Evidence Synthesis
Number 157

Screening for Ovarian Cancer: An Updated Evidence Review for the U.S. Preventive Services Task Force

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. HHSA-290-2015-00007-I-EPC5, Task Order No. 2

Prepared by:
Kaiser Permanente Research Affiliates Evidence-based Practice Center
Kaiser Permanente Center for Health Research
Portland, OR

Investigators:
Jillian T. Henderson, PhD
Elizabeth M. Webber, MS
George F. Sawaya, MD

AHRQ Publication No. 17-05231-EF-1
February 2018
Structured Abstract

**Importance:** Ovarian cancer, while not common, is the fifth-leading cause of cancer death among United States women. In 2012 the U.S. Preventive Services Task Force (USPSTF) determined that harms of ovarian cancer screening outweighed benefits based on trial evidence, and recommended against screening average-risk women.

**Objective:** To update the previous systematic review and inform USPSTF ovarian cancer screening guidance.

**Data Sources:** MEDLINE, PubMed, and the Cochrane Collaboration Registry of Controlled Trials from January 1, 2003, through January 31, 2017, and prior literature identified in the previous review conducted for the USPSTF.

**Study Selection:** English-language trials of benefits and harms of screening for ovarian cancer in average-risk women reporting health outcomes (e.g., mortality and quality of life). Interventions compared with the control condition were transvaginal ultrasound screening alone, ultrasound screening with cancer antigen 125 (CA-125) testing, and CA-125 screening alone—either with a single measurement threshold value or measures of change over time.

**Data Extraction and Synthesis:** Two investigators independently reviewed abstracts and full-text articles, and then extracted data from fair- and good-quality trials.

**Main Outcomes and Measures:** Ovarian cancer mortality and incidence (defined as ovarian, fallopian tube, and peritoneal cancer), ovarian cancer survival, harms associated with false positive test results, false positive surgery, screening and surgical complications.

**Results:** Four RCTs (n = 293,587) were included; three reported ovarian cancer mortality (KQ1) and all reported potential harms of screening (KQ2). Three trials were rated good-quality and the small trial (n= 549) reporting only on psychological harms of screening was rated fair-quality. Two trials were conducted in the United States and two in the United Kingdom, primarily with postmenopausal, average-risk women. The Prostate, Lung, Colorectal and Ovarian (PLCO) (n = 68,557) included 4-6 rounds of annual CA-125 (≥35 U/mL threshold) and transvaginal ultrasound screening, with up to 13 years of trial data. The U.K. Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) (n = 202,546) included 7–11 rounds of either annual transvaginal ultrasound screening or CA-125 screening using the Risk of Ovarian Cancer Algorithm with up to 14 years of trial data. A smaller U.K. Pilot trial (n = 21,935) included three rounds of annual screening with CA-125 (≥30 U/mL threshold) and up to 8 years of trial data. In all three screening trials, there was not a statistically significant difference in ovarian cancer mortality associated with screening. Mortality estimates from the PLCO (RR =1.18 [95% CI, 0.82 to 1.71]) or in either arm of the UKCTOCS: ultrasound (HR = 0.91 [95% CI, 0.76 to 1.09]) and CA-125 (HR = 0.89 [95% CI, 0.74 to 1.08]) were based on more rounds of screening and larger study populations. Harms of screening in these two large screening trials included surgical investigations among screen-positive women without cancer, which ranged from 1 percent of trial participants without cancer screened with CA-125 testing in the UKCTOCS, and 3.2 percent for the ultrasound arm of the UKCTOCS and in the PLCO screening intervention. Serious
surgical complications of occurred for just over 3 percent of women without cancer in the
UKCTOCS intervention arms, and in 15 percent of women in the PLCO intervention arm. In the
two largest trials, cumulative false-positive rates ranged from 9.8% to 44%. Evidence on
psychological harms was limited but nonsignificant, except in the case of repeat followup scans
and tests, which increased the risk of psychological morbidity in a subsample of the UKCTOCS
participants based on the General Health Questionnaire 12 (score ≥4) (OR 1.28 [95% CI, 1.18 to
1.39]).

Conclusions and Relevance: Since the previous review for the USPSTF, results from a large
trial conducted in the United Kingdom were published. Ovarian cancer mortality did not differ
between control and intervention screening conditions in any of the included trials, including two
good-quality studies with adequate power to detect differences. Harms of screening include
surgery following a false-positive test, often resulting in removal of one or both ovaries and/or
fallopian tubes, and the potential for major surgical complications. Reports from the UKCTOCS
of a potential delayed effect of screening on ovarian cancer mortality require further followup
data to evaluate, but the causal mechanism for a delayed screening effect is unclear. Major trials
of promising ovarian cancer screening tools have null findings to date among healthy average-
risk women, and there are considerable harms associated with screening.
# Table of Contents

## Chapter 1. Introduction
- Definition of Ovarian Cancer ................................................................. 1
- Prevalence and Burden ........................................................................... 1
- Etiology and Natural History ................................................................. 2
- Risk Factors ......................................................................................... 2
- Prevention .......................................................................................... 3
- Diagnosis and Treatment ...................................................................... 3
- Rationale for Screening ........................................................................ 4
- Screening Strategies ............................................................................ 4
- Current Clinical Practice in the United States ...................................... 5
- Previous USPSTF Recommendation .................................................... 6

## Chapter 2. Methods
- Scope and Purpose ................................................................................ 7
- Key Questions and Analytic Framework .............................................. 7
- Data Sources and Searches ................................................................ 7
- Study Selection .................................................................................... 7
- Quality Assessment and Data Abstraction ........................................... 8
- Data Synthesis and Analysis ................................................................. 9
- Grading the Strength of the Body of Evidence ..................................... 10
- Expert Review and Public Comment .................................................. 11
- USPSTF Involvement .......................................................................... 12

## Chapter 3. Results
- Description of Included Studies ............................................................. 13
- Included Populations ........................................................................... 13
- Results of Included Studies ................................................................ 13
  - KQ 1. Does Screening for Ovarian Cancer in Asymptomatic Women Using Single Tests or Combined Algorithms Reduce All-Cause or Disease-Specific Morbidity and Mortality? .................................................. 13
  - KQ 2. What Are the Harms of Screening for Ovarian Cancer, Including Harms of the Screening Test and of Diagnostic Evaluation? ................................................................................. 17

## Chapter 4. Discussion
- Summary of Evidence .......................................................................... 20
- Further Discussion of UKCTOCS ......................................................... 20
- Implications of Stage Shift and Treatment Findings in the Absence of a Mortality Benefit .................................................................................. 23
- Discussion of Harms ............................................................................ 24
- Limitations of the Review .................................................................. 25
- Future Research Needs ......................................................................... 27
- Conclusion .......................................................................................... 29

## References

### Figure
- Figure 1. Analytic Framework and Key Questions
Tables
Table 1. Characteristics of Ovarian Cancer Screening Randomized Trials
Table 2. Baseline Characteristics of Ovarian Cancer Screening Trial Participants
Table 3. Ovarian Cancer Screening Trial Protocols
Table 4. Adherence to Screening in Ovarian Cancer Screening Trials
Table 5. Benefits Reported in Ovarian Cancer Screening Trials: Ovarian Cancer Mortality
Table 6. Harms Reported in Ovarian Cancer Screening Trials: Positive Testing, False-Positive Testing, and Surgical Complications in Ovarian Cancer Screening
Table 7. Harms Reported in Ovarian Cancer Screening Trials: Psychological Effects of Screening
Table 8. Summary of Evidence

Appendixes
Appendix A. Detailed Methods
Appendix B. Included Studies
Appendix C. Excluded Studies
Appendix D. Benefits Reported in Ovarian Cancer Screening Trials: Ovarian Cancer Mortality in Analyses Excluding Peritoneal Cancer
Chapter 1. Introduction

Definition of Ovarian Cancer

Ovarian cancer includes cancers of the ovary, fallopian tubes, and peritoneum due to the origination from similar tissue types and similar clinical management and treatment.\(^1\) Epithelial ovarian cancer is classified into five subtypes, based on histology to identify the origin and degree of differentiation: high-grade serous carcinoma, endometrioid carcinoma, clear cell carcinoma, low-grade serous carcinoma, and mucinous carcinoma.\(^2\) Most ovarian cancers are included in these subtypes, with the remainder being rare non-carcinoma types include germ cell tumors, sarcomas, sex cord-stromal tumors. The most common subtype is serous carcinoma, and this subtype comprises the majority of advanced stage cancers.\(^3\) The staging of ovarian cancer follows National Cancer Institute definitions describing the extent of tissue involvement and spread with the terms in situ, localized, regional, distant (outside the peritoneum), and unknown.\(^4\) Detailed staging categories for ovarian cancer established by the International Federation of Gynecology and Obstetrics (FIGO) describe the specific characteristics used to assign stages I through IV, as well as staging within these levels.\(^5, 5\) As understanding of ovarian tumor pathogenesis has deepened, distinctions between type I and type II ovarian cancers have been outlined.\(^6, 7\) Type I ovarian cancers tend to be early stage low grade tumors that generally arise through a progression from benign precursor lesions, tending to remain indolent or slowly progressive. The majority of cases are classified as type II ovarian cancers, characterized by rapid metastatic progression and accounting for the overwhelming majority of ovarian cancer deaths.\(^6\)

Prevalence and Burden

Ovarian cancer is the fifth most common cause of cancer death in U.S. women and the leading cause of gynecologic cancer deaths despite having low incidence.\(^8\) Approximately 22,440 ovarian cancer cases and 14,080 deaths are estimated to occur in 2017.\(^9\) According to data from the Surveillance, Epidemiology, and End Results Program (SEER) and National Program of Cancer Registries, the average annual age-adjusted incidence of ovarian cancer in the U.S. was 11.4 cases per 100,000 women for 2010-2014 with a mortality rate of 7.4 per 100,000 women. The incidence of ovarian cancer has declined slightly since the mid-1970s with an average of 1.9 percent a year over the last 10 years.\(^2, 10-12\) The majority of women diagnosed with ovarian cancer are over age 45 (88%), with a median age of diagnosis of 63 years. The average age-adjusted (2010-2014) annual incidence of ovarian cancer varies by race and ethnicity, occurring most frequently in white women (11.8 per 100,000) followed by Hispanic women (10.3 per 100,000). Rates are similar for black women (9.2 per 100,000 women), Asian/Pacific Islander (9.1 per 100,000), and lowest for American Indian/Alaska Native women (8.3 per 100,000), although estimates are less precise for this subpopulation.\(^5, 10\)
Etiology and Natural History

Ovarian cancers can originate from ovarian, fallopian, or other tissue types (e.g., endometrium, peritoneum).\(^2\)\(^-\)\(^13\) Historically, ovarian carcinomas were assumed to derive from the ovarian surface epithelium; however, evidence increasingly indicates that high-grade intraepithelial lesions in the fallopian tubes may become malignant and spread to the ovarian epithelium and peritoneum.\(^6\) Two broad categories defined by shared clinical features have been developed to better represent distinct models of epithelial ovarian carcinogenesis.\(^14\) Type I tumors include low-grade, generally indolent tumors, that are often associated with somatic mutations in a number of genes (e.g., \textit{KRAS}, \textit{BRAF}, \textit{ERBB2}) and develop from benign extra-ovarian lesions implanted on the ovary.\(^6,\)\(^15,\)\(^16\) Type II tumors are more likely to derive from the fallopian tube or ovarian surface epithelium. These cancers are generally high grade and are genetically unstable, including high rates of \textit{TP53} and \textit{BRCA} mutations.\(^6,\)\(^14\)

Overall, mortality from ovarian cancer is high, with fewer than half (46%) of women surviving for at least 5 years following an ovarian cancer diagnosis.\(^8\) By comparison, overall 5-year survival rates for cancers of the breast (90%), endometrium (80%), and cervix (70%) are much higher. The aggressive nature and commonly advanced stage at diagnosis translates to a poorer prognosis for type II ovarian cancers, with 30 percent of patients surviving at 5 years, compared to 55 percent of patients with type I cancers.\(^14\) Non-Hispanic black women have the lowest 5-year survival rates, and the second-highest mortality rates (7 per 100,000), slightly lower than the mortality rates observed among non-Hispanic white women (8 per 100,000).\(^17\) Ovarian cancer mortality increases with age at diagnosis with the highest rates of death among women 65 to 74 and a median age of death of 70 years.\(^8\)

Risk Factors

Multiple modifiable and nonmodifiable factors have been associated with an increased risk for developing ovarian cancer including: increasing age, family history of ovarian cancer, inherited genetic mutations (e.g., \textit{BRCA1}/\textit{BRCA2}), obesity, nulliparity, use of hormone replacement therapy, and increased numbers of lifetime ovulatory cycles.\(^2,\)\(^18\) Most risk factors show significant heterogeneity across ovarian cancer subtypes.\(^19\) Approximately 20 percent of ovarian cancers are familial, with the presence of cancer in multiple first- or second-degree relatives being an indicator of inherited cancer syndrome.\(^1\) Inherited mutations are associated with 5 to 15 percent of all ovarian carcinomas. The most common high-risk genetic syndromes include hereditary breast and ovarian cancer (i.e., \textit{BRCA1} and \textit{BRCA2}), Lynch, and Peutz-Jeghers syndromes.\(^2,\)\(^20-24\) In addition, genome-wide association studies have identified as many as 17 common low-penetrance alleles associated with ovarian cancer.\(^2,\)\(^25-30\) Women with a greatly increased risk for developing ovarian cancer, defined by the presence of germline genetic mutations, may benefit from risk-reducing surgery or chemoprevention. For these women genetic counseling is recommended, including a discussion of the risks and benefits of prevention.\(^31-33\)

Several factors have been identified that are associated with a decreased risk of ovarian cancer. Among the most well established is the use of oral contraceptives, with an estimated 20 percent
Risk-prediction models have been developed to identify women at increased risk of developing ovarian cancer based on personal and family history. To date, for women at average genetic risk, these tools have not been found to have a strong predictive performance, likely due to the relative rarity of ovarian cancer and the modest effect size of known risk factors.2, 41-44

**Prevention**

Risk-reducing surgery, such as bilateral salpingo-oopherectomy (BSO), has generally been advocated for women at high genetic risk,2, 45 and some evidence46, 47 from observational studies suggests that it may also be associated with a decreased risk of ovarian cancer for women at average or unknown genetic risk. The risk reduction conferred, however, is not 100 percent, and has been associated with side effects and potential risks including: early menopause, osteoporosis, cardiovascular disease, and increased overall mortality.2, 48 Bilateral salpingectomy, even with ovarian retention, may be effective in preventing ovarian cancer as there are subtypes postulated to arise in the fallopian tubes. Salpingectomy may allow high-risk women to delay removal of the ovaries, and when performed during a planned hysterectomy, may reduce risk for average-risk women.2, 49, 50 Based on evidence of the distal fallopian tube epithelium as the site of origin for at least some cancers the Society for Gynecologic Oncology (SGO) and the American Congress of Obstetricians and Gynecologists (ACOG) have issued statements recommending consideration of opportunistic bilateral salpingectomy to reduce ovarian cancer mortality in the general population.2, 51, 52 Tubal ligation and hysterectomy have also been associated with reduced risk of ovarian cancer; however, no groups have recommended these procedures as prevention strategies.2, 53

In addition to surgical intervention, hormone-modulating prescription drugs such as oral contraceptives have been investigated for prevention of ovarian cancer.2, 34 The SGO has stated that appropriate counseling about side effects and contraindications for oral contraceptive use can allow patients to weigh the risks and benefits of their use for cancer prevention.54 Of note, a 2013 systematic review for the AHRQ Effective Healthcare Program did not find sufficient evidence to recommend oral contraceptive use for the sole purpose of ovarian cancer prevention.55

**Diagnosis and Treatment**

Definitive diagnosis and staging of ovarian cancer requires surgery.1, 2, 32 Most women with
newly diagnosed ovarian cancer undergo primary staging and debulking surgery to remove as much of the visible tumor as possible. This surgery may include hysterectomy, BSO, and omentectomy.\textsuperscript{1,32} Younger patients with early (stage I or II) and/or low-risk tumors who wish to preserve fertility may opt for unilateral salpingo-oophorectomy.\textsuperscript{1} Survival is improved for women with complete tumor resection. In cases where total resection is not possible or contraindicated, neoadjuvant chemotherapy can reduce tumor size and facilitate later resection. Most women respond well to initial treatment; however, the majority will experience recurrence of the disease, requiring a cycle of repeated surgeries and chemotherapy cycles.\textsuperscript{1,2} Care delivered within high-volume hospitals and by gynecologic oncologists has been associated with guideline-adherent treatment and improved survival.\textsuperscript{2,56-58}

**Rationale for Screening**

The high mortality and low 5-year survival among all women diagnosed with ovarian cancer is largely due to challenges detecting the disease at an early stage. Only 15 percent of cases are diagnosed at the local stage, when 5-year survival is favorable at 92 percent. Over 60 percent of cases are diagnosed after the cancer has distant metastases. With distant spread, the 5 year survival drops to 29 percent.\textsuperscript{8} Thus, screening for early-stage disease has been a focus of research. Screening women for symptoms of ovarian cancer poses challenges; symptoms are often nonspecific, including bloating, pelvic or abdominal pain, urinary symptoms, vaginal discharge, increased vaginal bleeding, or gastrointestinal problems.\textsuperscript{1,2} These symptoms are often not seen as symptoms of serious illness by women or providers.\textsuperscript{2} As many as 95 percent of all women in primary care report one of the symptoms of ovarian cancer in the previous year, with 72 percent of women having recurring symptoms, most commonly back pain, fatigue, indigestion, urinary tract problems, constipation, and abdominal pain.\textsuperscript{59} While women found to have ovarian cancer appear to have symptoms more frequently and with a higher severity, the frequency of reported symptoms in unaffected women poses a challenge for clinical detection.\textsuperscript{59} Efforts to generate clinical decision tools based on the presence of combinations of symptoms, such as the Ovarian Cancer Screening Index,\textsuperscript{60} have been found to have a higher sensitivity than individual symptoms; however, estimates for the accuracy of these tools indicate that they are not sufficiently specific for implementation in clinical practice.\textsuperscript{61} Because of the lack of specific symptoms, research has investigated the use of other strategies for early detection, including the use of biomarkers and imaging technologies.

**Screening Strategies**

The most widely tested screening approaches, with reasonable test performance characteristics, have broadly focused on identifying abnormalities in the physical structure of the ovary or detecting increased CA-125 levels or trends. The most complex screening strategies involve algorithms that use CA-125 levels measured over time to compute the likelihood of ovarian cancer and determine surveillance and surgical investigation protocols at different risk thresholds.\textsuperscript{2,62-65} Transvaginal ultrasound (TVU) is the most widely used imaging technique for gynecologic symptoms and pathologies, but the majority of adnexal masses identified by TVU are benign.\textsuperscript{2,66}
Elevated CA-125 levels have been noted in women with an advanced ovarian carcinoma at diagnosis, leading to its proposed use as a biomarker for early detection.\textsuperscript{2} Some limitations with regard to the assay’s specificity and sensitivity have been recognized, as CA-125 may be markedly elevated in patients with a variety of benign or non-ovarian malignant conditions. In addition, serum CA-125 has been found to be significantly elevated in only half of women diagnosed with stage I or II ovarian cancer.\textsuperscript{2, 67} Efforts to improve the performance of screening with CA-125 led to development of the Risk of Ovarian Cancer Algorithm (ROCA\textsuperscript{TM} [Abcodia]). The algorithm uses sequential CA-125 measures taken at annual screening visits to evaluate the trajectory of CA-125 serum over time following a baseline age-adjusted CA-125 measurement.\textsuperscript{68}

Another screening strategy that continues to be practiced\textsuperscript{69} but is not supported by clinical evidence is ovarian palpation with bimanual pelvic examination. The accuracy of this screening examination (sensitivity 5.1\%)\textsuperscript{70} does not support its use.\textsuperscript{71, 72} The practice has been discouraged\textsuperscript{71} or recognized as lacking evidence to recommend\textsuperscript{72} as a routine screening examination for ovarian cancer due to its high false positive rate, low positive predictive value\textsuperscript{73, 74} and potential physical and psychological harms in the absence of benefits.

Research continues into the discovery of other biomarkers and the use of alternative imaging strategies for the early identification of ovarian cancer; however, no other markers have been implemented and tested in an ovarian cancer screening trial.\textsuperscript{2}

\textbf{Current Clinical Practice in the United States}

No organizations currently recommend screening for ovarian cancer in the general population. Several groups have issued guidelines that do not recommend screening in asymptomatic, average-risk women, including the American Academy of Family Physicians,\textsuperscript{75} American Cancer Society,\textsuperscript{76, 77} American College of Radiology,\textsuperscript{78} and American Congress of Obstetricians and Gynecologists.\textsuperscript{73} Nonetheless, a 2012 nationally representative sample of over 1,000 family physicians, general internists, and obstetrician/gynecologists found that over one-third of physicians believed that ovarian cancer screening was effective and up to one-fourth routinely offered TVU and/or CA-125 screening to asymptomatic women.\textsuperscript{79} Additionally, a 2014 survey of 1,555 U.S. family physicians, general internists, and obstetrician-gynecologists found that 27 percent of physicians overestimated the ovarian cancer risk among women at the same risk as the general population and 65 percent underestimated ovarian cancer risk among women at much higher risk than the general population.\textsuperscript{80}

In 2016 the United States Food and Drug Administration (FDA) released a recommendation against using tests marketed for ovarian cancer screening. The FDA stated that there are no ovarian cancer screening tests approved by the FDA and no published clinical information to demonstrate that currently available tests are accurate and reliable in asymptomatic women. In particular the FDA stated that the ROCA algorithm has been marketed in the United States with no data to support its claims for ovarian cancer detection and improved cancer survival.\textsuperscript{81} Following this statement from the FDA the company marketing the ROCA test suspended its commercial availability in the United States.\textsuperscript{82}
Previous USPSTF Recommendation

In 2012, the U.S. Preventive Services Task Force (USPSTF) recommended against screening for ovarian cancer in women (D recommendation). This recommendation applies to asymptomatic women without a known genetic mutation that increases their risk for ovarian cancer (e.g., BRCA mutations). There was adequate evidence that annual screening with TVU and testing for CA-125 in women does not reduce ovarian cancer mortality. In addition, the disease occurs infrequently enough that most women with a positive screening test results will have a false positive result; therefore, screening for ovarian cancer can lead to important harms, including major surgical interventions and complications in women who do not have cancer. The USPSTF concluded that there was at least moderate certainty that the harms of screening for ovarian cancer outweighed the benefits.
Chapter 2. Methods

Scope and Purpose

This systematic review addresses the benefits and harms of screening for ovarian cancer in women at average risk for ovarian cancer. The USPSTF will use this review to update its 2012 recommendation on this topic.74

Key Questions and Analytic Framework

We developed an Analytic Framework (Figure 1) and two Key Questions (KQs) to guide the literature search, data abstraction, and data synthesis.

1. Does screening for ovarian cancer in asymptomatic women using a single test or combined algorithm (such as, but not limited to, testing for serum cancer antigen [CA–125] and ultrasonography) reduce all-cause or disease-specific morbidity and mortality?
2. What are the harms of screening for ovarian cancer, including harms of the screening test and of diagnostic evaluation?

Data Sources and Searches

In addition to considering all studies from the previous reviews on this topic83-85 for inclusion in the current review, we performed a comprehensive search of MEDLINE, PubMed Publisher-Supplied Records, and the Cochrane Collaboration Registry of Controlled Trials. We searched between January 1, 2003 and January 31, 2017, building upon the most recent full search for this topic.85

We worked with a research librarian to develop our search strategy, which was peer-reviewed by a second research librarian (Appendix A). All searches were limited to articles published in English. We managed literature search results using version X7 of Endnote® (Thomson Reuters, New York, NY), a bibliographic management software database.

To ensure comprehensiveness, we reviewed the reference lists of included studies and relevant systematic reviews and meta-analyses to identify relevant articles that were published before our search dates or were not identified in our literature searches.86, 87

Study Selection

Two reviewers independently reviewed the title and abstracts of all identified articles using Abstrackr88 to determine if the study met our a priori inclusion and exclusion criteria for design, population, intervention, and outcomes (Appendix A). Two reviewers then independently evaluated the full-text article(s) of all potentially relevant studies against the complete inclusion criteria.
and exclusion criteria. Disagreements in the abstract and/or full-text review were resolved by discussion.

For both KQs we considered randomized controlled trials in asymptomatic average risk women, including women of unknown risk, aged 45 and older. Studies focused on screening explicitly among high-risk populations (e.g., BRCA mutation carriers, individuals with first-degree relatives with ovarian cancer) were excluded from this review. We included any screening intervention conducted in primary care settings (including obstetrics and gynecology practices) compared with usual care or no screening. When available, as is the case for ovarian cancer screening, RCT evidence was prioritized, particularly when adequately powered to evaluate screening for a rare cancer. Trials that addressed only the accuracy of screening and cancer detection rates without reporting morbidity, mortality, or quality of life data were not included in this review.

For KQ 1 we considered a range of outcomes including: ovarian cancer-specific mortality (including primary peritoneal and fallopian tube cancer), cancer-related morbidity, and quality of life. KQ 2 included the following harms outcomes: mortality from other causes, rates of surgery, rates of false positive screening results, complications of diagnostic surgical procedures, and health and psychological effects of screening tests. Evidence of potential harms associated with screening was limited to trials reporting health outcomes. Effects of screening on quality of life are evaluated as a potential benefit or harm, but this outcome was reported only for harms. Importantly, direct trial evidence on the health outcomes associated with different screening programs (including not screening) offers the summary or net effect of harms and benefits of screening. Disease specific mortality is the key outcome for ovarian cancer rather than all-cause mortality because of the rarity of ovarian cancer, and because the mortality rates from ovarian cancer are quite high. Due to its rarity, ovarian cancer screening intervention effect sizes would not be expected to be large enough to have discernable effects on all-cause mortality. Nevertheless, it is important to compare mortality across study arms to determine whether there are any unexpected increases in other causes of mortality from the screening program, to assess for problems with coding or discernment of cause of death in the trial procedures, and to evaluate whether differential study continuation resulted in unrecognized study arm imbalances in participant risk of death.

Quality Assessment and Data Abstraction

Two reviewers independently assessed the methodological quality of each study using predefined criteria developed by the USPSTF. Disagreements in quality were resolved by discussion. Each study was given a final quality rating of good, fair, or poor.

Good-quality RCTs had adequate randomization procedures and allocation concealment, blinded outcome assessment, reliable outcome measures, similar baseline characteristics between groups, low attrition (≥90% of participants had followup data with <10 percentage point difference in loss to followup between groups), used intention-to-screen analysis, and reported diagnostic criteria for outcome ascertainment. We rated trials as fair quality if they were unable to meet the majority of the good-quality criteria. We rated trials as poor quality if attrition was greater than
40 percent or differed between groups by 20 percentage points, or if there were any other “fatal” flaws that seriously affected internal validity, as agreed upon by two independent investigators. Poor quality studies were excluded from this review.

We abstracted data from all included studies into standard evidence tables using Microsoft Word® (Microsoft Corporation, Redmond, WA). A second reviewer checked the data for accuracy. We abstracted information on study design, baseline data, intervention details, health outcomes, and adverse events.

**Data Synthesis and Analysis**

Given the small number of studies and heterogeneity of screening interventions, meta-analyses were not conducted for any of the KQs in this report. We instead conducted a narrative synthesis of the results by screening strategy. We generated summary tables and descriptive text detailing the populations and protocols of included studies, and the interventions and followup procedures of the included trials. The pre-specified outcomes sought for this review were abstracted from each study, by key question, and results were presented in groups defined by the screening strategy. The selected pre-specified outcomes were prioritized for abstraction even when they were not the primary outcome of the included studies. Some outcomes were calculated from raw data reported in study publications where needed to adhere to Task Force priorities or to facilitate comparability across trials. As a result, some results reported in this review may differ from the findings highlighted in the main results of the original publications.

In 2014, the World Health Organization revised the classification of tumors of the female reproductive system. Following these revisions, most cancers historically classified as peritoneal cancer would be reclassified as ovarian and tubal cancers. Analyses with both ovarian and peritoneal cancer were prioritized, since their presentation and treatment are not distinct in clinical practice and because they are often difficult to distinguish pathologically. Ovarian cancer was defined as primary ovarian, peritoneal, and fallopian tube cancers in the Prostate, Lung, Colorectal and Ovarian (PLCO). The U.K. Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) also reported results for this outcome, but excluded primary peritoneal cancer from primary analyses. We have noted where we have calculated results that included peritoneal cancer using raw numbers reported in the UKCTOCS. Data from the U.K. Pilot trial did not capture information related to peritoneal cancer; therefore, these results are limited to primary cancer of the ovary and fallopian tubes.

The number of participants stated throughout the report refers to the number analyzed, unless otherwise stated. The tables report both the n randomized and the n analyzed.

Data on false positive rates were calculated as the percent of those without ovarian cancer who received a positive test result. The PLCO defined a positive screen based on the results of the initial screening tests (i.e., positive on TVU or CA-125 screening test). In the UKCTOCS a positive screen was defined as a positive initial screen that eventually led to surgery. The definition used in the PLCO, where the screening test result defines test positives, is consistent with the USPSTF interest in evaluating how the initial test sets in motion a cascade of effects.
Each definition, however, provides slightly different insights, so we calculated and report false positive rates based on both definitions.

In accordance with the CONSORT scientific standards for the design and analysis of randomized controlled trials, when multiple statistical tests are reported to test comparisons between trial arms, we prioritize statistical tests defined a priori in the trial protocol.94

When reporting study results from the two large trials included in this review, we aimed to provide comparable statistics, and have noted where this required additional calculations.

**Grading the Strength of the Body of Evidence**

We graded the strength of the overall body of evidence for each KQ. We adapted the Evidence-based Practice Center approach95 which is based on a system developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group.96 Our method explicitly addresses four of the five Evidence-based Practice Center-required domains: consistency (similarity of effect direction and size), precision (degree of certainty around an estimate), reporting bias (potential for bias related to publication, selective outcome reporting, or selective analysis reporting), and study quality (i.e., study limitations). We did not address the fifth required domain—directness—as it is implied in the structure of the KQs (i.e., pertains to whether the evidence links the interventions directly to a health outcome).

Consistency was rated as reasonably consistent, inconsistent, or not applicable (e.g., single study). Precision was rated as reasonably precise, imprecise, or not applicable (e.g., no evidence). Reporting bias was rated as suspected, undetected, or not applicable (e.g., when there is insufficient evidence for a particular outcome). Study quality reflects the quality ratings of the individual trials and indicates the degree to which the included studies for a given outcome have a high likelihood of adequate protection against bias. The body of evidence limitations field highlights important restrictions in answering the overall KQ (e.g., lack of replication of interventions, nonreporting of outcomes important to patients).

We graded the overall strength of evidence as high, moderate, or low. “High” indicates high confidence that the evidence reflects the true effect and that further research is very unlikely to change our confidence in the estimate of effects. “Moderate” suggests moderate confidence that the evidence reflects the true effect and that further research may change our confidence in the estimate of effect and may change the estimate. “Low” indicates low confidence that the evidence reflects the true effect and that further research is likely to change our confidence in the estimate of effect and is likely to change the estimate. A grade of “insufficient” indicates that evidence is either unavailable or does not permit estimate of an effect. Two independent reviewers rated each KQ according to consistency, precision, reporting bias, and overall strength of evidence grade. We resolved discrepancies through consensus discussion involving more reviewers.
A draft Research Plan was posted for public comment on the USPSTF Web site from March 26, 2015, to April 22, 2015. The USPSTF received several comments about the population under consideration, particularly an interest in including women with specific genetic risk factors. Considerations for these populations are generally outside the scope of a review of a broad screening program among asymptomatic women at average risk for ovarian cancer; therefore, the included population for this topic was not changed. Several comments sought clarification on the analytic framework. In response, the USPSTF simplified the analytic framework to focus on direct evidence from screening trials. A few comments requested that specific screening interventions be evaluated. The USPSTF clarified the inclusion criteria to indicate that all screening tests and approaches evaluated in clinical trials will be included, and added a contextual question to consider whether there are promising screening approaches not yet evaluated by clinical trials. A final research plan was posted on the USPSTF’s Web site on March 3, 2015.

A draft report was shared with invited expert reviewers and federal partners (see Acknowledgements) and their comments were compiled and addressed (where appropriate) in our report posted for public comment. The preliminary draft of the full report was posted for public comment on the USPSTF Web site from July 18, 2017 through August 14, 2017. A few commenters, raised concerns about the manner in which the review characterized the CA-125 ROCA screening algorithm. Following usual USPSTF procedures for evaluating screening programs, the screening test women would experience each year in the screening program was considered the screening test, and any followup tests that occur on the basis of that result are considered followup on a positive result. Some commenters preferred evaluating the combined effects of screening and followup when considering the potential harms of false positive results. Revisions to the Evidence Synthesis include clearer explanation of the approach, and more detail on the differences in the extent to which trials followed strict followup protocols (as in UKCTOCS) versus those that referred women to community practices where the followup was inevitably more variable (as in PLCO). The initial report, and the final version included the false positive results for both the annual screening test (screening false positive rate) and the combined effect of screening and followup on rates of referral to diagnostic surgery (surgery false positive rate). Providing both of these outcomes, whether or not they were reported as key outcomes in the trials, allows comparability between trials and additional information for discussion.

Commentators also requested further explication of the choice of study outcomes (ovarian cancer as defined by international criteria including ovarian, fallopian tube, and peritoneal cases) and analyses, particularly from the UKCTOCS trial. Several exploratory analyses are discussed in our Final report, along with further justification for a focus on complete intention-to-treat statistical analysis, preferably with study arms comparisons with prespecified statistical tests. Nevertheless, the report highlights the importance of continuing to evaluate the results of trials as new information comes to light that would either strengthen or dampen potential effects observed in exploratory analyses.

Overall, in response to valuable input from public comment and federal partners, the report was revised for clarity, to further explain our methods and decisions regarding the scope and focus of
our synthesis.

USPSTF Involvement

We worked with USPSTF members at key points throughout this review, particularly when determining the scope and methods for this review and developing the Analytic Framework and KQs. After revisions reflecting the public comment period, the USPSTF members approved the final analytic framework, KQs, and inclusion and exclusion criteria. AHRQ funded this review under a contract to support the work of the USPSTF. An AHRQ Medical Officer provided project oversight, reviewed the draft report, and assisted in the external review of the report.
Chapter 3. Results

Description of Included Studies

Our literature search yielded 1,381 unique citations. From these, we reviewed the full text of 74 articles. After full text review and critical appraisal, we included 4 trials (published in 17 articles) (Appendix A Figure 1). Appendix B contains a full list of included studies.

Of the 74 articles that were reviewed, the most common reason for exclusion were not having an included study design (i.e., not RCT) or not reporting a relevant outcome. No trials were rated as poor quality. Appendix C contains a list of all excluded trials and their reasons for exclusion.

Four RCTs (n = 293,587) were included; three reported health outcomes (KQ1), and all reported potential harms of screening (KQ2) (Table 1). Two were conducted in the United States91, 97 and two in the United Kingdom.92, 98 Analyzed sample sizes ranged from 549 to 202,546 participants. The three trials91, 92, 98 reporting health outcomes and potential harms were rated good-quality and the small trial97 reporting only on psychological harms of screening was rated fair-quality.

Included Populations

The two U.K. trials (U.K. Pilot and UKCTOCS) were limited to postmenopausal women aged ≥45 and 50 to 74,92, 98 the PLCO to women aged 55 to 74,91 and the Quality of life, Education, and Screening Trial (QUEST) to women aged 30 or older97 (Table 1). Women with a history of bilateral oophorectomy or ovarian malignancy, at increased familial risk of cancer, or with an active malignancy where excluded from the largest U.K. trial.92 Exclusions were similar for the other trials included for KQ1,91, 98 but the PLCO additionally excluded women with a history of colorectal or lung cancer, or who had undergone an investigation or treatment for these cancers.91 The QUEST study, included only for KQ2, also excluded women with plans to become pregnant.97

Results of Included Studies

KQ 1. Does Screening for Ovarian Cancer in Asymptomatic Women Using Single Tests or Combined Algorithms Reduce All-Cause or Disease-Specific Morbidity and Mortality?

Summary

Three trials evaluating four screening strategies showed no effect of screening on ovarian cancer mortality on the basis of a priori per protocol statistical analyses testing. The screening tests evaluated were annual transvaginal ultrasound, annual transvaginal ultrasound and CA-125 serum testing, annual CA-125 testing, and annual CA-125 serum testing interpreted with an
algorithm (ROCA) that incorporates changes over time to inform triage and rescreening intervals.

Description of Included Studies

Three large, good-quality trials reported ovarian cancer incidence and mortality for women randomized to ovarian cancer screening versus no screening or usual care. The largest (n=202,546) and most recent trial is the UKCTOCS,92 which began enrolling trial participants in 2001 through 13 National Health Service centers in England, Wales, and Northern Ireland. Identification of women was centralized using primary care registers, and eligible women received personalized invitations. A smaller included trial (n=21,935) was conducted in the United Kingdom by the same research group in preparation for UKCTOCS (U.K. Pilot).98 This trial recruited women who had participated in a previous ovarian cancer screening study.99 The third included trial is the PLCO91 cancer screening randomized trial, begun in 1993 at 10 clinical screening centers in the United States (n = 68,557). The 10 screening centers developed individual recruitment methods. Women were primarily recruited via direct mail; however, other methods included community outreach and mass media.100

The demographic characteristics of randomized trial participants are described in Table 2. In the UKCTOCS and U.K. Pilot trials, the race/ethnic composition of the study population was over 95 percent white.92, 98 Eighty-eight percent of women in the PLCO91 were white, non-Hispanic (6% were black, non-Hispanic, 4% were Asian or Pacific Islander, and very few were Hispanic or Native American). The PLCO91 and UKCTOCS92 had similar participant age distributions, with over half of UKCTOCS participants between the ages of 56 and 66 at baseline and 65% in the PLCO between the ages of 55 and 64. The U.K. trials92, 98 included post-menopausal women as young as 45 or 50, whereas the PLCO91 did not include any women younger than age 55. Nineteen percent of women in UKCTOCS92 and 27 percent of PLCO91 participants had had a prior hysterectomy at baseline (without bilateral oophorectomy). Although women with a known elevated familial ovarian cancer risk were excluded, 1.6 percent of women in UKCTOCS92 reported maternal history of ovarian cancer, and 6.4 percent a maternal history of breast cancer. 17 percent of women in the PLCO 91 reported any family history of breast or ovarian cancer.

All three trials tested annual screening for ovarian cancer with CA-125 and/or transvaginal ultrasound (Table 3). The UKCTOCS92 had two intervention arms and a no-screening control arm (randomized 1:1:2, respectively). Women assigned to CA-125 risk algorithm screening received CA-125 serum testing, with triage and followup determined by application of the Risk of Ovarian Cancer Algorithm (ROCA). Three levels of followup, depending on the ROCA estimate were assigned: continue annual screening (low risk), repeat CA-125 in 12 weeks (intermediate risk), or repeat CA-125 testing and transvaginal ultrasound in 6 weeks (elevated risk). Women with persistent abnormalities were referred for surgical investigation at tertiary care health centers in the U.K. National Health Service system. Women in the ultrasound intervention arm had yearly transvaginal ultrasound testing, with repeat ultrasounds scheduled for unsatisfactory or abnormal results and referral to surgical diagnostic testing and recommended treatment through usual care in the U.K., as in the CA-125 screening arm. Women in the intervention arms of the UKCTOCS underwent a maximum of 11 screening rounds with a median of 11.1 years of follow up.92 The U.K. Pilot98 trial compared annual CA-125 screening
tests to no screening. Women with CA-125 serum levels of 30 U/mL or greater were considered screening test positive, and further evaluated using ultrasonography. Women in the intervention arm of the U.K. Pilot study underwent a maximum of 3 rounds of screening with up to 8 years of follow up. Women in the screening intervention arm of the PLCO received both CA-125 testing and ultrasonography. The cutoff designating an abnormal result on the CA-125 test was serum level of 35 U/mL or greater. Women with abnormal results on either the CA-125 test or ultrasound were referred for additional evaluation by notifying the woman and her usual community physician. The original trial design for PLCO called for screening for only four annual ovarian cancer screening rounds. A modification of the protocol in 1998 extended screening to 6 annual rounds, with the purpose of increasing the power of the trial. During this 2-round extension, only CA-125 was used for screening. Therefore, women in the intervention arm of PLCO underwent a maximum of 6 rounds of screening (4 with CA-125 and ultrasound, 2 with CA-125 alone) with a median of 12.4 years of follow up. Additionally, in the first 5 years of the PLCO, palpation of the ovaries with a bimanual pelvic examination was also included in the screening protocol. This component of the intervention was dropped, however, as no cancers were identified solely on the basis of this screening component.

Screening Adherence and Contamination

In the UKCTOCS, 95 percent of women in the ultrasound arm completed at least one screen, and 99 percent of women in the CA-125 ROCA intervention arm completed at least one screen (Table 4). Overall 81 percent of screens were attended in the CA-125 ROCA intervention arm and 78 percent were attended in the ultrasound intervention arm. Adherence varied by round, ranging from 47 to 98 percent in the CA-125 ROCA group and 36 to 95 percent in the ultrasound group. In the U.K. Pilot trial, 86 percent of randomized participants completed at least one screen and 71 percent completed all three screening rounds. Adherence in the PLCO was similar, with 78 to 84 percent of participants attending ultrasound screening and 73 to 85 percent of women attending CA-125 screening depending on the screening round. Rates of screening contamination in the control group were not reported in the U.K. Pilot trial, but were below 5 percent in the UKCTOCS and PLCO.

Definition of Outcomes in the Trials

The following results from the PLCO and UKCTOCS include cases of ovarian, fallopian, and primary peritoneal cancer cases, the prioritized outcome for this review. Data from the U.K. Pilot trial did not capture information related to peritoneal cancer; therefore, these results are limited to primary cancer of the ovary and fallopian tubes.

Ovarian Cancer Incidence

The incidence of ovarian cancer did not differ significantly between study arms in any of the included trials. In the UKCTOCS ovarian cancer was diagnosed 0.7 percent of those in the CA-125 ROCA arm (354 cases), and 0.6 percent of those in the ultrasound arm (324 cases) and control arm (645 cases). The incidence rate of ovarian cancer was 6.4 per 10,000 person-years in the CA-125 ROCA screening arm, and 5.9 per 10,000 person-years in both the ultrasound and no screening control arms. In the PLCO, ovarian cancer was diagnosed in 0.6 percent (212 cases)
of those in the intervention arm and 0.5 percent (176 cases) of those in the usual care arm (RR 1.21 95% CI, 0.99 to 1.48). In the U.K. Pilot trial, ovarian cancer was diagnosed in 0.1 percent (16 cases) of women in intervention arm and 0.2 percent (20 cases) in the no-screening arm. The low incidence may be due to smaller sample sizes, or the fact that women underwent a prevalence screen 10 years prior to the study. Incidence rates were not reported for the U.K. Pilot trial.98

**Ovarian Cancer Mortality**

*CA-125 Screening*

The UKCTOCS92 and U.K. Pilot98 trial included intervention arms that compared CA-125 screening to no screening *(Table 5)*. Although both of the trials used an initial CA-125 screening test to determine followup, the U.K. Pilot study used a single cutpoint level to refer participants to further evaluation (CA-125 ≥ 30 u/ml), and the UKCTOCS used the ROCA algorithm103 to triage intermediate and high CA-125 ROCA results for repeat screening and ultrasound. The ROCA CA-125 screening arm was described as multimodal screening in the UKCTOCS trial publications, and included a standard protocol for all followup screening tests. The ROCA algorithm is more complex than single cutpoint CA-125 testing approaches because it incorporates changes in CA-125 changes over time for individual women. In the UKCTOCS, ovarian cancer mortality (calculated to include peritoneal cancers) in the intervention arm and control arm was similar (IG: 0.32% versus CG: 0.35%) and in survival analysis was 2.9 ovarian cancer deaths per 10,000 person-years in the intervention group and 3.3 ovarian cancer deaths per 10,000 person-years in the control group. This difference was not statistically significant based on a Cox proportional hazards model (HR = 0.89 [95% CI, 0.74 to 1.08]).92

The smaller U.K. Pilot98 trial (n = 21,935) was designed to assess feasibility and performance of screening and was not powered to test mortality differences. There were 9 ovarian cancer deaths in the intervention group (0.08%) and 18 in the no-screening comparison group (0.16%); the difference was not statistically significant (RR 0.50 [95% CI, 0.22 to 1.11]). A statistically significant difference in survival between women with index cancers in the IG and the CG was observed when computed from the date of randomization (IG median 72.9 months, CG median 41.8 months; p=.01). This finding was based on a small number of events, and only 6 of 16 index cancers identified in the intervention arm were screen detected. Survival in the control group noted by the study authors as being unusually poor.

*Ultrasound Screening*

The ultrasound screening intervention arm in the UKCTOCS92 did not reduce ovarian cancer mortality compared with no screening *(Table 5)*. Ovarian cancer mortality in the intervention arm and control arm was similar (IG 0.32% versus CG 0.35%) and in survival analysis was 3.0 per 10,000 person-years in the intervention group and 3.3 per 10,000 person-years in the comparison group (HR 0.91[95% CI, 0.76 to 1.09]).92
Combined CA-125 and Ultrasound Screening

The incidence of ovarian cancer mortality in the PLCO\textsuperscript{91} was 3.1 per 10,000 person-years in the intervention arm and 2.6 per 10,000 person-years in the usual care comparison arm. There were 118 deaths in the intervention group (0.34\%) and 100 deaths in the control group (0.29\%), a not statistically significant difference (RR 1.18 [95\% CI, 0.82 to 1.71]). Survival with ovarian cancer did not differ significantly between study arms when estimated from the date of diagnosis (p=0.18) and from the date of randomization (p=0.67), which better accounts for possible lead-time bias (Table 5).\textsuperscript{91}

Quality of Life

Two studies addressed changes in quality of life associated with ovarian cancer screening,\textsuperscript{97, 104} both with regard to potential anxiety associated with screening. These results are therefore reported for Key Question 2 as potential harms.

Ovarian Cancer Morbidity

There were no differences in treatment approach by study arm in the PLCO; 81 percent in the intervention group received surgery plus systemic therapy, compared with 80 percent in the usual care group.\textsuperscript{91}

KQ 2. What Are the Harms of Screening for Ovarian Cancer, Including Harms of the Screening Test and of Diagnostic Evaluation?

Summary

Four trials reported on the harms of ovarian cancer screening. False positive rates and surgical harms were highest for screening programs using transvaginal ultrasound as an initial test. Major surgical complications as estimated in the two largest trials occurred in women with investigations from screening that did not lead to a cancer diagnosis, ranging from 3 to 15 percent of surgeries. The screening tests themselves resulted in minor complications, at rates widely ranging based on study specific definitions from 0.86 to 58.3 per 10,000 screens/women for CA-125 test blood draws (e.g., fainting, bruising) and from 1.86 to 3.3 per 10,000 screens/women for ultrasound testing (e.g., pain, discomfort, infection, bruising).

Description of Included Studies

Evidence on false positive rates and surgical harms of screening were included from the three trials\textsuperscript{91, 92, 98} included for KQ1. A substudy of the UKCTOCS\textsuperscript{104} (n=23,374) and an additional fair-quality trial (QUEST)\textsuperscript{97} (n=549) aimed at evaluating the effects of ovarian cancer screening on quality of life (QoL) and psychological outcomes are described below.
False Positive Rates and Complications

**CA-125 Screening**

False positive rates calculated for screening with CA-125 across all rounds of the U.K. Pilot trial were 4.2 percent. In the UKCTOCS the false positive rate in the prevalence round of screening was 9.0 percent and cumulatively across all subsequent incidence rounds of the UKCTOCS the false positive rate rose to 44.2 percent with a total of 44.3 percent of women screened receiving a positive test result at least once (Table 6). Screening test complications were minor for CA-125 screening, including bruising, fainting, and pain. These complications were reported to occur for 0.86 per 10,000 screens in UKCTOCS and 58.3 per 10,000 women in the PLCO.

False positive surgeries, defined as surgery following a positive screen among women found to have normal or benign adnexa, occurred in nearly 1 percent of those in the CA-125 ROCA arm of the UKCTOCS (n = 488) and 0.2 percent of those screened in the U.K. Pilot (n=23). In the UKCTOCS, one or more major complication occurred in 3.1 percent of false positive surgeries in the CA-125 ROCA arm and included: infection, injury to hollow viscus, anesthetic complications, and cardiovascular and pulmonary events. No surgical complications were reported among the women undergoing surgical investigations in the U.K. Pilot trial.

**Ultrasound Screening**

False positive rates calculated based on the results of the initial screening using ultrasound occurred in 11.9 percent in the prevalence screening round of the UKCTOCS (Table 6). The false positive rate from subsequent rounds was not reported for this arm. Complications from transvaginal ultrasound screening in the UKCTOCS included pain, cystitis or other infection, discomfort, bruising, and fainting, as well as others (unspecified), among 1.86 per 10,000 screens. In the PLCO complications from transvaginal ultrasound occurred in 3.3 per 10,000 women.

False positive surgeries (benign findings on a screen positive investigation) occurred in 3.2 percent of women in the ultrasound arm of the UKCTOCS. Major complications occurred in 3.5 percent of these surgeries and included: injury to hollow viscus, hemorrhage, anesthetic complication/myocardial infarction, hernia, deep vein thrombosis/pulmonary embolism, wound breakdown, bowel obstruction, wound/supravaginal hematoma, infection, and pain with readmission or further operation.

**Combined CA-125 and Ultrasound Screening**

The combined CA-125 and ultrasound intervention used for the PLCO had a 9.6 percent cumulative false positive rate across all rounds of screening (Table 6). Complications from each screening component are discussed above.

False positive surgeries (benign findings on a screen positive investigation) occurred in 3.2 percent of women undergoing combination screening in the PLCO. Major complications occurred in 15.1 percent of surgeries, with benign findings in the PLCO including: infection,
direct surgical harms, cardiovascular/pulmonary events, and other adverse events that were not specified. 91

Deaths From Other Causes

The UKCTOCS provided data on causes of death other than ovarian cancer, and the PLCO provided data on causes of death other than ovarian, colorectal, and lung cancers. There were no statistically significant differences in causes of death in screening arms compared with control arms.

Psychological Harms of Screening

A study of the psychological morbidity associated with ovarian cancer screening was undertaken within the UKCTOCS (n=23,374).104 A random sample of women was drawn at baseline from each trial arm (n=1,339) and a survey was administered annually for 7 years to obtain variables for evaluating psychological effects of screening. Similar data were also collected for women in the screening arms who were at any time recalled for followup testing (CA-125 ROCA: n = 12,357; ultrasound: n = 9,678). For the event group, upon abnormal results and recall, questionnaires were administered, and thereafter on an annual basis.104 The main measures were the Spielberger State/Trait Anxiety Inventory (STAI)105 and the General Health Questionnaire 12 (GHQ-12)106 for evaluating psychological morbidity. A small but statistically significant difference in the education levels of women between the ROCA screening and ultrasound group of the random samples was observed, and in the event sample more women reported hormone replacement therapy use than in the random sample (21% versus 14%, p<.001).104 In light of these unexpected imbalances, the study authors recommended cautious interpretation of findings. In adjusted analyses with linear and logistic regression, no statistically significant differences in mean STAI or the risk of psychological morbidity (GHQ-12 ≥ 4) were observed between the control and intervention arms in the random sample. In the analysis of women with recall screening events, there was a statistically significant increased risk of psychological morbidity among women recalled for higher-level screening (adjusted OR 1.28 [95% CI, 1.18 to 1.39]).104

The QUEST97 trial analyzed 549 average-risk women age 30 years or older in the United States to examine the effect of ovarian cancer screening on cancer worry and QoL. Women were randomized to ovarian cancer screening, risk counseling, or a screening/risk counseling combination compared to a usual care-only protocol consisting of annual pelvic examination and routine education by a woman’s primary care physician. We report only on results from the screening and usual care control arms of the trial (n = 442). Ovarian cancer screening consisted of alternating CA-125 measurement and TVU every 6 months for a maximum of 4 screening rounds. Overall, women had a high level of education, with 99 percent completing high school and 95 percent attending college. The QUEST97 study found no statistically significant differences for QoL measured with the SF-36, distress measured with the Impact of Events Scale, or cancer worry measured with a modified Lerman cancer worry scale. The only significant effect observed in participants was a higher level of cancer worry after 2 years among those who had experienced any abnormal test results.97
Chapter 4. Discussion

This review considered direct trial evidence of the health benefits and harms of ovarian cancer screening. The known availability of evidence from large trials and previous recommendations against ovarian cancer screening based on trial evidence motivated this focus on the overarching effects of screening on women’s health outcomes. The rarity of ovarian cancer necessitated a focus on mortality from ovarian cancer rather than on overall mortality from all-causes because the effects of screening on overall mortality would be minor given that ovarian cancers represent a very small proportion of deaths overall. Because ovarian cancer is rare, large trials are necessary to evaluate effects of screening on ovarian cancer morbidity and mortality in average-risk women.

Summary of Evidence

Since the previous review of this topic, mortality results from the large, well-designed UKCTOCS were published. Thus, there were three trials reporting mortality outcomes after ovarian cancer screening. The PLCO and UKCTOCS were designed with statistical power to detect a 30 to 35 percent difference in mortality from this relatively rare but often fatal cancer, and had null findings in primary analyses. The U.K. Pilot trial reported mortality outcomes, but was designed to examine the feasibility of screening and was underpowered to detect a mortality difference. The small QUEST trial evaluated quality of life and psychological effects of ovarian cancer screening. Given evidence that there is no mortality benefit from ovarian cancer screening, the harms associated with screening merit focused consideration. For women in the ovarian cancer screening programs evaluated, positive tests and followup can lead to surgery and surgical complications in women without disease.

Further Discussion of UKCTOCS

Despite a primary null finding from the UKCTOCS on ovarian cancer mortality, the investigators included statistical analyses suggestive of a possible long-term benefit of the CA-125 ROCA screening intervention on ovarian cancer mortality (excluding peritoneal cancers), based on their observation that Kaplan-Meier cumulative mortality curves appear to diverge approximately 10 years after randomization. Consistent with best practices, the trialists were transparent in their reporting of protocol-specified versus post hoc exploratory analyses. There are several reasons our review did not focus on these secondary analyses. First, we prioritized analyses with both ovarian and peritoneal cancer included, since their presentation and treatment is not distinct in clinical practice and because they are often difficult to distinguish pathologically. More of the cancers identified in the CA-125 ROCA screening arm of UKCTOCS were coded as peritoneal cancers than in the ultrasound and no screening arms. It is not surprising that excluding these cases that have very high mortality increases the difference in mortality between the CA-125 ROCA and control arm, Nevertheless, excluding peritoneal cancer cases did not alter the null findings of the pre-specified statistical analyses. Second, we focused on statistical tests that were specified a priori through publication of a
A statistically significant benefit for the CA-125 ROCA intervention was also detected when peritoneal cancer cases were excluded and a weighted Log Rank test was used, which assigns greater weight in the survival analysis to later years of the trial (HR 0.78 [95% CI, 0.62 to 0.97]). This report focuses on the prespecified statistical analysis. Beyond the 10 year followup range of the trial, far fewer of the study participants have accrued followup data, and statistical analyses weighting the years far beyond the end of the screening period may overemphasize the observed later divergence of the ROCA arm. Nevertheless, publication of UKCTOCS trial data incorporating additional years of completed followup will be important contributions, as will causal theories to explain delayed screening effects, if observed. A subgroup analysis of the data that obtained statistically significant findings endeavored to remove cases that were prevalent at the outset of the trial, using imputation, modeling and stored CA-125 data when available. Our review focused on intention-to-treat analysis of all participants, since these findings are most robust and applicable to the implementation of a screening program, and its cumulative effects.

Third, the divergence of the trial arms later in the study period are more difficult to attribute to the original randomized condition and screening, as the longer a study continues, the more opportunities there are for measured and unmeasured differences in the study arms to accrue. As noted, there were also substantially fewer women at risk included in the analyses beyond 10 years, because women recruited into the study later have not yet accrued followup for inclusion in the analysis of the longer time horizon. Thus, data from the later years of the trial (>10 years) are based on incomplete data and should be cautiously interpreted.

Differential reasons for censoring could lead to some divergence in the ovarian cancer mortality curves as followup times lengthen. In this study, there were no differences across arms in participant followup (censoring) or other causes of death, but there may have been differences between arms and changes over time in the proportion of participants in the trial with two ovaries intact. Those with both ovaries rather than just one, by definition, have higher ovarian cancer risk. In other words, the usual-care screening arm may have had a net surplus of ovaries at risk, despite a similar proportion of women at risk. We calculated the overall proportion of women having an oophorectomy during the trial, based on surgical investigation of screening results (true and false positives) or other indications based on numbers reported for the UKCTOCS. Approximately 4.4 percent of women in the ultrasound arm, 2.3 percent in the CA-125 ROCA arm, and 1.4 percent in the no-screening arm had oophorectomies (commonly bilateral, with or without salpingectomy). Others have suggested that the potential prophylactic effect of ovary and fallopian tube removal might influence the UKCTOCS results, especially in the long term. To date, there is no overall difference in the incidence of ovarian cancer by arm, suggesting that a prophylactic effect is not present, but as more years of followup data are available for more of the enrolled participants, additional analysis of the cumulative cancer incidence rate by study arm can be undertaken. The UKCTOCS team have received grant funding to continue follow up through 2018.

Further Discussion of PLCO

The PLCO was the only U.S.-based study for directly assessing potential net benefits or harms of
screening. The trial protocol consisted of annual transvaginal ultrasound for 4 years and annual CA-125 testing for 6 years. Annual bimanual ovarian palpation by trained examiners was also included during the first 4 years of the trial. Depending on when women entered the study, they received 0 to 4 physical palpation examinations as part of the screening program, consequently 20,872 women in the screening arm received at least one ovarian palpation examination (61%). This screening modality was dropped from the trial, however, because no cancer cases were identified solely on the basis of palpation (i.e., all cancer cases were also positive on CA-125 and/or TVU), and test sensitivity (defined as cancer diagnosed within 1 year of screening positive with a palpable adnexal mass) was very low (5.1%).

Overall, in the PLCO, the ovarian cancer mortality rate was greater in the intervention arm compared with usual care over 13-years of followup, although the difference was not statistically significant. A recently published analysis added up to 6 additional years of post-trial followup mortality data (mean 2.3 years) and did not find evidence of a late-emerging benefit of screening. The post-trial data were obtained through a different, centralized system rather than through the trial screening centers, and upon completion of the trial centralized followup was refused by 16 percent of usual care women and 12 percent of intervention arm participants. Consequently, total followup time was shorter for refusers than for those willing to participate in ongoing surveillance. Overall, however, followup times did not differ across arms with mean followup of approximately 15 years in both groups. The rate ratio moved toward null (RR 1.06 [95% CI, 0.87 to 1.30]) from 1.18 during the trial period. There was also no difference in ovarian cancer specific survival by arm in the trial or its extended followup (p=0.16).

An additional supplemental analysis of PLCO data aimed to determine whether use of the ROCA algorithm on CA-125 measurements collected for the PLCO would have had better performance for identifying ovarian cancer cases compared with the PLCO screening protocol. The analysis employed a best-case scenario assumption biased toward finding a ROCA test benefit over the PLCO protocol. Namely, all cancers in the trial that would have had a positive ROCA screening test occurring earlier than a positive screening test with the PLCO protocol were assumed to have avoided mortality from ovarian cancer. Modeling with PLCO data suggested that application of the ROCA algorithm used in the UKCTOCS would have led to earlier diagnosis of cancer in 32 percent of cases that were detected using CA125 screening using a single cutoff. This analysis obtained a mortality relative risk of 0.90 (95% CI, 0.69 to 1.17), and the authors concluded that even under the most lenient assumptions, the ROCA algorithm would not necessarily have resulted in a beneficial trial finding.

Another recent secondary analysis of PLCO participants with a family history of breast or ovarian cancer found a non-significant trend towards diagnosis of stage I or II cancers in the screened arm compared with the usual care arm (29% versus 17%; p=0.085) and improved survival in these patients with ovarian cancer detected by screening compared to usual care; however, this apparent improvement in survival did not result in improved ovarian cancer mortality.

The control condition was described as “usual care” in the U.S. PLCO and as “no screening” in the U.K. trials. If usual care included any practices that might affect ovarian cancer detection and treatment, this could potentially reduce differences between the study arms in the PLCO.
Surveys have found that a majority of U.S. primary care and reproductive specialty clinicians conduct bimanual pelvic examinations as part of their routine gynecological care for women, believing it to be an effective way to screen for ovarian cancer.69,115 During the early years of the PLCO, the screening protocol included ovarian palpation with bimanual pelvic examinations, but this element was dropped from the intervention protocol 5 years into the trial. Estimates of the test performance of the bimanual examination derived from the PLCO and systematic reviews of the effectiveness of routine pelvic examinations116 have found limited evidence on its effectiveness, and have brought to light its poor accuracy. Thus, it is unlikely that ovarian palpation in the screening or control arm of the PLCO would influence the results observed. Essentially, women in both arms of the PLCO likely received pelvic examinations with ovarian palpation over the course of the study, and the CA-125 and TVU screening intervention can be validly compared against usual care without routine CA-125 and TVU screening. Overall rates of TVU and CA-125 testing contamination in the control arm were similar for the PLCO and U.K.-based studies, however, suggesting that differences in the control condition did not contribute to widely divergent practices in the control arms.

**Implications of Stage Shift and Treatment Findings in the Absence of a Mortality Benefit**

As a contextual question in this review, we examined included studies for evidence of a cancer stage or type shift. We focused the evaluation of stage shift on comparisons in the trial arms between women diagnosed with localized disease (stage I) and those with regional or distant disease (stages II-IV). These comparisons are relevant because of the higher survival rates associated with disease diagnosed at the localized stage and clinical differences in treatment strategies between stage I and higher-stage ovarian cancer (i.e., need for adjuvant radiation therapy).1 Patients with stage II ovarian cancer have high recurrence rates. As a result, in 2009 the Gynecologic Oncology Group has recommended that stage II patients be included in trials that assess treatments for advanced-stage disease.1 Tests for differences in proportions were conducted in Stata.

Detection of a higher proportion of localized cancers in the screening arms compared with control arms was reported in the UKCTOCS.92 A statistically significant (p<0.005) greater proportion of cases was identified at the localized stage (stage I) in the CA-125 ROCA (36%) and ultrasound (31%) arms than in the control arm (23%) (calculated from reported data on cancer stage). The overall differences by arm and stage were also statistically significant when comparing localized and regional cancers (stages I and II) to more advanced stages (stages III and IV). Differences between the CA-125 ROCA and ultrasound arms were not statistically significant.92 In the PLCO91 there was not a statistically significant difference in the proportion of cases identified at the localized stage in the intervention versus usual care group (15% versus 10%, p =0.08) and comparisons were not statistically different when comparing localized and regional cancer cases to more advanced cancers. In addition, an analysis of data from PLCO found that, compared with the usual care arm, fewer cancers in the intervention arm were categorized as type II ovarian cancers. Additionally, within the screening arm, fewer type II cancers were screen-detected. Overall, type II tumors were less likely diagnosed from screening and were diagnosed at a later stage. The authors suggested that overdiagnosis of more indolent
cancer types could, in part, account for the lack of a mortality benefit from ovarian cancer screening in the trial.117

Only the UKCTOCS reported cancer cases by type and study arm, but did not conduct statistical tests of type differences. A greater proportion of cancers occurring in the control condition were type II cancers, which tend to be more aggressive; 60 percent cases in the TVU and CA-125 ROCA screening arms were type II compared to 64 percent in the control arm.92 Conversely, more cases in the screening arms were non-epithelial and borderline cancer types, which tend to progress more slowly or remain indolent; 16 percent in the CA-125 ROCA arm, 20 percent in the TVU arm, and 11 percent in the no screening arm. Borderline and non-epithelial types are more likely to be early stage cancers, and have high rates of survival.7 Thus, detection of more cases of indolent disease with screening may have had limited impact on ovarian cancer survival, even though it appears to contribute to an observed shift in disease stage. Among invasive epithelial and peritoneal cancer cases found, however, the CA-125 ROCA screening test appeared to identify slightly more cases at an earlier stage. Notably, however, even stage I cancers in some type II high grade epithelial carcinomas may be associated with microscopic metastases, as cancer cells can be present in ascites (stage Ic).2, 3, 5, 6 In addition, there were no differences in treatment approach by study arm in the PLCO; 81 percent in the intervention group received surgery plus systemic therapy, compared with 80 percent in the usual care group.91 Treatment outcomes for participants in the UKCTOCS92 have not yet been published. While there is some evidence of a stage shift for the CA-125 ROCA and ultrasound intervention arms in the UKCTOCS, this shift did not confer a statistically significant mortality benefit.

The absence of a mortality benefit in large, well-conducted trials has contributed to concerns the most common cancer type spreads so rapidly that it cannot be identified at an early enough stage to substantially shift mortality. The stage shift in UKCTOCS trial would seem to counter this, but the lack of mortality benefit may indicate that the early stage tumors detected are more aggressive tumor phenotypes with poor prognosis even when identified at earlier stages. Recent work to refine the distinctions among ovarian cancer molecular, pathological, and clinical characteristics highlight this point in noting that survival differences are more likely attributable to type than to stage at diagnosis, with the most common type II cancers being particularly lethal regardless of stage, likely owing to microscopic lesions that are not detectable before significant spread has occurred.6

Discussion of Harms

The UKCTOCS employed a more nuanced approach to CA-125 testing and triage by using an algorithm that incorporates CA-125-level trajectories, assigning three levels of risk to direct surveillance and triage tests. This was aimed at reducing rates of surgical investigation, and indeed surgery rates were lower in the CA-125 ROCA arm than in the ultrasound-only arm of the trial. Accordingly, false positive surgery rates in the CA-125 ROCA arm of the UKCTOCS were markedly lower than in the PLCO (1% versus 3%).

The surgical complication rates differed considerably for the PLCO and UKCTOCS, with 15 percent of women who underwent false positive surgery experiencing a major complication in
the PLCO and just over 3 percent having a major complication from false positive surgery in the UKCTOCS. Differences in the study settings could account in part for this difference, as diagnostic testing in the PLCO was conducted through referrals to women’s routine sources of care, and not necessarily specialized tertiary care settings. In contrast, in the U.K. all women referred for diagnostic testing were seen at National Health Services tertiary care surgical centers. It is unclear whether the complication rates observed in the PLCO would be observed in current U.S. community surgical practices for women referred for diagnostic testing from primary care. Current U.S. data on the complication rates for diagnostic oophorectomy (alone) are not available, but it is likely that rates vary by setting, region, and clinician characteristics. Regardless of the complication rates, high rates of surgery and removal of an ovary or ovaries in the absence of disease occurred in both trials, although rates of surgery were lowest in the CA-125 ROCA screening intervention in the UKCTOCS trial.

False positive surgical investigations in the included trials were reported to often include bilateral salpingo-oophorectomy (BSO), based on common practice for the investigation of suspected ovarian cancer. Given that these screening interventions were undertaken in healthy women, potential harms of unnecessary removal of the ovaries in postmenopausal women deserves scrutiny. A recent systematic review evaluated the general health consequences of BSO at the time of hysterectomy for benign indications. Health consequences for women undergoing surgical investigation for a false positive ovarian cancer screen receiving BSO would likely be comparable. Although evidence was somewhat limited, the review found reductions in ovarian cancer and in rates of reoperation for women who underwent BSO, particularly for women younger than 45 years old. There was also evidence of potential adverse effects on cardiovascular health and all-cause mortality. Sexual function may also be negatively impacted by BSO conducted at the time of benign hysterectomy. Thus, the removal of ovaries and/or fallopian tubes at the time of a surgical investigation for a false positive screening test result may have downstream harms beyond those owing to the direct effects of surgery on health outcomes. Evidence on the effects of BSO among older postmenopausal women similar to those included in this trial is limited, and firm conclusions about adverse or beneficial health or social effects are not possible.

Limitations of the Review

Although a body of evidence on the test performance of various screening strategies exists, the most promising approaches using ultrasound and CA-125 have been assessed in trials. Our review did not consider observational evidence, where some tools have appeared promising in early investigations.

Given the rarity of ovarian cancer, and the invasiveness of diagnostic surgery for positive screening results, the mortality reduction from screening relative to an unscreened group is key evidence for this condition, as it summarizes the net effect of screening, detection, and treatment. An effective ovarian cancer screening program among asymptomatic average-risk women would be hypothesized to save lives through lower rates of death from ovarian cancer. Observation of a cancer stage shift toward more localized cancers may lead to less morbid treatments and could underlie an observed screening mortality benefit, but stage shifts do not
necessarily confer a mortality benefit. Namely, a screening test might identify cancers that would not have progressed or earlier treatment might not sufficiently change the survival rates of women to make a difference relative to an unscreened group. Even in the absence of an organized screening program, asymptomatic ovarian cancers may be detected opportunistically. Evidence from randomized trials of ovarian cancer-screening programs that report mortality outcomes can establish whether specific screening protocols result in better health outcomes (e.g., reduced mortality) than usual care or the absence of the screening program. This direct evidence is available from two large trials, and neither provides evidence of a screening benefit with any of the screening protocols tested.

Given the low incidence of ovarian cancer, very large trials are necessary to determine whether the benefits of a screening program outweigh the harms of diagnostic testing, which for ovarian cancer necessarily involves surgery and ovarian removal. We are confident that our review identified all relevant trials with ovarian cancer-mortality outcomes reported. Two additional trials that did not meet our inclusion criteria on the basis of study design and outcome reporting, but enrolled large samples of women to ovarian cancer screening would not have changed our findings had they been included.

The Shizuoka Cohort Study of Ovarian Cancer Screening and Shizuoka Cancer Registry (SCOSCS trial) randomized asymptomatic postmenopausal women in Shizuoka, Japan, to screening using ultrasonography and CA-125 (with a cutoff of 35 U/ml) (n= 41,688) or followup without screening (n=40,799).121 Screenings were repeated yearly for an average of 5.4 screens and a mean followup of 9.2 years. Information on the impact of screening on ovarian cancer diagnosis was published in 2008, with no significant difference between the number of women with ovarian cancer detected between the screening and control groups. Differences in the percent of patients detected with stage I cancer were also not statistically significant (63% versus 38%, p=0.23). Mortality data have not been published from this trial, and it is unclear if additional analyses are planned.121 The SCOSCS was not included in the results of this review because health outcomes (e.g., mortality) have not been reported for this trial.

The University of Kentucky Ovarian Cancer Screening Trial was initiated in 1987 to assess the use of annual transvaginal ultrasound to detect ovarian cancer in asymptomatic women aged 50 or older and women aged 25 or older with a family history of ovarian cancer. 122 The study was a controlled trial, but did not randomize women to a control condition, instead comparing the screened cohort to a cohort of women with ovarian cancer outcomes reported in the Kentucky Tumor Registry. Data published on the effect of screening TVU on over 37,000 women from 1987 to 2011 reported that in the screening cohort 47 percent of cancers were detected at stage I and 70 percent were detected with stage I or II disease. In contrast, only 27 percent of those entered into the Kentucky Tumor Registry during the same time period had stage I or II disease (p<0.01). The 5-year disease-free survival rate for those cancers detected by screening was 85 percent compared with 54 percent of unscreened women treated at the same cancer center using the same surgical and chemotherapeutic protocols (p<0.001).122 This study also reported fewer surgical complications (10% of subjects) compared with the PLCO.123 Our review did not include this evidence because the screening and comparison cohorts included both average-risk and high-risk women, and the participants were not randomly assigned to study groups. Further analyses of this investigation have not been reported, and there are limits with regard to
conclusions that can be drawn about the effectiveness of a TVU screening program based on a cohort comparison study design.

The scope of this review was limited to the type of evidence that would be necessary to inform a change in clinical practice in accordance with USPSTF standards. While some topics evaluate the effectiveness of screening through an indirect pathway logic model, considering the performance of a screening test separately from the effectiveness of treatment. Evidence that tests the effect of screening compared with the absence of screening on intended health outcomes in a randomized design does not require as much inference across heterogeneous bodies of evidence. Nevertheless, trial evidence can have limitations in terms of generalizability and applicability to usual care. The PLCO\textsuperscript{91} is more applicable to a U.S. setting than the UKCTOCS,\textsuperscript{92} since the PLCO referred women to usual care settings. The low surgical complication rates from surgery seen in the UKCTOCS\textsuperscript{92}, for example, may have been due to the receipt of care in tertiary care centers which is standard in the U.K. health system. Similarly, screening tests offered through a trial might be more accurate than screening performance in routine care settings, or surgical investigations might be more common in the absence of trial protocols.

### Future Research Needs

Given null findings from two major, well-powered trials, future research may focus on identifying women at elevated risk of ovarian cancer that could benefit from screening or prophylactic interventions. More work is needed, however, to develop approaches for assessing family history and ovarian cancer risk in primary care and to optimize women’s use of genetic counselors. Research is also needed to identify new markers with greater sensitivity and specificity for detection of ovarian cancer in average-risk women. In 2016, the IOM recommended additional research should focus on the development and assessment of early detection strategies that extend beyond current imaging modalities and biomarkers and reflect the pathobiology of each ovarian cancer subtype.\textsuperscript{2}

The appearance of diverging cumulative mortality curves in the later years of the UKCTOCS has been a focus of the study investigators and critical commentaries. Given the absence of any effective screening modality to reduce mortality from ovarian cancer, any hint at a possible benefit merits close attention. As the UKCTOCS investigators point out, a few more years of data from the trial are needed to accrue followup data on women who entered the trial later in the enrollment period so that the complete findings are included in the later years of the analysis. Nevertheless, questions can be asked in the meantime about the mechanism that might underlie a screening intervention benefit for ovarian cancer occurring several years after the screening program ended. Given the natural history of ovarian cancer, it is unclear how a screening intervention aimed at identifying ovarian cancer and intervening at a more treatable stage would have a delayed effect. One possible explanation is that the screening activities resulted in the removal of the abnormal appearing but nonmalignant ovaries and fallopian tubes of women, and some of these otherwise might have gone on to develop ovarian cancer in later years. Given the relatively high false positive surgery rates seen in this trial, it is possible that prophylactic removal of selected women’s ovaries and fallopian tubes would have effects on the later
divergence of the cumulative mortality curves. A sensitivity analysis of the trial data using ovaries at risk, rather than women at risk as the dominator, might shed light on this potential effect on the trial results. The absence of an overall difference in ovarian cancer incidence suggests that the higher rate of oophorectomy may not be influential, but analyses examining the cumulative ovarian cancer incidence over the course of the trial and with more followup data could be informative.

More detailed data on the surgeries following screening would shed further light on possible interpretations of the diverging cumulative risk curves highlighted by the study authors. While rates of ovarian cancer have been found to be lower among women opting for removal at the time of hysterectomy for benign pathology or elective reasons, reductions in the risk of rare ovarian cancers have to be weighed against possible negative effects of oophorectomy for other conditions, and risks associated with surgery.

Evidence of health risks associated with removal of the ovaries, although limited among postmenopausal women, includes negative consequences for cardiovascular health, sexual function, and some mental health outcomes. Given growing recognition that many ovarian cancers originate in the fallopian tubes, prophylactic salpingectomy with ovarian conservation at the time of surgery for other indications is viewed as a potentially promising preventive strategy for women undergoing pelvic surgery for other reasons. Overall, in the United States, rates of opportunistic salpingectomy are increasing, although opportunistic BSO is far more common. Nevertheless, ACOG does not endorse routine salpingectomy at the time of benign indication hysterectomy or sterilization until rigorous observational or trial evidence on this intervention is available. Data from a large community-based health system in Northern California reported an increase from 15 percent to 72 percent between 2011 and 2014 in the practice of salpingectomy at the time of elective hysterectomy for benign indications following the distribution of resources related to the potential benefits of the practice.

Differences in the oophorectomy rates were also estimated in the PLCO, based on a supplemental survey following the screening period (66% of participants responded). More women in the intervention arm reported an oophorectomy (7.7% versus 5.8%, RR 1.33 [95% CI, 1.24 to 1.43]). The PLCO did not find a long-term benefit of screening; the cumulative mortality effect estimate was in the direction of harms rather than benefits of screening. In the longer-term followup data from the trial, however, the intervention effect estimate moved closer to null.

We identified no ongoing randomized trials of ovarian cancer screening using new screening tools. While some tools in development may hold promise for the future (e.g., microRNA), currently there are no new screening tools (i.e., biomarkers, instruments) exhibiting levels of test performance beyond what is observed for the screening tools evaluated in trials.

The UKCTOCS trialists are engaged in efforts to improve upon the ROCA algorithm, adding other protein markers along with CA-125 to new prediction models derived using data from the UKCTOCS data. These models would require further validation and testing to ascertain whether they truly represent improvements on the ROCA algorithm that would potentially attain clinical benefits for ovarian cancer detection and treatment. In any case, given the absence of a single marker or screening device that is effective for ovarian cancer, research is likely to increasingly
Methods for identifying women at high-risk for ovarian cancer may help to direct preventive interventions. In 2016, a report from the Institute of Medicine recommended that research focus on developing and validating dynamic risk assessment tools for identifying those at high risk of cancer, including increasing rates of genetic counseling, and to exam the risk-benefit balance of nonsurgical and surgical prevention strategies in these populations.2

Conclusion

Since the previous review, results from a large trial conducted in the United Kingdom were published. Ovarian cancer mortality did not differ between control and intervention screening conditions in any of the included trials, including two good-quality studies with adequate power to detect differences. Harms of screening include surgery resulting from a false positive. These surgeries often result in the removal of one or both ovaries and/or fallopian tubes, and can lead to major surgical complications. Reports from the UKCTOCS of a potential delayed effect of screening on ovarian cancer mortality require further follow data to evaluate, but the causal mechanism for a delayed screening effect is unclear. Major trials of promising ovarian cancer screening tools have null findings to date among healthy average-risk women, and there are considerable harms associated with screening. Further analysis of existing trials and research on new biomarkers, new risk-assessment tools, and trials of prophylactic interventions may ultimately be found to be useful in reducing ovarian cancer mortality and will need to be weighed against known screening harms.
References

1. National Cancer Institute. Ovarian Epithelial, Fallopian Tube, and Primary Peritoneal

2. Institute of Medicine, Committee on the State of the Science in Ovarian Cancer Research,
Board on Health Care Services, et al. Ovarian Cancers: Evolving Paradigms in Research

3. Prat J, FIGO Committee on Gynecologic Oncology. Staging classification for cancer of
PMID: 24219974. http://dx.doi.org/10.1016/j.iijo.2013.10.001


6. Kurman RJ, Shih le M. The Dualistic Model of Ovarian Carcinogenesis: Revisited,
http://dx.doi.org/10.1016/j.ajpath.2015.11.011

7. McCluggage WG. Morphological subtypes of ovarian carcinoma: a review with emphasis
http://dx.doi.org/10.1097/PAT.0b013e328348a6e7


2016.

http://dx.doi.org/10.1155/2014/934261

15. Reade CJ, McVey RM, Tone AA, et al. The fallopian tube as the origin of high grade

16. Kurman RJ, Shih le M. Molecular pathogenesis and extraovarian origin of epithelial
ovarian cancer--shifting the paradigm. Hum Pathol. 2011;42(7):918-31. PMID:
21683865. http://dx.doi.org/10.1016/j.humpath.2011.03.003


47. Parker WH, Broder MS, Chang E, et al. Ovarian conservation at the time of hysterectomy and long-term health outcomes in the nurses' health study. Obstet Gynecol. 2009;113(5):1027-37. PMID: 19384117. [http://dx.doi.org/10.1097/AOG.0b013e3181a11c64](http://dx.doi.org/10.1097/AOG.0b013e3181a11c64)


1. Does screening for ovarian cancer in asymptomatic women using a single test or combined algorithm [such as, but not limited to, testing for serum cancer antigen (CA–125) and ultrasonography] reduce all-cause or disease-specific morbidity and mortality?
2. What are the harms of screening for ovarian cancer, including harms of the screening test and of diagnostic evaluation?
Table 1. Characteristics of Ovarian Cancer Screening Randomized Trials

<table>
<thead>
<tr>
<th>Trial, year of publication</th>
<th>UKCTOCS, 2016&lt;sup&gt;92&lt;/sup&gt;</th>
<th>PLCO, 2011&lt;sup&gt;91&lt;/sup&gt;</th>
<th>QUEST, 2007&lt;sup&gt;97&lt;/sup&gt;</th>
<th>U.K. Pilot, 1999&lt;sup&gt;98&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Good</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>Included key questions</td>
<td>1, 2</td>
<td>1, 2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>Randomization allocation</td>
<td>1:1:2 (CG)</td>
<td>1:1</td>
<td>1:1</td>
<td>1:1</td>
</tr>
<tr>
<td>N randomized</td>
<td>202,638</td>
<td>78,216†</td>
<td>592</td>
<td>21,955</td>
</tr>
<tr>
<td>N analyzed</td>
<td>202,546</td>
<td>68,557†</td>
<td>549</td>
<td>21,935</td>
</tr>
<tr>
<td>Key outcomes reported‡</td>
<td>KQ1: Ovarian cancer incidence and mortality</td>
<td>KQ1: Ovarian cancer incidence and mortality</td>
<td>KQ2: Psychological harms of screening program participation</td>
<td>KQ1: Ovarian cancer incidence and mortality§</td>
</tr>
<tr>
<td></td>
<td>KQ2: Screening false positive rates and surgical complications</td>
<td>KQ2: Screening false positive rates and surgical complications</td>
<td>KQ2: Screening false positive rates and surgical complications</td>
<td></td>
</tr>
<tr>
<td>Enrollment/ recruitment source</td>
<td>National Health Service catchments of 13 regional centers in Wales, England, and Northern Ireland; women recruited from 27 primary care service groups in the regions.</td>
<td>Community volunteers from the catchment areas of 10 screening centers</td>
<td>Population volunteers, physician referral</td>
<td>Community volunteers and postal invitations to 40 primary care practices in England, Scotland, and Wales</td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td>Post-menopausal, age 50-74</td>
<td>Aged 55-74</td>
<td>Age ≥30</td>
<td>Post-menopausal, age ≥45</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>Self-reported history of bilateral oophorectomy or ovarian malignancy, increased risk of familial ovarian cancer, active nonovarian malignancy</td>
<td>Previous bilateral oophorectomy; history of lung, colorectal, or ovarian cancer; current treatment for cancer other than nonmelanoma skin cancer; colonoscopy, sigmoidoscopy, or barium enema in past 3 years; pervious surgical removal of lung or entire colon; participation in other screening trial‖</td>
<td>High risk of ovarian cancer¶; cancer diagnosis in past year; plans to become pregnant in the following 2 years</td>
<td>History of bilateral oophorectomy, ovarian cancer, or any active malignancy</td>
</tr>
</tbody>
</table>

* Additional mortality data published through 2012<sup>112</sup>
† 9,659 women excluded from analysis due to oophorectomy prior to trial entry (included in n randomized because they were screened for other cancers in PLCO)
‡ Ovarian cancer was defined according to the 2014 World Health Organization and Fédération Internationale de Gynécologie Obstétrique definitions which includes ovarian, tubal, and peritoneal cancers.<sup>90</sup> Cancer cases were abstracted or calculated using this definition when possible, even if it was not the primary outcome reported in the trial.
§ Data from the U.K. Pilot trial did not capture information related to peritoneal cancer; therefore, these results are limited to primary cancer of the ovary and fallopian tubes.
‖ Exclusion based on CRC screening began in April 1995. Trial initially excluded women with of previous oophorectomy (dropped in 1996) and current tamoxifen use (dropped in 1999)
¶High risk of ovarian cancer: reported family history predicted at least a 10% probability of a germline mutation in the BRCA1 or BRCA2 genes or Amsterdam criteria for hereditary nonpolyposis colorectal cancer syndrome<sup>129</sup>

Abbreviations: CG = control group; N = number; NR = not reported; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; QUEST = Quality of life, Education, and Screening Trial; U.K. = United Kingdom; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening; U.S. = United States.
<table>
<thead>
<tr>
<th>Trial, year of publication</th>
<th>UKCTOCS, 2016&lt;sup&gt;92&lt;/sup&gt;</th>
<th>PLCO, 2011&lt;sup&gt;91&lt;/sup&gt;</th>
<th>QUEST, 2007&lt;sup&gt;97&lt;/sup&gt;</th>
<th>U.K. Pilot, 1999&lt;sup&gt;98&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Good</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>Mean age, years</td>
<td>61</td>
<td>NR</td>
<td>NR</td>
<td>56</td>
</tr>
<tr>
<td>Age distribution</td>
<td>IQR: 56.0-66.2</td>
<td>55-59: 34.2%</td>
<td>30-49: 54.1%*</td>
<td>45-54: 40.6%†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60-64: 30.3%</td>
<td>&gt;49: 45.9%*</td>
<td>55-64: 48.4%†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65-69: 21.9%</td>
<td></td>
<td>65-74: 10.0%†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70-74: 13.6%</td>
<td></td>
<td>&gt;74: 1.0%†</td>
</tr>
<tr>
<td>Race (%)</td>
<td>White: 96.4</td>
<td>White, non-Hispanic: 88.5</td>
<td>White, non-Hispanic: 95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black: 1.4</td>
<td>Black, non-Hispanic: 5.7</td>
<td>Other: 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asian: 0.9</td>
<td>Hispanic: 1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other: 0.8</td>
<td>Asian/Pacific Islander: 3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>American Indian/Alaskan Native: 0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior hysterectomy (%)</td>
<td>18.8</td>
<td>27.2</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Personal history of cancer (%)</td>
<td>6.0</td>
<td>Breast cancer: 3.6</td>
<td>9.9*</td>
<td>NR</td>
</tr>
<tr>
<td>Family history of breast or ovarian cancer (%)</td>
<td>Maternal ovarian: 1.6</td>
<td>17.4</td>
<td>17.1</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Maternal breast: 6.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Percent in screening arms only
† Data from initial prevalence screen study, n=22,000

Abbreviations: IQR = interquartile range; NR = not reported; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; QUEST = Quality of life, Education, and Screening Trial; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening
## Table 3. Ovarian Cancer Screening Trial Protocols

<table>
<thead>
<tr>
<th>Trial, year of publication</th>
<th>UKCTOCS, 2016&lt;sup&gt;92&lt;/sup&gt;</th>
<th>PLCO, 2011&lt;sup&gt;91&lt;/sup&gt;</th>
<th>U.K. Pilot, 1999&lt;sup&gt;98&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality</strong></td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>
| **Screening intervention**| Arm 1: Initial CA-125 testing with ROCA algorithm used to determine risk and followup testing and assessment through clearly defined, trial specific protocols.; followup included repeat CA-125 test (intermediate risk) or repeat CA-125 and ultrasound (elevated risk) based on ROCA.  
Arm 2: Ultrasound (primarily transvaginal); followup ultrasound for unsatisfactory or abnormal ultrasounds. | Ultrasound (mainly TVU) and CA-125§ | Initial CA-125 testing; followup included ultrasound for elevated CA-125 levels. |
| **Study definition of positive screening result** | Referral to surgery for suspected ovarian cancer following positive screening and clinical assessment. | Abnormal CA-125 or ultrasound | Referral to surgery for suspected ovarian cancer. |
| **Followup protocol for women with positive screening test result** | Clinical assessment and surgical investigation conducted by trial clinicians. | Notification of patients and their primary care physicians; community care. | Referral through family physician to a gynecologist for surgical investigation. |
| **Comparison group**      | No screening              | Standard community care | No screening                |
| **Screening frequency**   | Annual                   | Annual                 | Annual                      |
| **Maximum number of screening rounds** | 6 (original protocol)  
7-11 (extended screening based on interim analysis) | CA-125: 4-6 depending on enrollment date. TVU: 4 | 3# |
| **Median length of followup, years (range)** | 11.1 (IQR 10.0 - 11.0)  
(Range 0 – 13.6) | 12.4 (IQR 10.9 -13)  
(Range 0 - 8) | NR |
| **Positive test result definitions** | CA-125 ROCA: Intermediate risk (risk ≥1/1818), elevated risk (risk ≥1/500)*  
Ultrasound: One or both ovaries with complex morphology, simple cysts greater 60 cm³, or ascites. | CA-125: ≥35 U/mL  
Ultrasound: Ovarian volume >10 cm³; Cyst volume >10 cm³; Any solid area or papillary projection extending into the cavity of a cystic ovarian tumor of any size; or any mixed (solid and cystic) component within a cystic ovarian tumor. | CA-125 ≥30 U/mL |
| **Definition of cancer**  | Primary outcome: Malignant neoplasms of the ovary including epithelial and nonepithelial ovarian cancer and malignant neoplasms of the fallopian tube† and undesignated malignancies of the ovaries, fallopian tube, or peritoneum.  
Secondary outcome: Ovarian (as defined above) and primary peritoneal cancer‡. | Ovarian, primary peritoneal, and fallopian tube cancers‖¶ | Invasive primary epithelial cancers of the ovary and fallopian tube. |
Table 3. Ovarian Cancer Screening Trial Protocols

<table>
<thead>
<tr>
<th>Identification of ovarian cancer cases and deaths</th>
<th><strong>UKCTOCS, 2016</strong>&lt;sup&gt;92&lt;/sup&gt;</th>
<th><strong>PLCO, 2011</strong>&lt;sup&gt;91&lt;/sup&gt;</th>
<th><strong>U.K. Pilot, 1999</strong>&lt;sup&gt;98&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative records (e.g., National Health Service), cancer registries, followup questionnaires, direct communication with participants, their families, and physicians, death certificates.</td>
<td>Annual questionnaire to participants, National Death Index, and population-based cancer registries (when possible).</td>
<td>National Health Service, followup questionnaires, communication with physicians and participant families, death certificates.</td>
<td></td>
</tr>
</tbody>
</table>

| Confirmation of ovarian cancer cases and deaths | Blinded review of diagnosis and mortality by members of a designated panel (2 pathologists, 2 gynecological oncologists) | Blinded review of mortality by qualified (i.e., epidemiology, surgery, medicine, radiation oncology) member of a designated panel | Blinded pathology review of diagnosed cases |

* CA-125 levels were changed in 2005 to maintain the % in each risk level (intermediate ≥1/3500, elevated ≥1/1000). 84.6% of screens were classified using pre-2005 cutoffs.
† ICD-10 C56, C57
‡ ICD-10 C48.1, C48.2
§ Annual bimanual clinical examination of the ovaries discontinued in 1998 because no cases identified solely with this screening test
‖ ICD for Classification of Diseases for Oncology, Revision 2: C569, C481, C482, C570
¶ Borderline tumors were considered false positive screens
# All of the women in this trial (including the control group) had undergone a previous round of screening approximately 10 years prior

**Abbreviations:** CA-125 = cancer antigen 125; cm = centimeter; IQR = interquartile range; NR = not reported; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; ROCA = Risk of Ovarian Cancer Algorithm TVU = transvaginal ultrasound; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening; U/mL = units per milliliter
Table 4. Adherence to Screening in Ovarian Cancer Screening Trials

<table>
<thead>
<tr>
<th>Trial, Year of publication</th>
<th>UKCTOCS, 2016&lt;sup&gt;92&lt;/sup&gt;</th>
<th>PLCO, 2011&lt;sup&gt;91&lt;/sup&gt;</th>
<th>U.K. Pilot, 1999&lt;sup&gt;98&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Completed at least one screening round (%)</td>
<td>CA-125 ROCA: 98.9 ultrasound: 95.3</td>
<td>NR</td>
<td>85.5</td>
</tr>
<tr>
<td>Screening adherence (%)</td>
<td>CA-125 ROCA: 80.8 (% of screens attended) ultrasound: 78.0 (% of screens attended) Adherence in each round: CA-125 ROCA: 47.2-98.4 ultrasound: 35.9-94.9</td>
<td>ultrasound: 78-84 CA-125: 73-85</td>
<td>1st round: 79.7 2nd round: 79.3 3rd round: 77.4 (70.7% completed all 3 screens)</td>
</tr>
<tr>
<td>Screening contamination in control group %</td>
<td>4.3*</td>
<td>ultrasound: 2.3 - 3.2 per year CA-125: 2.7 - 4.6 per year</td>
<td>NR</td>
</tr>
</tbody>
</table>

* Self-reported ultrasound or CA-125 screening in the control arm over the entire trial period

**Abbreviations:** CA-125 = cancer antigen 125; NR = not reported; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; U.K. = United Kingdom; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening
Table 5. Benefits Reported in Ovarian Cancer Screening Trials: Ovarian Cancer Mortality*

<table>
<thead>
<tr>
<th>Trial, Year of publication</th>
<th>UKCTOCS, 2016(^92) (CA-125 ROCA arm)</th>
<th>UKCTOCS, 2016(^92) (ultrasound arm)</th>
<th>PLCO, 2011(^91)</th>
<th>U.K. Pilot, 1999(^98)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>N analyzed per arm</td>
<td>IG: 50,624</td>
<td>IG: 50,623</td>
<td>IG: 34,253</td>
<td>IG: 10,958</td>
</tr>
<tr>
<td>Ovarian cancer cases, n (%)</td>
<td>IG: 354 (0.7)</td>
<td>IG: 324 (0.6)</td>
<td>IG: 212 (0.6)</td>
<td>IG: 16 (0.1)</td>
</tr>
<tr>
<td></td>
<td>CG: 645 (0.6)</td>
<td>CG: 645 (0.6)</td>
<td>CG: 176 (0.5)</td>
<td>CG: 20 (0.2)</td>
</tr>
<tr>
<td>Ovarian cancer incidence rate</td>
<td>IG: 6.4 per 10,000 p-y</td>
<td>IG: 5.9 per 10,000 p-y</td>
<td>IG: 5.7 per 10,000 p-y</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>CG: 5.9 per 10,000 p-y</td>
<td>CG: 5.9 per 10,000 p-y</td>
<td>CG: 4.7 per 10,000 p-y</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer incidence between group difference</td>
<td>NR</td>
<td>NR</td>
<td>RR: 1.21 (95% CI, 0.99 to 1.48)</td>
<td>NR</td>
</tr>
<tr>
<td>Ovarian cancer mortality, n (%)</td>
<td>IG: 160 (0.32)</td>
<td>IG: 163 (0.32)</td>
<td>IG: 118 (0.34)‡</td>
<td>IG: 9 (0.082)</td>
</tr>
<tr>
<td></td>
<td>CG: 358 (0.35)</td>
<td>CG: 358 (0.35)</td>
<td>CG: 100 (0.29)‡</td>
<td>CG: 18 (0.16)</td>
</tr>
<tr>
<td>Ovarian cancer mortality rate</td>
<td>IG: 2.9 per 10,000 p-y</td>
<td>IG: 3.0 per 10,000 p-y</td>
<td>IG: 3.1 per 10,000 p-y</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>CG: 3.3 per 10,000 p-y</td>
<td>CG: 3.3 per 10,000 p-y</td>
<td>CG: 2.6 per 10,000 p-y</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer mortality between group difference</td>
<td>HR: 0.89 (95% CI, 0.74 to 1.08), p=0.23†</td>
<td>HR: 0.91 (0.76 to 1.09), p=0.31†</td>
<td>RR: 1.18 (95% CI, 0.82 to 1.71) p=NR§</td>
<td>RR: 0.50 (95% CI, 0.22 to 1.11) p=0.083¶</td>
</tr>
<tr>
<td>Ovarian cancer survival</td>
<td>NR</td>
<td>NR</td>
<td>Survival difference from date of randomization p = 0.67</td>
<td>IG: survival 72.9 months (median)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Survival difference from date of diagnosis (lead time bias), p=0.18.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Survival difference from date of randomization, p=0.011#</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer treatment</td>
<td>NR</td>
<td>NR</td>
<td>Surgery with systemic chemotherapy IG: 171 (81%)</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG: 140 (80%) p=NR, NS</td>
<td></td>
</tr>
</tbody>
</table>

* Includes ovarian, fallopian, and primary peritoneal cancers if reported
† Cox model
‡ Extended mortality data reported in results text
§ Sequentially adjusted
‖ Does not include peritoneal cancer
¶ Calculated (article reports RR calculated in in terms of increased relative risk)
# Log rank

**Abbreviations:** CG = control group; CI = confidence interval; HR = hazard ratio; IG = intervention group; n = number; NR = not reported; NS = not statistically significant; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; p-y = person-years; RR = rate ratio; U.K. = United Kingdom; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening
Table 6. Harms Reported in Ovarian Cancer Screening Trials: Positive Testing, False Positive Testing, and Surgical Complications in Ovarian Cancer Screening*

<table>
<thead>
<tr>
<th>Trial, year of publication</th>
<th>UKCTOCS, 2016 (CA-125 ROCA arm)</th>
<th>UKCTOCS, 2016 (ultrasound arm)</th>
<th>PLCO, 2011</th>
<th>U.K. Pilot, 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Prevalence of positive screening test results, by round</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% (n positive test/N screened)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative prevalence of positive screening test results</td>
<td></td>
<td>NR</td>
<td>Overall screening program: 9.8 (3,358/34,253)</td>
<td>Overall screening program: 4.3 (468/10958)</td>
</tr>
<tr>
<td>% (n with positive test/N screened)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2 to 11: 44.3 (20,485/46,237)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False positive rate: Women without cancer who had a positive screening result</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% (n with false positive screen/N women without cancer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence round: 9.0 (4,513/50,031)§</td>
<td>Prevalence round: 11.9 (5,734/48,177)§</td>
<td>Overall screening program: 9.6 (3,285/34,041)</td>
<td>Overall screening program: 4.2 (462/10,942)§</td>
<td></td>
</tr>
<tr>
<td>% (n surgery/N women without cancer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False positive surgery rate: Women without cancer undergoing surgery</td>
<td>0.97 (488/50,270)</td>
<td>3.25 (1634/50,299)</td>
<td>3.17 (1,080/34,041)</td>
<td>0.2 (23/10,942)§</td>
</tr>
<tr>
<td>Screening test complications</td>
<td>0.86 per 10,000 screens†</td>
<td>1.86 per 10,000 screens†</td>
<td>CA-125: 58.3 per 10,000 women§§</td>
<td>NR</td>
</tr>
<tr>
<td>Women without cancer with surgical complications</td>
<td>3.07 (15/488)¶</td>
<td>3.49 (57/1,634)‡‡</td>
<td>15.09 (163/1,080)‖</td>
<td>0 (0)</td>
</tr>
<tr>
<td>% (n complication/N with false positive surgery)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deaths from other causes</td>
<td>IG: 3376 (6.7%) 61.5 per 10,000 p-y CG: 6658 (6.6%) 60.7 per 10,000 p-y RR: 0.99, p=0.65**</td>
<td>IG: 3262 (6.4%) 59.4 per 10,000 p-y CG: 6658 (6.6%) 60.7 per 10,000 p-y RR: 0.99, p=0.65**</td>
<td>IG: 2924 (76.6 per 10,000 p-y)¶¶ CG: 2914 (76.2 per 10,000 p-y)¶¶ RR: 1.01 (95% CI, 0.96 to 1.06) p=NR</td>
<td>NR</td>
</tr>
<tr>
<td>§ Does not include peritoneal cancers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>† Includes: Bruising (13), pain (8), hematoma (3), fainting (1), cystitis/infection (1), other (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Includes ovarian, fallopian, and primary peritoneal cancers if reported
† Patient experience of first positive screening test result leading to additional triage/followup (including repeated testing due to unsatisfactory results).
‡ Among women with false positive results/benign findings who underwent surgery
§ Does not include peritoneal cancers
‖ Includes: Bruising (13), pain (8), hematoma (3), fainting (1), cystitis/infection (1), other (4)
### Table 6. Harms Reported in Ovarian Cancer Screening Trials: Positive Testing, False Positive Testing, and Surgical Complications in Ovarian Cancer Screening*

<table>
<thead>
<tr>
<th>Harm Category</th>
<th>Number of Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anesthetic complications</td>
<td>1</td>
</tr>
<tr>
<td>Injury to hollow viscus</td>
<td>2 gastrointestinal, 1 bladder</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>2</td>
</tr>
<tr>
<td>Deep vein thrombosis</td>
<td>1</td>
</tr>
<tr>
<td>Bowel obstruction</td>
<td>4</td>
</tr>
<tr>
<td>Wound breakdown</td>
<td>1</td>
</tr>
<tr>
<td>Significant ileus</td>
<td>1</td>
</tr>
<tr>
<td>Uterine perforation</td>
<td>1</td>
</tr>
<tr>
<td>Infection</td>
<td>1</td>
</tr>
</tbody>
</table>

- Includes: anesthetic complications (1), injury to hollow viscus (2 gastrointestinal, 1 bladder), hemorrhage (2), deep vein thrombosis (1), bowel obstruction (4), wound breakdown (1), significant ileus (1), uterine perforation (1), infection (1)

# Excludes ovarian and primary peritoneal cancer deaths

** RR mortality ratio for no screening group vs CA-125 ROCA and ultrasound groups combined

†† Includes: Pain (20), cystitis/infection (11), discomfort (5), bruising (2), fainting (1), other (22)

‡‡ Includes: Injury to hollow viscus (4 gastrointestinal, 3 bladder, 1 ureter), hemorrhage (11), anesthetic complication/myocardial infarction (3), hernia (6), deep vein thrombosis/pulmonary embolism (3), wound breakdown (6), bowel obstruction (4), wound/supravaginal hematoma (4), infection (6), pain with readmission or further operation (3)

§§ Minor complications (e.g., fainting, bruising)

‖‖ 222 total complications in 163 patients. Includes: Infection (89), direct surgical harms (63), cardiovascular or pulmonary events (31), or other (39)

¶¶ Excludes deaths from ovarian, colorectal, and lung cancer

**Abbreviations:** CA-125 = cancer antigen 125; CG = control group; CI = confidence interval; IG = intervention group; n = number; NR = not reported; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; P-Y = person-years; RR = relative risk; TVU = transvaginal ultrasound; U.K. = United Kingdom; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening
Table 7. Harms Reported in Ovarian Cancer Screening Trials: Psychological Effects of Screening

<table>
<thead>
<tr>
<th>Trial, Author, Year of publication</th>
<th>UKCTOCS, 92 Barrett, 2014194</th>
<th>QUEST,97 Andersen, 200797</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Fair</td>
</tr>
<tr>
<td>Measures</td>
<td>Spielberger State/Trait Anxiety Inventory, General Health Questionnaire 12</td>
<td>QoL: SF-36 Mental and Physical Health scores Distress: Impact of Events Scale Cancer worry: Modified Lerman cancer worry scale</td>
</tr>
<tr>
<td>Psychological effect of screening</td>
<td>Random sample: no evidence of difference in state anxiety between screening and control groups</td>
<td>QoL, Distress, Cancer Worry: No statistically significant differences between study arms</td>
</tr>
<tr>
<td>Psychological effects of positive test results or repeat screening tests</td>
<td>Random sample: No evidence of change in anxiety or psychological morbidity due to repeat screenings compared with annual screen. Event sample: Compared with a single repeat screen: evidence of higher anxiety for multiple repeat scans (p&lt;0.010) (small absolute effect); greater odds of psychological morbidity (GHQ-12: score ≥4) with higher level referral screening: OR 1.28 (95% CI, 1.18 to 1.39) Women with abnormal test results (n=32) compared with women with no abnormal results more likely to report cancer worry at 2 year followup (OR 2.8; 95% CI, 1.1 to 7.2)</td>
<td></td>
</tr>
</tbody>
</table>

*All women in the screening arms recalled for repeat screening (excluding those in the random sample)

**Abbreviations:** CG = control group; CI = confidence interval; GHQ-12 = General Health Questionnaire 12; OR = odds ratio; QoL = quality of life; QUEST = Quality of Life, Education, and Screening Trial; SF-36 = 36-Item Short Form Survey; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening
<table>
<thead>
<tr>
<th>Test</th>
<th># studies (k), sample size (n) Design</th>
<th>Summary of Findings by Outcome</th>
<th>Consistency/Precision</th>
<th>Reporting Bias</th>
<th>Quality</th>
<th>Body of Evidence Limitations</th>
<th>EPC Assessment of Strength of Evidence for the KQ</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KQ1: Effectiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual screening with CA-125 testing</td>
<td>k=2, n=173,858 RCT</td>
<td>Ovarian cancer mortality (k=2, n=173,858). Screening with CA-125 did not result in improved ovarian cancer mortality compared with no screening (UKCTOCS\textsuperscript{92} HR 0.89, 95% CI, 0.74 to 1.08, U.K. Pilot\textsuperscript{98} RR=0.50, 95% CI, 0.22 to 1.11)</td>
<td>Consistency reasonably consistent</td>
<td>Reasonably precise</td>
<td>Undetected</td>
<td>Good</td>
<td>Followup not available beyond 10 years for a substantial proportion of UKCTOCS trial participants. Reporting bias undetected.</td>
<td>Moderate</td>
</tr>
<tr>
<td>Annual screening with transvaginal ultrasound examination</td>
<td>k=1, n=151,922 RCT</td>
<td>Ovarian cancer mortality (k=1, n=151,922). TVU screening did not result in improved ovarian cancer mortality compared with usual care (UKCTOCS\textsuperscript{92} HR 0.91, 95% CI, 0.76 to 1.09)</td>
<td>Consistency NA</td>
<td>Reasonably precise</td>
<td>Undetected</td>
<td>Good</td>
<td>Followup data incomplete beyond 10 years for a substantial proportion of trial participants.</td>
<td>Moderate</td>
</tr>
<tr>
<td>Annual screening with CA-125 testing and transvaginal ultrasound examination</td>
<td>k=1, n=68,557 RCT</td>
<td>Ovarian cancer mortality (k=1, n=68,557). No reduction was found in ovarian cancer mortality from combined TVU and CA-125 screening compared with usual care (PLCO\textsuperscript{91} RR 1.18, 95% CI, 0.82 to 1.71).</td>
<td>Consistency NA</td>
<td>Reasonably precise</td>
<td>Undetected</td>
<td>Good</td>
<td>Changes to protocol, ovarian palpation dropped after first 4 trial years</td>
<td>Moderate</td>
</tr>
<tr>
<td>Test</td>
<td># studies (k), sample size (n)</td>
<td>Design</td>
<td>Summary of Findings by Outcome</td>
<td>Consistency/ Precision</td>
<td>Reporting Bias</td>
<td>Quality</td>
<td>Body of Evidence Limitations</td>
<td>EPC Assessment of Strength of Evidence for the KQ</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------</td>
<td>--------</td>
<td>--------------------------------</td>
<td>------------------------</td>
<td>-----------------</td>
<td>---------</td>
<td>-----------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>KQ2: Harms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual screening with CA-125 testing</td>
<td>k=3, n=242,415</td>
<td>RCT</td>
<td>False positive rate from screening (k=2, n=173,858). False positive rate over multiple rounds of screening in the largest trial was 44%. Complications from screening (k=2, n=220,480). Complications from CA-125 testing were generally minor and ranged from 0.86 per 10,000 screens to 58.3 per 10,000 women. False positive surgery (k=2, n=173,858). False positive surgeries occurred in 0.2% to 1% of those screened with CA-125.</td>
<td>Reasonably consistent or NA</td>
<td>Undetected</td>
<td>Good</td>
<td>Psychological harms measured only for subsets of trial participants</td>
<td>Moderate (Low for psychological harms)</td>
</tr>
<tr>
<td>Psychological effects of screening (k=1, n=13,413). Psychological harms reported in a subset of 1 trial. No statistically significant differences were found in psychological outcomes between the screening and no screening arms; increased psychological morbidity risk among women recalled for higher level screening.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 8. Summary of Evidence

<table>
<thead>
<tr>
<th>Test</th>
<th># studies (k), sample size (n) Design</th>
<th>Summary of Findings by Outcome</th>
<th>Consistency/ Precision</th>
<th>Reporting Bias</th>
<th>Quality</th>
<th>Body of Evidence Limitations</th>
<th>EPC Assessment of Strength of Evidence for the KQ</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual screening with transvaginal ultrasound examination</td>
<td>k=2 n =220,479 RCT</td>
<td>False positive rate and complications from screening (k=1, n=151,922). A false positive rate of 11.9% was reported in initial screening round. Complications from screening (k=2, n=220,479). Complications from screening with TVU ranged from 1.86 per 10,000 screens to 3.3 per 10,000 women. False positive surgery (k=1, n=151,922). False positive surgeries occurred in 3.2% of those screened with TVU. Complications from false positive surgery (k=1, n=151,922). Complications occurred in 3.5% of false positive surgeries. Psychological effects of screening (k=1, n=10,716). Psychological harms were reported in a subset of 1 trial. No statistically significant differences were found in psychological outcomes between the screening and no screening arms.</td>
<td>Reasonably consistent or NA Reasonably precise</td>
<td>Undetected</td>
<td>Good</td>
<td>Psychological harms measured only for subsets of trial participants Data on cumulative false positive rate not reported</td>
<td>Moderate (Low for psychological harms)</td>
<td>Screening conducted in specialized trial centers Treatment for cancer (in all study arms) was through the centralized National Health Service system in U.K. and in community care settings in U.S.</td>
</tr>
</tbody>
</table>
Table 8. Summary of Evidence

<table>
<thead>
<tr>
<th>Test</th>
<th># studies (k), sample size (n)</th>
<th>Design</th>
<th>Summary of Findings by Outcome</th>
<th>Consistency/ Precision</th>
<th>Reporting Bias</th>
<th>Quality</th>
<th>Body of Evidence Limitations</th>
<th>EPC Assessment of Strength of Evidence for the KQ</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual screening with CA-125 testing and transvaginal ultrasound examination</td>
<td>k=2, n=69,106 RCT</td>
<td>RCT</td>
<td>False positive rate and complications from screening (k=1, n=68,557). False positive screening rate of 9.8% over the course of the screening program.</td>
<td>Consistency NA</td>
<td>Undetected</td>
<td>Fair to Good</td>
<td>Psychological harms measured only for subsets of trial participants</td>
<td>Moderate (Low for psychological harms)</td>
<td>U.S. based, multisite trial Pragmatic trial with usual care control condition and referral to community clinicians for screen positives Majority white, non-Hispanic participants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Complications from screening (see complication rates for individual components).</td>
<td>Reasonably precise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>False positive rate for screen positive surgery (k=1, n=68,557). False positive surgeries occurred in 3.17% of those screened.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Complications from false positive surgery (k=1, n=68,557). Complications occurred in 15.09% of false positive surgeries.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Psychological effects of screening (k=1, n=549). Women with abnormal test results (n=32) compared with women with no abnormal results more likely to report cancer worry at 2 year followup (OR 2.8; 95% CI, 1.1 to 7.2).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** CA-125 = cancer antigen 125; CI = confidence interval; EPC = evidence-based practice center; FDA = U.S. Food and Drug Administration; HR = hazard ratio; KQ = key question; NA = not applicable; OR = odds ratio; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; QUEST = Quality of life, Education, and Screening Trial; RCT = randomized controlled trial; RR = rate ratio; TVU = transvaginal ultrasound; UKCTOCS = U.K. Collaborative Trial of Ovarian Cancer Screening; U.S. = United States
Appendix A. Detailed Methods

Literature Search Strategies for Primary Literature

Key:
/ = MeSH subject heading
$ = truncation
* = truncation
? = wildcard
ab = word in abstract
adj# = adjacent within x number of words
ae = adverse effects
hw = subject heading word
id = identifier
kw = keyword
md = methodology
near/# = adjacent within x number of words
ti = word in title

Cochrane Central Register of Controlled Trials (CENTRAL)
#1 (ovar* or (fallopian next tub*) or adenx*):ti,ab,kw near/4 (cancer* or neoplas* or tumo* or malignant* or carcinoma* or adenocarcinoma* or mass*):ti,ab,kw
#2 screen*:ti,ab,kw
#3 detect*:ti
#4 (sonog* or ultraso*):ti,ab,kw
#5 (tumo* next marker*):ti,ab,kw
#6 (serum next cancer next antigen*):ti,ab,kw
#7 "CA 125":ti,ab,kw
#8 101-#7
#9 #1 and #8 Publication Year from 2003 to 2016, in Trials

MEDLINE
1 Ovarian Neoplasms/
2 Fallopian Tube Neoplasms/
3 ((ovar$ or fallopian tub$ or adenx$) adj4 (cancer$ or neoplas$ or tumo$ or malignant$ or carcinoma$ or adenocarcinoma$ or mass$)).ti,ab.
4 or/1-3
5 Mass screening/
6 "Early detection of cancer"/
7 (screen$:ti adj5 (ovar$ or fallopian tub$ or adenx$)).ti,ab.
8 detect$:ti.
9 Ultrasonography/
10 (sonog$ or ultraso$).ti,ab.
11 Tumor Markers, Biological/
12 tumo?r marker$.ti,ab.
13 serum cancer antigen$.ti,ab.
14 CA 125.ti,ab.
15 algorithm$.ti,ab.
16 ROCA.ti,ab.
17 or/5-16
18 4 and 17
19 Ovarian Neoplasms/us [Ultrasonography]
20 Fallopian Tube Neoplasms/us [Ultrasonography]
21 18 or 19 or 20
22 clinical trials as topic/ or controlled clinical trials as topic/ or randomized controlled trials as topic/ or meta-analysis as topic/
23 (clinical trial or controlled clinical trial or meta analysis or randomized controlled
Appendix A. Detailed Methods

trial).pt.
24     Random$.ti,ab.
25     control groups/ or double-blind method/ or single-blind method/
26     clinical trial$.ti,ab.
27     controlled trial$.ti,ab.
28     meta analy$.ti,ab.
29     or/22-28
30     21 and 29
31     Animals/ not (Humans/ and Animals/)
32     30 not 31
33     limit 32 to (english language and yr="2003 -Current")
34     remove duplicates from 33

PUBMED, publisher-supplied records
#14   Search (((#13) AND publisher[sb]) AND (^"2003/01/01"[Date - Publication] : ^"3000"[Date - Publication])) AND English[Language]
#13   Search #8 AND #12
#12   Search #9 OR #10 OR #11
#11   Search (control[tiab] OR controls[tiab] OR controlled[tiab] OR controled[tiab]) AND (trial[tiab] OR trials[tiab])
#10   Search "clinical trial"[tiab] OR "clinical trials"[tiab] OR random*[tiab]
#9     Search systematic review[sb] OR metaanaly*[tiab] OR meta analysis[tiab]
#8     Search #1 AND #7
#7     Search #2 OR #3 OR #4 OR #5 OR #6
#6     Search CA 125[tiab]
#5     Search serum cancer antigen*[tiab]
#4     Search tumor* marker*[tiab]
#3     Search sonog*[tiab] or ultraso*[tiab]
#2     Search screen*[tiab] OR detect*[tiab]
#1     Search (ovar*[tiab] or fallopian tub*[tiab] or adenx*[tiab]) AND (cancer*[tiab] or neoplas*[tiab] or tumor*[tiab] OR tumour*[tiab] or malignan*[tiab] or carcinoma*[tiab] or adenocarcinoma*[tiab] or mass*[tiab])
## Appendix A Table 1. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Included</th>
<th>Excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>Screening for ovarian cancer in a primary care setting (alone or as part of a clinical examination)</td>
<td>Screening for ovarian cancer in selected high-risk populations, such as women who are BRCA mutation carriers or patients of a specialty practice, such as oncology</td>
</tr>
<tr>
<td>Populations</td>
<td>Asymptomatic, average risk women, ages 45 years and older</td>
<td>Trials enrolling only women who are selected based on an increased risk for ovarian cancer (e.g. known predisposing genetic syndromes, strong family history)</td>
</tr>
<tr>
<td>Screening tests</td>
<td>Screening tests and approaches evaluated in clinical trials such as, but not limited to: testing for serum cancer antigen (CA–125), transvaginal ultrasonography, and combined screening approaches or algorithms</td>
<td>Screening tests not evaluated in clinical trials</td>
</tr>
<tr>
<td>Comparisons</td>
<td>Comparison of screening with usual care or no screening; comparison of different included screening methods or programs</td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>KQ 1: Ovarian cancer–specific mortality (including primary peritoneal and fallopian tube cancer), all-cause mortality, cancer-related morbidity, and quality of life. KQ 2: Surgery rate, rates of false-positive screening results, complications of diagnostic surgical procedures, and health and psychological effects of screening tests</td>
<td></td>
</tr>
<tr>
<td>Settings</td>
<td>Primary care settings, including obstetrics/gynecology practices</td>
<td>Specialty practice settings, such as oncology</td>
</tr>
<tr>
<td>Study designs</td>
<td>Randomized, controlled trials</td>
<td>Cohort studies, case-controls, case reports, case series, and decision analyses</td>
</tr>
<tr>
<td>Study quality</td>
<td>Good and fair quality according to USPSTF criteria and supplemented quality measures</td>
<td>Poor quality according to USPSTF criteria and supplemental quality measures</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
<td>Non–English language studies</td>
</tr>
</tbody>
</table>

**Abbreviations:** KQ = key question; USPSTF = United States Preventive Services Task Force
### Appendix A Table 2. Quality Assessment Criteria of Randomized Controlled Trials

<table>
<thead>
<tr>
<th>USPSTF quality rating criteria&lt;sup&gt;120&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Initial assembly of comparable groups employs adequate randomization, including first concealment and whether potential confounders were distributed equally among groups</td>
</tr>
<tr>
<td>• Maintenance of comparable groups (includes attrition, crossovers, adherence, contamination)</td>
</tr>
<tr>
<td>• Important differential loss to followup or overall high loss to followup</td>
</tr>
<tr>
<td>• Measurements: equal, reliable, and valid (includes masking of outcome assessment)</td>
</tr>
<tr>
<td>• Clear definition of the interventions</td>
</tr>
<tr>
<td>• All important outcomes considered</td>
</tr>
<tr>
<td>• Intention-to-treat analysis</td>
</tr>
</tbody>
</table>

**Abbreviations:** KQ = key question; USPSTF = United States Preventive Services Task Force
Appendix B. Included Studies

Below is a list of included studies and their ancillary publications (indented below main results publication):


Appendix B. Included Studies


Appendix C. Excluded Studies

<table>
<thead>
<tr>
<th>Reason for Exclusion*</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1. Study aim</td>
</tr>
<tr>
<td>E2. Study design (not RCT)</td>
</tr>
<tr>
<td>E3a. Setting: Not conducted in a 'very high' HDI country</td>
</tr>
<tr>
<td>E3b. Setting: Not primary care</td>
</tr>
<tr>
<td>E4a. Population: High-risk (genetic or family history)</td>
</tr>
<tr>
<td>E4b. Population: Symptomatic</td>
</tr>
<tr>
<td>E5. No relevant outcomes or incomplete outcomes</td>
</tr>
<tr>
<td>E5a. Detection rate only</td>
</tr>
<tr>
<td>E6. Study quality: Poor quality rating</td>
</tr>
<tr>
<td>E7. Language: Publication not in English</td>
</tr>
<tr>
<td>E8. Intervention is not a screening test (i.e., staging, prognostic testing)</td>
</tr>
</tbody>
</table>

* Assigned at abstract and full-text phase

**Abbreviations:** E = exclude; HDI = human development index; KQ = key question; RCT = randomized, controlled trial

Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies

## Appendix D. Benefits Reported in Ovarian Cancer Screening Trials: Ovarian Cancer Mortality in Analyses Excluding Peritoneal Cancer*

<table>
<thead>
<tr>
<th>Trial, Year of publication</th>
<th>UKCTOCS, 2016 (CA-125 ROCA arm)⁵²</th>
<th>UKCTOCS, 2016 (ultrasound arm)⁵²</th>
<th>PLCO, 2011⁹¹</th>
<th>U.K. Pilot, 1999⁸⁸</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>N analyzed per arm</td>
<td>IG: 50,624</td>
<td>IG: 50,623</td>
<td>IG: 34,253</td>
<td>IG: 10,958</td>
</tr>
<tr>
<td>Ovarian cancer incidence, n (%)</td>
<td>IG: 338 (0.7)</td>
<td>IG: 314 (0.6)</td>
<td>IG: 183 (0.5)</td>
<td>IG: 16 (0.1)</td>
</tr>
<tr>
<td></td>
<td>CG: 630 (0.6)</td>
<td>CG: 630 (0.6)</td>
<td>CG: 158 (0.5)</td>
<td>CG: 20 (0.2)</td>
</tr>
<tr>
<td>Ovarian cancer incidence rate</td>
<td>IG: 6.2 per 10,000 p-y</td>
<td>IG: 5.7 per 10,000 p-y</td>
<td>IG: 4.9 per 10,000 p-y</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>CG: 5.7 per 10,000 p-y</td>
<td>CG: 5.7 per 10,000 p-y</td>
<td>CG: 4.7 per 10,000 p-y</td>
<td>NR</td>
</tr>
<tr>
<td>Between group difference</td>
<td>p=NR</td>
<td>p=NR</td>
<td>p=NR</td>
<td>NR</td>
</tr>
<tr>
<td>Ovarian cancer mortality, n (%)</td>
<td>IG: 148 (0.29)</td>
<td>IG: 163 (0.30)</td>
<td>NR</td>
<td>IG: 9 (0.082)</td>
</tr>
<tr>
<td></td>
<td>CG: 347 (0.34)</td>
<td>CG: 347 (0.34)</td>
<td></td>
<td>CG: 18 (0.16)</td>
</tr>
<tr>
<td>Ovarian cancer mortality rate</td>
<td>IG: 2.7 per 10,000 p-y</td>
<td>IG: 2.8 per 10,000 p-y</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>CG: 3.2 per 10,000 p-y</td>
<td>CG: 3.2 per 10,000 p-y</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Ovarian cancer mortality between group difference</td>
<td>HR: 0.85 (95% CI, 0.70 to 1.03), p=0.10†</td>
<td>HR: 0.89 (0.73, 1.07), p=0.21†</td>
<td>NR</td>
<td>RR: 0.50 (95% CI, 0.22 to 1.11) p=0.083‡</td>
</tr>
</tbody>
</table>

* Includes ovarian and fallopian cancers
† Cox proportional hazards model
‡ Calculated (article reports RR calculated in terms of increased relative risk)

**Abbreviations:** CG = control group; CI = confidence interval; HR = hazard ratio; IG = intervention group; NR = not reported; NS = not significant; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; P-Y = person-years; ROCA = Risk of Ovarian Cancer Algorithm; RR = relative risk; U.K. = United Kingdom; UKCTOCS: U.K. Collaborative Trial of Ovarian Cancer Screening