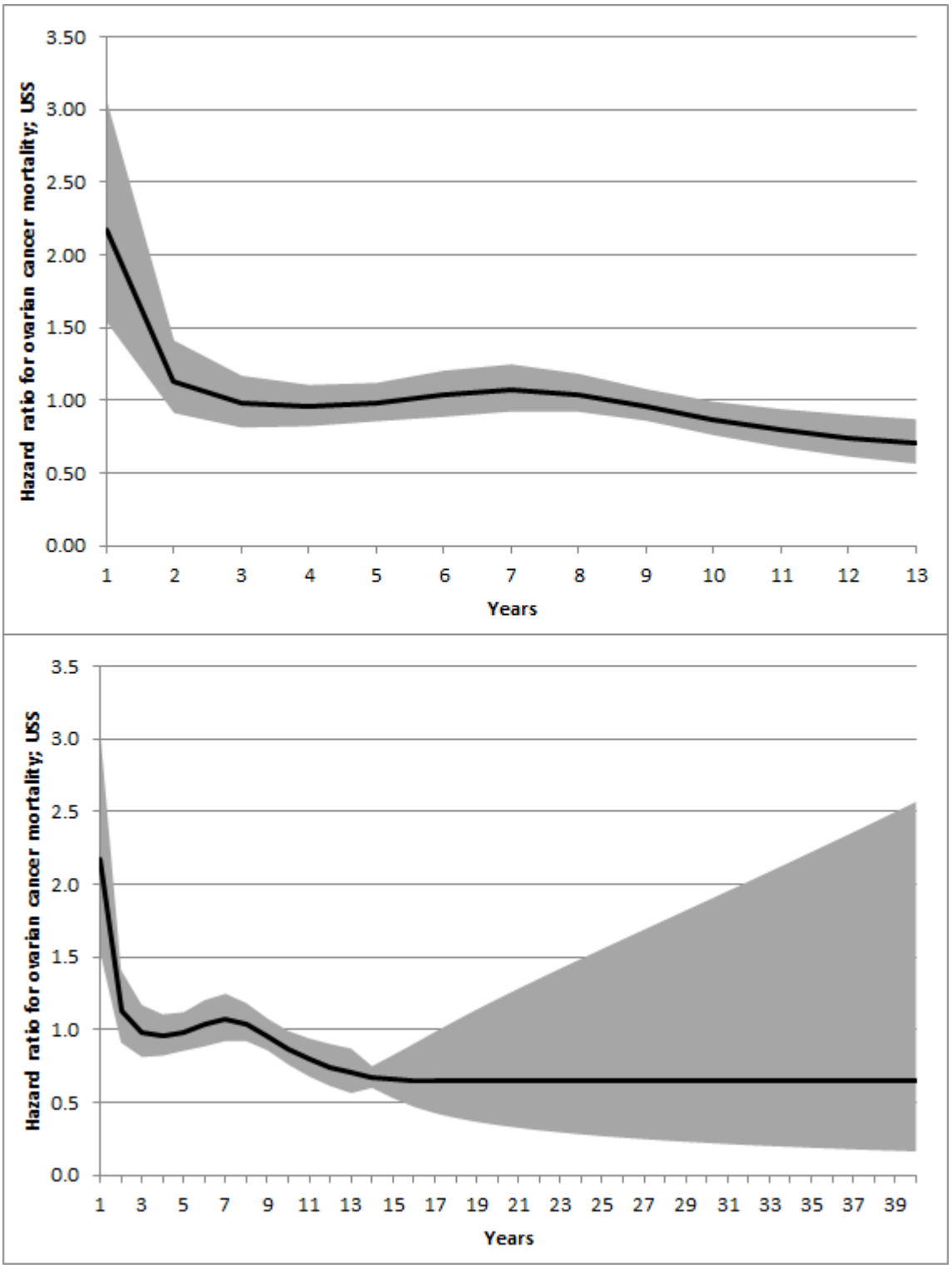


Figure A4.2: Cumulative ovarian cancer mortality for ultrasound screening

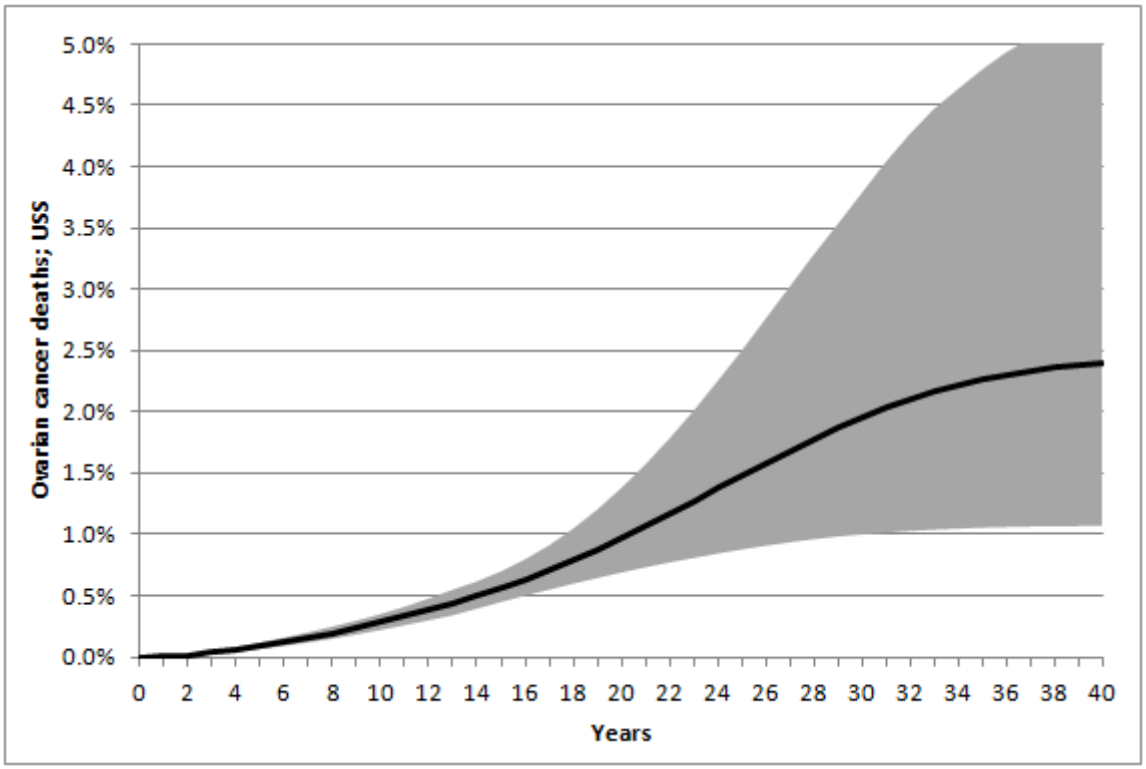
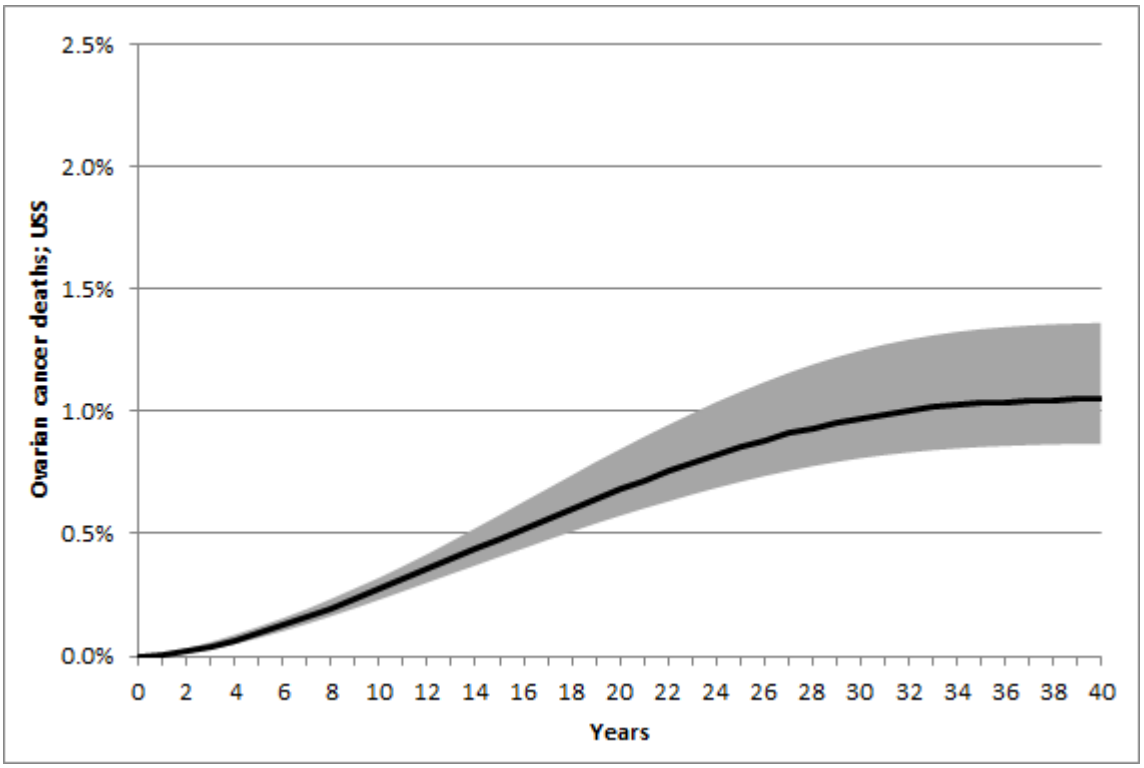


Figure A4.3: Cumulative ovarian cancer mortality for ultrasound screening; separate parametric models for each trial arm (log-Normal for ultrasound)



The log-Normal distribution is always used for USS, regardless of if the parametric models are restricted to be the same across all three trial arms.

Figure A4.4: **Cumulative ovarian cancer mortality for no screening; top-panel: separate parametric models for each trial arm (Weibull); bottom-panel: the same parametric model for each trial arm (log-Normal)**

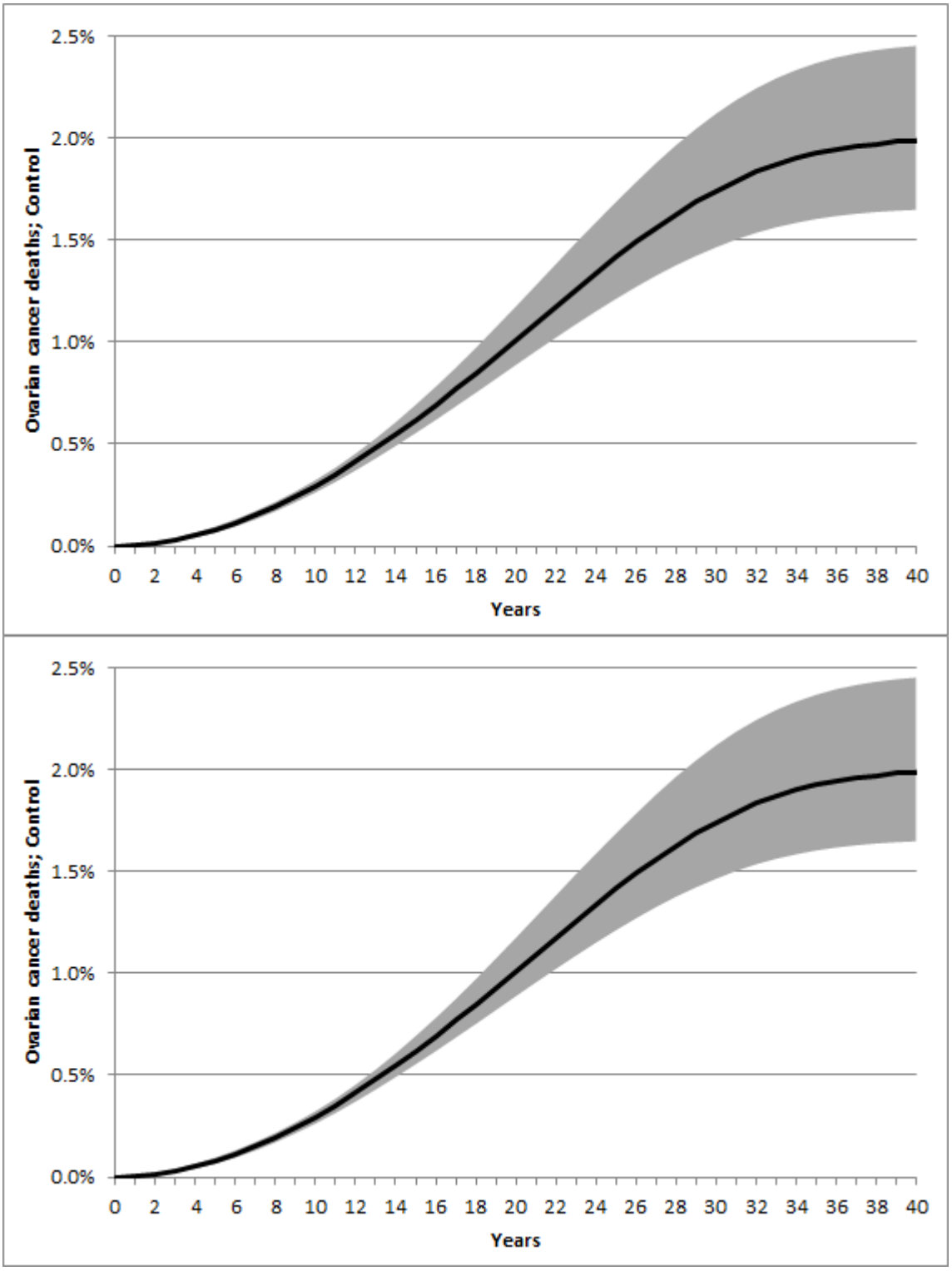
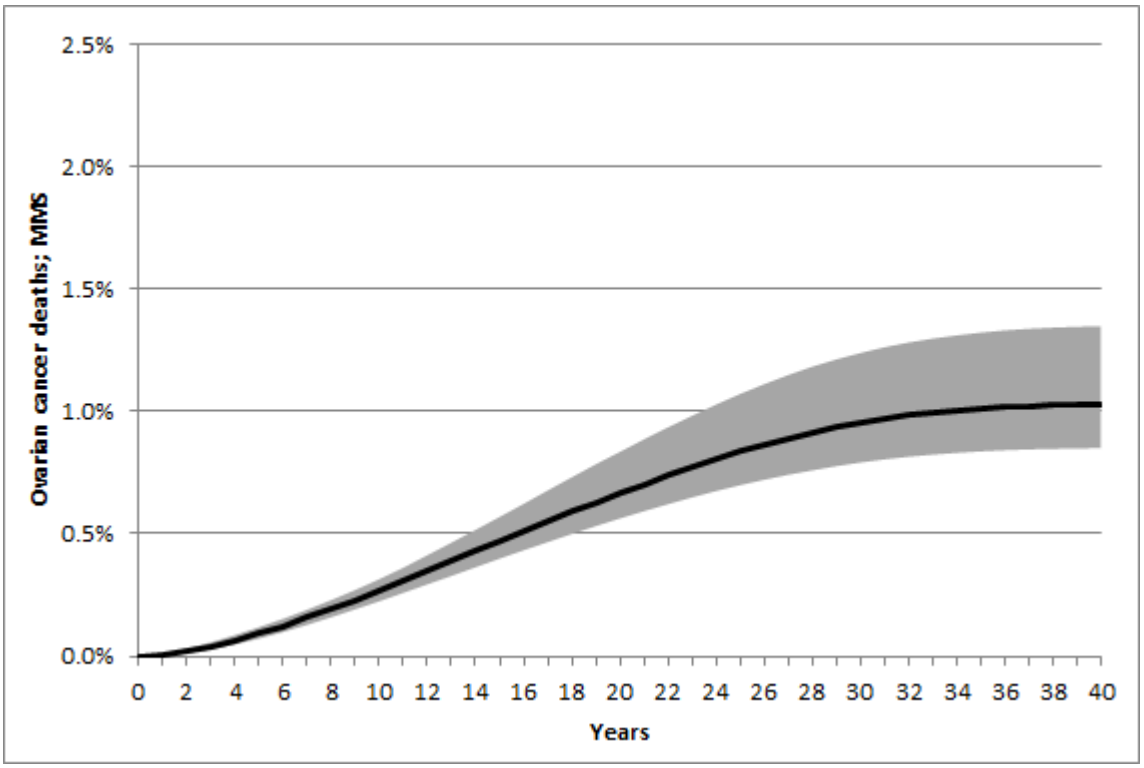


Figure A4.5: Cumulative ovarian cancer mortality for multimodal screening; separate or same parametric models for each trial arm (log-Normal for multimodal)



The log-Normal distribution is always used for MMS, regardless of if the parametric models are restricted to be the same across all three trial arms.

Figure A4.6: **Cumulative ovarian cancer mortality for multimodal screening (top panel) and ultrasound screening (bottom panel); model discrepancy approach: cumulative shrinkage of 1% per year.**

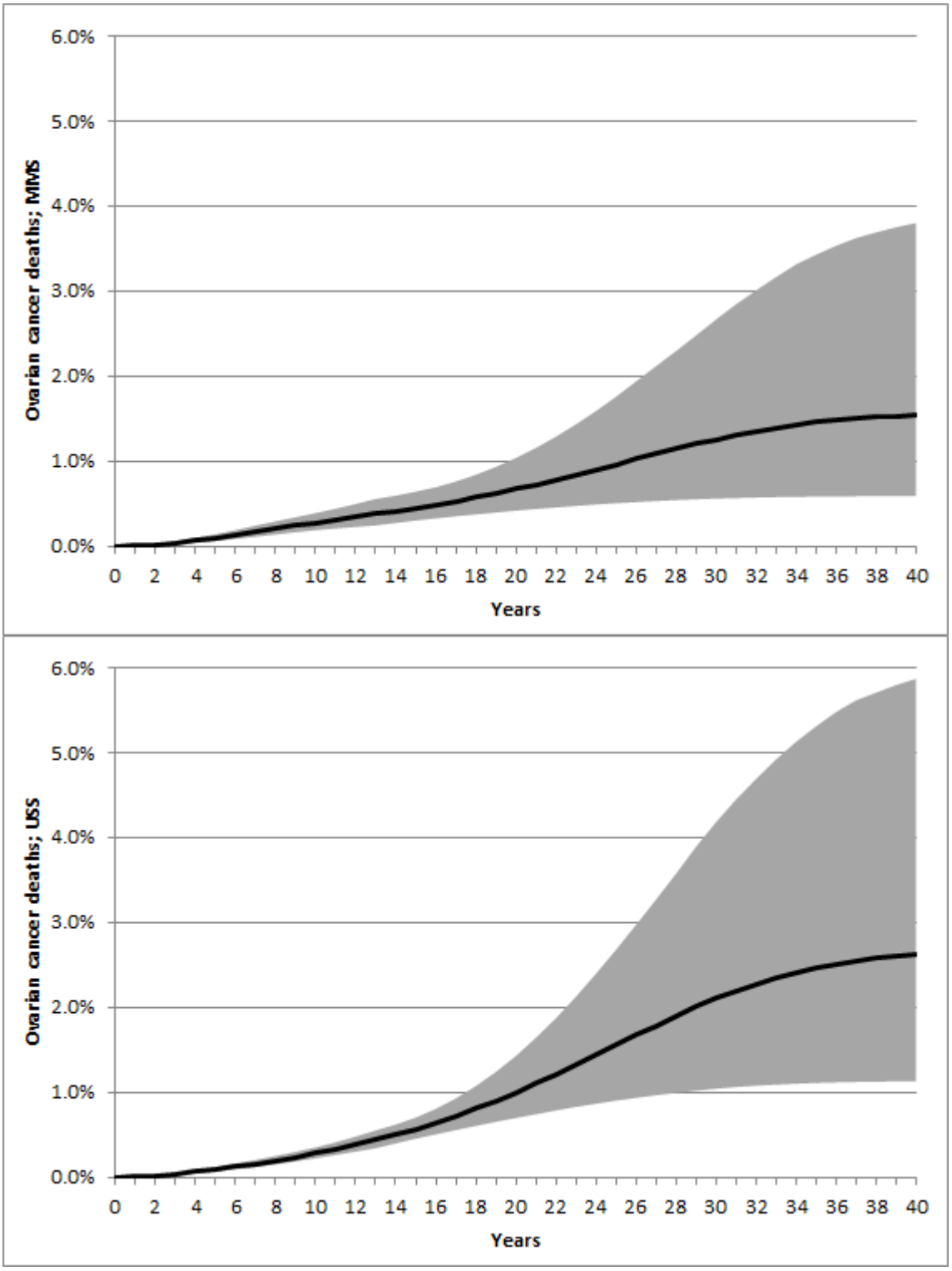
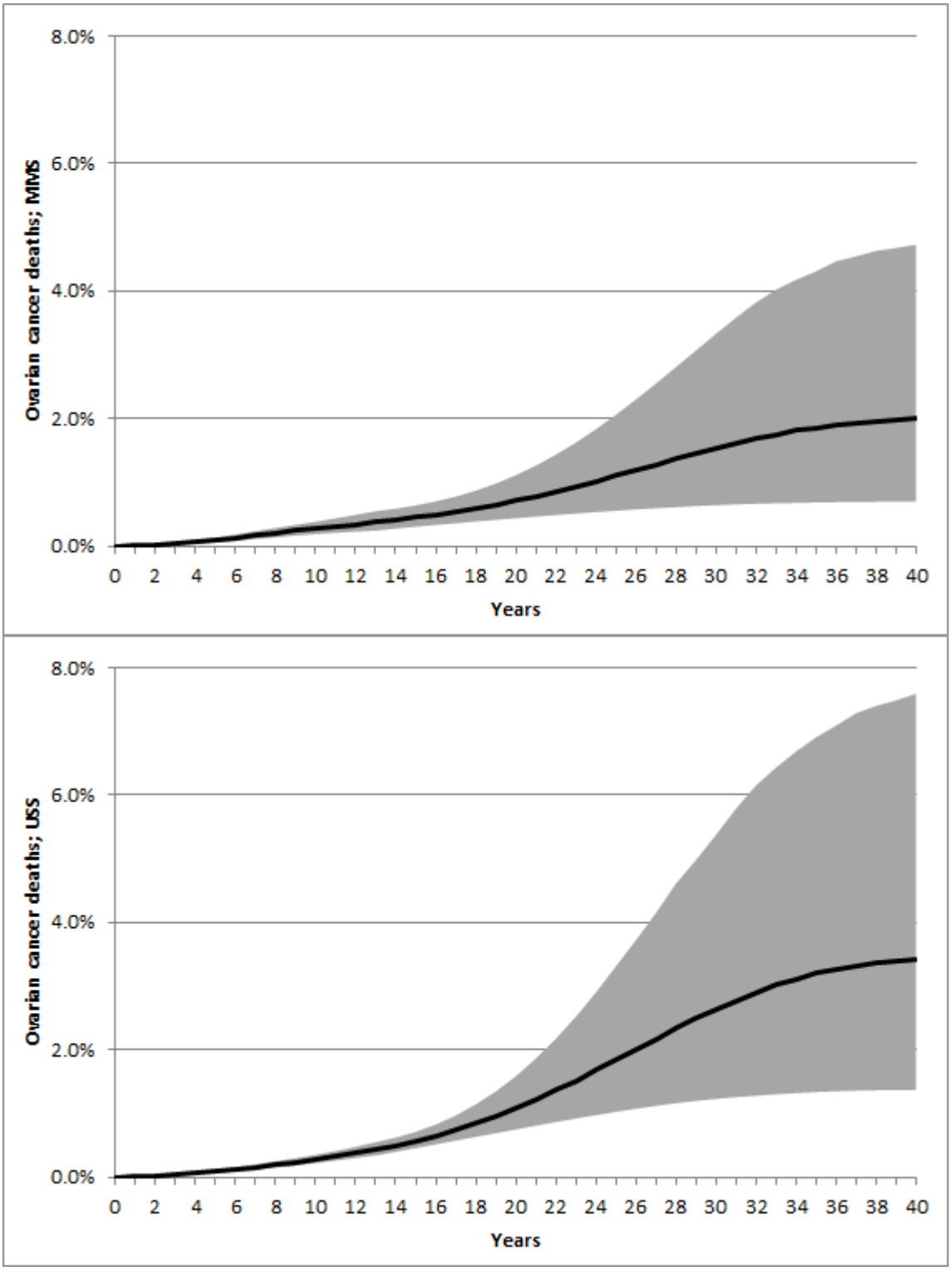


Figure A4.6: **Cumulative ovarian cancer mortality for multimodal screening (top panel) and ultrasound screening (bottom panel); model discrepancy approach: cumulative shrinkage of 5% per year.**



The model discrepancy approach does not alter estimates for no screening, so is only presented for MMS and USS.

Appendix 5: Consolidated health economic evaluation reporting standards (CHEERS) checklists for the three unique economic evaluations.

Havrilesky *et al* 2008[53].

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Neither, although only considers hypothetical screening strategies.
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Abstract, apart from perspective.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Introduction section.
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Methods, 1st paragraph.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Not explicitly mentioned.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Page 182 (below key assumptions).
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen	Last paragraph on page 181.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Results, 1st paragraph.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 182, no justification given.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Not explicitly stated.

Measurement of effectiveness	11a	*Single study-based estimates*: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	N/A – hypothetical effectiveness.
	11b	*Synthesis-based estimates*: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	
Measurement and valuation of preference based outcome	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	N/A
Estimating resources and costs	13a	*Single study-based economic evaluation*: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
	13b	*Model-based economic evaluation*: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Page 182 and Table 2.
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate	Page 182.
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Page 180 and Figure 1.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Pages 180 to 182.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Pages 180 to 182.
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Tables 1 and 2, text in pages 180 to 182.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Table 4.

Characterising uncertainty	20a	*Single study-based economic evaluation*: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	
	20b	*Model-based economic evaluation*: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Table 4 and Figure 4. Also pages 183 to 185.
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	N/A
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Page 186.
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 187 (role of funder not specified).
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Page 187.

Drescher *et al* 2012[12].

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Title
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Unstructured.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Introduction.
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Methods – overview and natural history component. Page 1016.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Introduction.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Not stated.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen	Methods – screening component. Pages 1018 to 1019.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Methods – natural history component. Page 1016.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Methods – overview. Page 1016. No justification given.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Methods – overview. Page 1016.
Measurement of effectiveness	11a	*Single study-based estimates*: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Methods – screening component. Pages

			1018 to 1019 and supplementary material.
	11b	*Synthesis-based estimates*: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	
Measurement and valuation of preference based outcome	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	N/A
Estimating resources and costs	13a	*Single study-based economic evaluation*: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
	13b	*Model-based economic evaluation*: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Methods: cost component and Table 1.
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate	Table 1.
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Materials and methods: overview. No figure provided.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Supplementary material
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Supplementary material
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Tables 1 and 2.
Incremental costs and	19	For each intervention, report mean values for the main categories of estimated costs and outcomes	Table 3.

outcomes		of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	
Characterising uncertainty	20a	*Single study-based economic evaluation*: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	
	20b	*Model-based economic evaluation*: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Supplementary table S1.
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	N/A
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Discussion, page 1022.
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 1023 (role of funder not described).
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Page 1023.

Wiwanitkit 2013[55].

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Title (although English poor).
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Structured abstract, although perspective, setting, methods and uncertainty analyses not provided.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Broader context in introduction, but study question and relevance not provided.
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Not stated
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Only state location (Thailand) in 'Materials and Methods'
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Not stated
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen	Introduction and 'Materials and

			Methods'
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Not stated
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Not stated
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Unclear
Measurement of effectiveness	11b	*Synthesis-based estimates*: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Not stated
Measurement and valuation of preference based outcome	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Not applicable
Estimating resources and costs	13b	*Model-based economic evaluation*: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	'Materials and Methods'
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate	Year not stated. Conversion rate in 'Materials and Methods'
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Not stated/provided
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Not stated
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Not stated
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Not perfomed/stated
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report	Tables 1 and 2 (incremental

		incremental cost-effectiveness ratios.	results not provided)
Characterising uncertainty	20b	*Model-based economic evaluation*: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Not stated
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Not provided
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Not stated
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Provided
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Provided

References

1. Trent Cancer Registry. Overview of Ovarian Cancer in England: Incidence, Mortality and Survival (November 2012). 2012.

2. National Cancer Intelligence Network. Cancer survival in England by stage. 2014.

3. Gentry-Maharaj A., Menon U. Screening for ovarian cancer in the general population. *Best Practice & Research Clinical Obstetrics & Gynaecology* 2012; 26(2):243-256.

4. Menon U., Griffin M., Gentry-Maharaj A. Ovarian cancer screening - current status, future directions. *Gynecologic Oncology* 2014; 132(2):490-495.

5. Reade C.J., Riva J.J., Busse J.W., Goldsmith C.H., Elit L. Risks and benefits of screening asymptomatic women for ovarian cancer: a systematic review and meta-analysis. *Gynecol Oncol* 2013; 130(3):674-681.

6. Menon U., Gentry-Maharaj A., Hallett R., Ryan A., Burnell M., Sharma A. et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *The Lancet Oncology* 2009; 10(4):327-340.

7. Jacobs I.J., Menon U., Ryan A., Gentry-Maharaj A., Burnell M., Kalsi J.K. et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet* 2015.

8. Cancer Research UK. Types of ovarian cancer. 2014.http://www.cancerresearchuk.org/cancer-help/type/ovarian-cancer/about/types-of-ovarian-cancer

9. Scottish Intercollegiate Guidelines Network. SIGN 135: Management of Epithelial Ovarian Cancer. 2013.

10. National Institute for Health and Clinical Excellence. Recognition and initial management of ovarian cancer. 2005.

11. Maringe C., Walters S., Butler J., Coleman M.P., Hacker N., Hanna L. et al. Stage at diagnosis and ovarian cancer survival: Evidence from the International Cancer Benchmarking Partnership. *Gynecologic Oncology* 2012; 127(1):75-82.

12. Drescher C.W., Hawley S., Thorpe J.D., Marticke S., McIntosh M., Gambhir S.S. et al. Impact of screening test performance and cost on mortality reduction and cost-effectiveness of multimodal ovarian cancer screening (Structured abstract). *Cancer Prevention Research* 2012; 5:1015-1024.

13. Lim A.W.W., Mesher D., Gentry-Maharaj A., Balogun N., Widschwendter M., Jacobs I. et al. Time to diagnosis of Type I or II invasive epithelial ovarian cancers: a multicentre observational study using patient questionnaire and primary care records. *BJOG: An International Journal of Obstetrics & Gynaecology* 2015.

14. Public Health England. Data from the English Cancer Registries. 2015.

15. The World Health Organisation. International Classification of Diseases - ICD 10. 2015.www.who.int/whosis/icd10/

16. Lalwani N., Shanbhogue A.K., Vikram R., Nagar A., Jagirdar J., Prasad S.R. Current update on borderline ovarian neoplasms. *American Journal of Roentgenology* 2010; 194(2):330-336.

17. Sherman M.E., Mink P.J., Curtis R., Cote T.R., Brooks S., Hartge P. et al. Survival among women with borderline ovarian tumors and ovarian carcinoma. *Cancer* 2004; 100(5):1045-1052.

18. Croswell J.M., Ransohoff D.F., Kramer B.S. Principles of cancer screening: lessons from history and study design issues. *Seminars in Oncology* 2010; 37(3):202-215.

19. van Nagell J.R., DePriest P.D., Ueland F.R., DeSimone C.P., Cooper A.L., McDonald J.M. et al. Ovarian cancer screening with annual transvaginal sonography. *Cancer* 2007; 109(9):1887-1896.

20. Ferlay J., Soerjomataram I., Dikshit R., Eser S., Mathers C., Rebelo M. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* 2015; 136(5):E359-E386.

21. Jayson G.C., Kohn E.C., Kitchener H.C., Ledermann J.A. Ovarian cancer. *The Lancet* 2014; 384(9951):1376-1388.

22. Sundar S., Neal R.D., Kehoe S. Diagnosis of ovarian cancer. *Bmj* 2015; 351:h4240.

23. Barclay M., Gildea C., Poole J., Hirschowitz L., Menon U., Nordin A. Factors Affecting Short-term Mortality in Women With Ovarian, Tubal, or Primary Peritoneal Cancer: Population-Based Cohort Analysis of English National Cancer Registration Data. *International Journal of Gynecological Cancer* 2016; 26(1):56-65.

24. National Cancer Intelligence Network. One, Five and Ten-Year Cancer Prevalence. 2010.

25. Office for National Statistics. National Life Tables, United Kingdom Statistical bulletins. 2016.http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/previousReleases

26. National Cancer Intelligence Network. Routes to Diagnosis. 2016.http://www.ncin.org.uk/publications/routes_to_diagnosis

27. Anderson M.C. Malignant potential of benign ovarian cysts: the case'for'. *Ovarian Cancer Biological and Therapeutic Challenges, Sharp, F , Mason, WP & Leake, RE (Eds)* 1990;187-190.

28. Crayford T.J., Campbell S., Bourne T.H., Rawson H.J., Collins W.P. Benign ovarian cysts and ovarian cancer: a cohort study with implications for screening. *The Lancet* 2000; 355(9209):1060-1063.

29. Kurman R.J., Shih I.M. The Origin and pathogenesis of epithelial ovarian cancer-a proposed unifying theory. *The American Journal of Surgical Pathology* 2010; 34(3):433.

30. Vang R., Shih I.M., Kurman R.J. Fallopian tube precursors of ovarian low-□and high-□grade serous neoplasms. *Histopathology* 2013; 62(1):44-58.

31. Jacobs I.J., Skates S.J., MacDonald N., Menon U., Rosenthal A.N., Davies A.P. et al. Screening for ovarian cancer: a pilot randomised controlled trial. *The Lancet* 1999; 353(9160):1207-1210.

32. Kobayashi H., Yamada Y., Sado T., Sakata M., Yoshida S., Kawaguchi R. et al. A randomized study of screening for ovarian cancer: a multicenter study in Japan. *International Journal of Gynecological Cancer* 2008; 18(3):414-420.

33. Buys S.S., Partridge E., Black A., Johnson C.C., Lamerato L., Isaacs C. et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *Jama* 2011; 305(22):2295-2303.

34. Karlan B.Y., Thorpe J., Watabayashi K., Drescher C.W., Palomares M., Daly M.B. et al. Use of CA125 and HE4 serum markers to predict ovarian cancer in elevated-risk women. *Cancer Epidemiology Biomarkers & Prevention* 2014; 23(7):1383-1393.

35. Barrett J., Jenkins V., Farewell V., Menon U., Jacobs I., Kilkerr J. et al. Psychological morbidity associated with ovarian cancer screening: results from more than 23 000 women in the randomised trial of ovarian cancer screening (UKCTOCS). *BJOG: An International Journal of Obstetrics & Gynaecology* 2014.

36. Parkes C.A., Smith D., Wald N.J., Bourne T.H. Feasibility study of a randomised trial of ovarian cancer screening among the general population. *Journal of Medical Screening* 1994; 1(4):209-214.

37. Tabor A., Jensen F.R., Bock J.E., HÃ¸gdall C.K. Feasibility study of a randomised trial of ovarian cancer screening. *Journal of Medical Screening* 1994; 1(4):215-219.

38. Taylor K.L., Shelby R., Gelmann E., McGuire C. Quality of life and trial adherence among participants in the prostate, lung, colorectal, and ovarian cancer screening trial. *Journal of the National Cancer Institute* 2004; 96(14):1083-1094.

39. Menon U., Skates S.J., Lewis S., Rosenthal A.N., Rufford B., Sibley K. et al. Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. *Journal of Clinical Oncology* 2005; 23(31):7919-7926.

40. Johnson D.B. The effects of an abnormal cancer screening test on health related quality of life. *International Journal of Cancer Research* 2006; 2(3):277-289.

41. Andersen M.R., Drescher C.W., Zheng Y., Bowen D.J., Wilson S., Young A. et al. Changes in cancer worry associated with participation in ovarian cancer screening. *Psycho-Oncology* 2007; 16(9):814-820.

42. Partridge E., Greenlee R.T., Xu J.L., Kreimer A.R., Williams C., Riley T. et al. Results from four rounds of ovarian cancer screening in a randomized trial. *Obstetrics and Gynecology* 2009; 113(4):775.

43. Pinsky P.F., Miller A., Kramer B.S., Church T., Reding D., Prorok P. et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *American Journal of Epidemiology* 2007; 165(8):874-881.

44. Menon U., Gentry-Maharaj A., Jacobs I. Ovarian cancer screening and mortality. *Jama* 2011; 306(14):1544-1545.

45. Black A., Buys S., Berg C. Ovarian Cancer Screening and Mortality—Reply. *Journal of the American Medical Association* 2011; 306(14):1544-1545.http://jama.jamanetwork.com/article.aspx?articleid=1731834

46. Pinsky P.F., Zhu C., Skates S.J., Black A., Partridge E., Buys S.S. et al. Potential effect of the risk of ovarian cancer algorithm (ROCA) on the mortality outcome of the Prostate, Lung, Colorectal and Ovarian (PLCO) trial. *International Journal of Cancer* 2013; 132(9):2127-2133.

47. Gupta S., ZAUBER A.G., VAN BALLEGOOIJEN M., KUNTZ K.M. Will Test-Specific Adherence Predict the Best Colorectal Cancer Screening Strategy? Authors' reply. *Annals of Internal Medicine* 2009; 150(5):359-360.

48. Kearns B., Whyte S., Seaman H.E., Snowball J., Halloran S.P., Butler P. et al. Factors associated with completion of bowel cancer screening and the potential effects of simplifying the screening test algorithm. *British Journal of Cancer* 2016.

49. Menon U., Gentry-Maharaj A., Ryan A., Sharma A., Burnell M., Hallett R. et al. Recruitment to multicentre trials- lessons from UKCTOCS: descriptive study. *Bmj* 2008; 337.

50. Szucs T.D., Wyss P., Dedes K.J., Szucs T.D., Wyss P., Dedes K.J. Cost-effectiveness studies in ovarian cancer. [Review] [47 refs]. 2003; 13 Suppl 2:212-219.http://onlinelibrary.wiley.com/store/10.1046/j.1525-1438.2003.13038.x/asset/j.1525-1438.2003.13038.x.pdf?v=1&t=i0ghzehw&s=26114a470c605edfa2ea65a55111cc0c72ab51cb

51. Manchanda R.B. Population based testing for BRCA mutations is highly cost-effective compared to family history based approach: A health-economic decision analytical model from the GCaPPS trial. 2013; Conference:188-189.

52. Ding Y.H. Cost-effectiveness analysis of multimodal screening for ovarian cancer. 2010; Conference:A37.

53. Havrilesky L.J., Sanders G.D., Kulasingam S., Myers E.R. Reducing ovarian cancer mortality through screening: is it possible, and can we afford it? (Structured abstract). 2008; 111:179-187.http://onlinelibrary.wiley.com/o/cochrane/cleed/articles/NHSEED-22009100148/frame.html http://ac.els-cdn.com/S0090825808005052/1-s2.0-S0090825808005052-main.pdf?_tid=5aa306fa-43f0-11e4-81ca-00000aab0f27&acdnat=1411566212_10017b85b5cd1a4748381d7bb3f0c676

54. Drescher C.W., Hawley S., Thorpe J.D., Marticke S., McIntosh M., Gambhir S.S. et al. Impact of Screening Test Performance and Cost on Mortality Reduction and Cost-effectiveness of Multimodal Ovarian Cancer Screening. *Cancer Prevention Research* 2012; 5(8).

55. Wiwanitkit V. CA-125 and risk of malignancy index for screening for malignancy in fertile aged females with ovarian cyst, which is more cost effectiveness? 2013; 34:72-73.http://www.ijmpo.org/article.asp?issn=0971-5851;year=2013;volume=34;issue=2;spage=72;epage=73;aulast=Wiwanitkit

56. Skates S.J., Menon U., Macdonald N., Rosenthal A.N., Oram D.H., Knapp R.C. et al. Calculation of the risk of ovarian cancer from serial CA-125 values for preclinical detection in postmenopausal women. *J Clin Oncol* 2003; 21(10 Suppl):206s-210s.PM:12743136

57. Urban N., Drescher C., Etzioni R., Colby C. Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening (Structured abstract). *Controlled Clinical Trials* 1997; 18(3).

58. Havrilesky L.J., Broadwater G., Davis D.M., Nolte K.C., Barnett J.C., Myers E.R. et al. Determination of quality of life-related utilities for health states relevant to ovarian cancer diagnosis and treatment. 2009; 113:216-220.<Go to ISI>://WOS:000265400400014

59. Myers E.R., Havrilesky L.J., Kulasingam S.L., Sanders G.D., Cline K.E., Gray R.N. et al. Genomic tests for ovarian cancer detection and management. 2006.

60. Havrilesky L.J., Sanders G.D., Kulasingam S., Chino J.P., Berchuck A., Marks J.R. et al. Development of an ovarian cancer screening decision model that incorporates disease heterogeneity. *Cancer* 2011; 117(3):545-553.

61. Menon U., Skates S.J., Lewis S., Rosenthal A.N., Rufford B., Sibley K. et al. Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. *J Clin Oncol* 2005; 23(31):7919-7926.PM:16258091

62. Skates S.J., Horick N., Yu Y., Xu F.J., Berchuck A., Havrilesky L.J. et al. Preoperative sensitivity and specificity for early-stage ovarian cancer when combining cancer antigen CA-125II, CA 15-3, CA 72-4, and macrophage colony-stimulating factor using mixtures of multivariate normal distributions. *J Clin Oncol* 2004; 22(20):4059-4066.PM:15381683

63. Skates S.J., Singer D.E. Quantifying the potential benefit of CA 125 screening for ovarian cancer. *Journal of Clinical Epidemiology* 1991; 44(4):365-380.

64. Urban N.D. Use of stochastic stimulation model to identify an efficient protocol for ovarian cancer screening. 1997; 18:251-270.

65. Abbas F.M., Sert M.B., Abbas F.M., Sert M.B. Cost, quality of life and outcome measures in ovarian cancer. [Review] [86 refs]. 1998; 9:859-867.

66. Luckett T., King M., Butow P., Friedlander M., Paris T., Luckett T. et al. Assessing health-related quality of life in gynecologic oncology: a systematic review of questionnaires and their ability to detect clinically important differences and change. [Review]. 2010; 20:664-684.

67. Chase D.M., Wenzel L., Chase D.M., Wenzel L. Health-related quality of life in ovarian cancer patients and its impact on clinical management. [Review]. 2011; 11:421-431.

68. Friedlander M.L., King M.T., Friedlander M.L., King M.T. Patient-reported outcomes in ovarian cancer clinical trials. [Review]. 2013; 24 Suppl 10:x64-x68.http://annonc.oxfordjournals.org/content/24/suppl_10/x64.full.pdf

69. Donovan K.A., Donovan H.S., Cella D., Gaines M.E., Penson R.T., Plaxe S.C. et al. Recommended Patient-Reported Core Set of Symptoms and Quality-of-Life Domains to Measure in Ovarian Cancer Treatment Trials. 2014; 106:i.http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2012642797&site=ehost-live

70. Montazeri A., McEwen J., Gillis C.R., Montazeri A., McEwen J., Gillis C.R. Quality of life in patients with ovarian cancer: current state of research. [Review] [60 refs]. 1996; 4:169-179.

71. Jones G.L.L. The impact of treatment for gynecological cancer on health-related quality of life (HRQoL): A systematic review. 2006; 194:26-42.

72. Grzankowski K.S., Carney M., Grzankowski K.S., Carney M. Quality of life in ovarian cancer. [Review]. 2011; 18:52-58.

73. Lockwood-Rayermann S. Survivorship issues in ovarian cancer: a review. 2006; 33:553-562.http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2009169502&site=ehost-live

74. Kiss N.S. Quality of life and patient preferences in platinum sensitive ovarian cancer. 2012; Conference:A429.

75. Roland K.B., Rodriguez J.L., Patterson J.R., Trivers K.F. A literature review of the social and psychological needs of ovarian cancer survivors. 2013; 22:2408-2418.http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2012346461&site=ehost-live

76. Hess L.M., Stehman F.B., Hess L.M., Stehman F.B. State of the science in ovarian cancer quality of life research: a systematic review. [Review]. 2012; 22:1273-1280.

77. Ferrell B., Cullinane C.A., Ervin K., Melancon C., Uman G.C., Juarez G. Perspectives on the impact of ovarian cancer: women's views of quality of life. 2005; 32:1143-1149.http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2009059356&site=ehost-live

78. Mirabeau-Beale K.L., Kornblith A.B., Penson R.T., Lee H., Goodman A., Campos S.M. et al. Comparison of the quality of life of early and advanced stage ovarian cancer survivors. 2009; 114:353-359.http://ac.els-cdn.com/S0090825809003527/1-s2.0-S0090825809003527-main.pdf?_tid=8c031598-43d0-11e4-b53b-00000aab0f6b&acdnat=1411552551_101f2675b4941dd91c544606d066615a

79. Holman L.L., Lu K.H., Bast Jr R.C., Hernandez M.A., Bodurka D.C., Skates S. et al. Risk perception, worry, and test acceptance in average-risk women who undergo ovarian cancer screening. 2014; 210:257.http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2012486744&site=ehost-live

80. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. 2013.http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9

81. Department of Health. NHS reference costs 2012 to 2013. 2015.https://www.gov.uk/government/publications/nhs-reference-costs-2012-to-2013

82. Incisive Health. Saving lives, averting costs. 2014.

83. NHS Careers. Agenda for change - pay rates. 2015.http://www.nhscareers.nhs.uk/working-in-the-nhs/pay-and-benefits/agenda-for-change-pay-rates/

84. Curtis L. PSSRU Unit Costs of Health & Social Care 2013. *University of Kent (United Kingdom)* 2013.

85. Whyte S., Chilcott J., Halloran S. Reappraisal of the options for colorectal cancer screening in England. *Colorectal Disease* 2012; 14(9):e547-e561.

86. Guest J.F., Ruiz F.J., Greener M.J., Trotman I.F. Palliative care treatment patterns and associated costs of healthcare resource use for specific advanced cancer patients in the UK. *European Journal of Cancer Care* 2006; 15(1):65-73.

87. Tappenden P., Chilcott J., Brennan A., Squires H., Stevenson M. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. *Value in Health* 2012; 15(8):1127-1136.

88. Kehoe S., Hook J., Nankivell M., Jayson G.C., Kitchener H., Lopes T. et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *The Lancet* 2015.

89. Kang S. Neoadjuvant chemotherapy for ovarian cancer: do we have enough evidence? *The Lancet* 2015; 386(9990):223-224.

90. Whyte S., Harnan S. Effectiveness and cost-effectiveness of an awareness campaign for colorectal cancer: a mathematical modeling study. *Cancer Causes & Control* 2014; 25(6):647-658.

91. Whyte S., Walsh C., Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Medical Decision Making* 2011; 31(4):625-641.

92. Guyot P., Ades A.E., Ouwens M.J., Welton N.J. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology* 2012; 12(1):9.

93. Royston P., Parmar M.K. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 2002; 21(15):2175-2197.

94. Chatfield C. The analysis of time series: an introduction. 2013.

95. Hyndman R.J., Khandakar Y. Automatic time series for forecasting: the forecast package for R. 2007.

96. Briggs A., Sculpher M., Claxton K. Decision modelling for health economic evaluation. 2006.

97. Strong M., Oakley J.E., Brennan A., Breeze P. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample A Fast Nonparametric Regression-Based Method. *Medical Decision Making* 2015;0272989X15575286.

98. The Office for National Statistics. Annual Mid-year Population Estimates. 2016.http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest

99. Rustin G.J., van der Burg M.E., Griffin C.L., Guthrie D., Lamont A., Jayson G.C. et al. Early versus delayed treatment of relapsed ovarian cancer (MRC OV05/EORTC 55955): a randomised trial. *The Lancet* 2010; 376(9747):1155-1163.

100. Aaronson N.K., Ahmedzai S., Bergman B., Bullinger M., Cull A., Duez N.J. et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute* 1993; 85(5):365-376.

101. Tappenden P., Chilcott J.B. Avoiding and Identifying Errors and Other Threats to the Credibility of Health Economic Models. *Pharmacoeconomics* 2014; 32(10):967-979.

102. McCabe C., Claxton K., Culyer A.J. The NICE cost-effectiveness threshold. *Pharmacoeconomics* 2008; 26(9):733-744.

103. Claxton K., Martin S., Soares M., Rice N., Spackman E., Hinde S. et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. 2015.

104. Newcombe R.G. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine* 1998; 17(8):857-872.

105. StataCorp. Stata Statistical Software: Release 14. 2015.