Evidence Synthesis
Number 147

Screening for Gynecologic Conditions With Pelvic Examination: A Systematic 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-290-2012-00015-I-EPC4, Task Order No. 4

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

Investigators:
Janelle M. Guirguis-Blake, MD
Jillian T. Henderson, PhD, MPH
Leslie A. Perdue, MPH
Evelyn P. Whitlock, MD, MPH

AHRQ Publication No. 15-05220-EF-1
March 2017
This report is based on research conducted by the Kaiser Permanente Research Affiliates Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HHSA-290-2012-00015-I-EPC4, Task Order No. 4). The findings and conclusions in this document are those of the authors, who are responsible for its contents; and do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Smyth Lai, MLS, for creating and conducting the literature searches; Kevin Lutz, MFA, for editorial assistance; Nadia Redmond, MSPH, for assistance with data abstraction; Jennifer S. Lin, MD, for review of the final report; Tina Fan, MD, MPH, at AHRQ; current and former members of the U.S. Preventive Services Task Force who contributed to topic deliberations; and Hanna Bloomfield, MD, MPH, Melinda A. Cavicchia, MD, MPH, David Chelmow, MD, Esther Eisenberg, MD, MPH, Tara Jatlaoui, MD, MPH, Barnett Kramer, MD, MPH, Giang Nguyen, MD, MPH, Mary Roary, PhD, Mona Saraiya, MD, MPH, Sherri Stewart, PhD, and Naomi Tepper, MD, MPH, for their expert feedback on this report. The authors thank Barnett Kramer, MD, MPH, Paul Pinsky, PhD, and Maryam Doroudi, PhD, for providing unpublished data from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial for participants undergoing ovarian palpation.

Suggested Citation

Structured Abstract

**Background:** In light of recent guideline changes in the periodicity of cervical cancer screening, there is uncertainty surrounding the role of the routine screening pelvic examination during annual prevention visits.

**Purpose:** We conducted this systematic review to support the U.S. Preventive Services Task Force in creating its recommendation on the periodic screening pelvic examination. Our review addresses three questions: 1) What is the direct evidence for the effectiveness of the pelvic examination in reducing all-cause mortality, cancer- and disease-specific morbidity and mortality, and improving quality of life? 2) What are the test performance characteristics of the pelvic examination in screening for gynecologic cancers and other gynecologic conditions? 3) What are the adverse effects of screening using the pelvic examination?

**Data Sources:** We searched MEDLINE, PubMed (publisher-supplied references only), and the Cochrane Central Register of Controlled Trials to identify literature that was published from the earliest date indexed (1946 for MEDLINE) to January 13, 2016. We supplemented our searches with reference lists from relevant existing systematic reviews, suggestions from experts, and ClinicalTrials.gov to identify ongoing trials.

**Study Selection:** Two investigators independently reviewed identified abstracts and full-text articles against a set of a priori inclusion and quality criteria.

**Data Analysis:** One investigator abstracted data into an evidence table and a second investigator confirmed these data. We qualitatively synthesized the data for each key question; quantitative synthesis was not appropriate due to heterogeneity and the low number of trials for any given intervention and outcome.

**Results:** We found no studies that assessed the effectiveness of the pelvic examination in reducing all-cause mortality, reducing cancer- and disease-specific morbidity and mortality, or improving quality of life. For four conditions, we identified a total of eight diagnostic accuracy studies that examined test characteristics for the screening pelvic examination: ovarian cancer (k=4), bacterial vaginosis (k=2), trichomoniasis (k=1), and genital herpes (k=1). These eight studies also provided information on the harms of screening using the pelvic examination (false-positive and false-negative results). One large good-quality randomized, controlled trial reported additional diagnostic workup, surgeries, and any complications occurring 1 year after abnormal ovarian palpation. An additional cohort study also assessed harms (urinary symptoms). The low prevalence of ovarian cancer in the general population consistently resulted in low positive predictive values. In these four ovarian cancer screening studies, with more than 26,000 screened patients, more than 98 percent of the positive test results were false positives, depending on the study design and management protocols. Surgery rates resulting from an abnormal pelvic examination ranged from 5 to 36 percent at 1 year, with the largest study reporting an 11 percent surgery rate and 1 percent complication rate within 1 year of an abnormal screening pelvic examination. Each diagnostic accuracy study for bacterial vaginosis, trichomoniasis, and genital herpes was performed in a high-prevalence population with high proportions of symptomatic patients and reported accuracy characteristics for individual physical examination findings.
thereby limiting any conclusions that could be made regarding the screening accuracy or adverse effects of the pelvic examination in asymptomatic primary care populations.

**Conclusions:** There is no direct evidence on the overall benefits and harms of the pelvic examination as a one-time or periodic screening test. In addition, there is limited evidence regarding the diagnostic accuracy and harms of the routine screening pelvic examination to guide practice in asymptomatic primary care populations. Research is needed to illuminate how recent changes in cervical cancer screening periodicity may influence women’s access to other evidence-based preventive services in the primary care setting and to create best practices for achieving high rates of uptake for these recommended services.
Table of Contents

Chapter 1. Introduction ................................................................. 1
  Background .............................................................................. 1
  Current Clinical Practice in the United States ......................... 2
  Recommendations of Other Groups ........................................ 3
  Previous Related USPSTF Topics ........................................... 4

Chapter 2. Methods ..................................................................... 5
  Scope and Purpose ................................................................. 5
  Key Questions and Analytic Framework .................................. 5
    Key Questions ....................................................................... 5
  Data Sources and Searches ..................................................... 6
  Study Selection ....................................................................... 6
  Quality Assessment and Data Abstraction .............................. 6
  Data Synthesis and Analysis .................................................. 7
    Key Question 1 ................................................................. 7
    Key Questions 2 and 3 ....................................................... 7
    Key Question 3 ................................................................. 8
  Expert Review and Public Comment ....................................... 8
  USPSTF Involvement .............................................................. 8

Chapter 3. Results ..................................................................... 9
  Key Question 1. What Is the Direct Evidence for the Effectiveness of the Pelvic Examination in a) Reducing All-Cause Mortality, b) Reducing Cancer and Disease-Specific Morbidity and Mortality, and c) Improving Quality of Life? ....................................... 9
  Key Questions 2 and 3. What Are the Test Performance Characteristics of the Pelvic Examination in Screening for Gynecologic Cancers and Other Gynecologic Conditions? What Are the Adverse Effects of Screening Using the Pelvic Examination? ................................................................. 9
    Ovarian Cancer ................................................................. 9
    BV .................................................................................. 12
    Genital Herpes (Herpes Simplex Virus-1 or -2) ..................... 13
    Trichomoniasis ................................................................. 15
    Other Harms ..................................................................... 16

Chapter 4. Discussion ............................................................... 17
  Screening Accuracy ............................................................... 17
    Ovarian Cancer ............................................................... 17
    Infectious Diseases ........................................................... 18
  Harms .................................................................................. 18
  Clinical Implications ........................................................... 19
  Limitations .......................................................................... 19
  Research Gaps ..................................................................... 20
  Conclusions ......................................................................... 21

References ............................................................................... 22
Figures
Figure 1. Analytic Framework

Tables
Table 1. Epidemiology of Gynecologic Cancers and Conditions
Table 2. Recommendations on the Periodic Pelvic Examination for Asymptomatic Adult Women
Table 3. U.S. National Guidelines and Statements on Screening for Individual Gynecologic Conditions in Unselected Adult Women Who Are Asymptomatic and Not Pregnant
Table 4. USPSTF Screening Recommendations for Malignant and Benign Gynecologic Conditions
Table 5. Study Characteristics, Ovarian Cancer Screening
Table 6. Summary of Diagnostic Accuracy of Pelvic Examination for Ovarian Cancer
Table 7. Diagnostic Procedures Within 1 Year of a Positive Palpation Examination in Women Without an Ovarian Cancer Diagnosis: PLCO Trial Only
Table 8. Study Characteristics, Infectious Diseases
Table 9. Summary of Diagnostic Accuracy of Pelvic Examination for Infectious Disease
Table 10. Summary of Evidence by Key Question and Condition

Appendixes
Appendix A. Detailed Methods
Appendix B. Excluded Studies
Chapter 1. Introduction

In recent years, evidence reviews and expert consensus have been the basis for changes to longstanding practices of preventive gynecologic screening. There have been several monumental changes in women’s health in the past decade, including a shift away from annual cervical cancer screening and a move toward urine-based screening for sexually transmitted infections (STIs) (specifically, chlamydia and gonorrhea) among young women. Later initiation of cervical cancer screening and longer intervals between Papanicolaou (Pap) tests were recommended by the American College of Obstetricians and Gynecologists (ACOG) in 2009. Current guidelines from the U.S. Preventive Services Task Force (USPSTF) and ACOG recommend screening for cervical cancer beginning at age 21 years and every 3 years thereafter until age 30 years; after age 30 years, 5-year intervals are recommended for most women not at high risk of this disease. Prior to these changes, annual visits for cervical cancer screening provided an opportunity for routine examination of the external and internal reproductive organs. As the intervals for cervical cancer screening have been extended, the independent clinical value of the pelvic examination has been increasingly questioned and debated and a variety of recommendations have been issued. Currently, a central question in women’s primary health care is whether women attending routine visits without gynecologic symptoms would benefit from a screening pelvic examination.

Background

The pelvic examination consists of visual and physical assessments of female reproductive organs. The pelvic examination may be performed for the purpose of screening for a specific condition, diagnostic evaluation of gynecological symptoms, or disease surveillance. Typically, the screening pelvic examination for asymptomatic women includes a visual inspection of the external genitalia; a speculum examination of the vagina and cervix; bimanual examination of the adnexa, uterus, and cervix; and may include a rectovaginal examination. In addition, tests for cervical cancer screening (i.e., Pap test, human papillomavirus [HPV] test) may be collected during a routine pelvic examination. Historically, screening pelvic examinations were part of routine annual gynecological examinations during which Pap test collections occur. Even after most professional societies endorsed less frequent cervical cancer screening, many women continued to present annually for routine gynecologic care. Routine pelvic examination is a longstanding practice that some patients and providers may view as an opportunity to discuss a broad range of sexual and reproductive health issues.

In contrast to most screening tests, the pelvic examination does not identify a unique disease entity. Multiple gynecologic conditions (malignant and benign) could plausibly be detected by pelvic examination or are cited by providers as reasons for conducting a pelvic examination, including: cervical, endometrial, ovarian, vaginal, and vulvar cancer; bacterial vaginosis (BV), candidiasis, chlamydia, gonorrhea, genital warts, genital herpes, pelvic inflammatory disease (PID), and trichomoniasis; and atrophic vaginitis, cervical polyps, endometriosis, ovarian cysts, pelvic organ prolapse, uterine fibroids, and vulvar lichen sclerosis (Table 1). Each disease can be considered individually for evidence-based screening recommendations by weighing the
potential benefits and risks based on test-, disease-, and population-specific factors. Specifically, each condition can be evaluated for prevalence and burden, typical clinical presentation, screening test accuracy during the asymptomatic phase, and treatment benefits in early-stage disease. Moreover, the gynecologic conditions potentially detectable with the pelvic examination vary by target population—some conditions occur only in specific age groups (adolescent, young adult, pregnant, premenopausal, or postmenopausal) or primarily among women at increased risk based on behavioral or genetic factors.

The value of early detection of asymptomatic disease for these gynecologic conditions varies considerably. For example, identifying and treating screen-detected asymptomatic BV or vaginal candidiasis in nonpregnant women may have little clinical benefit compared to diagnosis and treatment during symptomatic stages. Likewise, for asymptomatic atrophic vaginitis, cervical polyps, or uterine fibroids, the clinical significance, and therefore the role for early treatment, is unclear in the absence of symptoms. For some gynecologic conditions, such as cervical cancer, gonorrhea, and chlamydia, there are alternative and well-established evidence-based screening tests with superior accuracy compared to the pelvic examination (i.e., Pap/HPV test for cervical cancer and nucleic acid amplification tests [NAATs] for gonorrhea and chlamydia). In theory, some gynecologic cancers, such as those of the ovaries, vulva, and vagina, might have an improved treatment prognosis if detected in earlier, asymptomatic stages and there is currently no alternative effective screening strategy. In contrast, endometrial cancer is frequently symptomatic in its early stages, and the screening pelvic examination is unlikely to detect early-stage cancer since it is not palpable or visible on examination. Some conditions, like endometriosis, clinically present in the context of dysmenorrhea or infertility diagnostic workup. Likewise, pelvic floor dysfunction or pelvic organ prolapse may be diagnosed and graded for severity only after history taking reveals urinary incontinence or retention. Furthermore, advances in ultrasound technology over the past few decades, which is more sensitive than pelvic examination for detecting pelvic masses, have transformed the detection and surveillance of pelvic masses.

**Current Clinical Practice in the United States**

In 2012, 44.2 million pelvic examinations were performed in outpatient visits in the United States. Sixty-eight percent of surveyed U.S. obstetrician-gynecologists routinely perform a pelvic examination, and 78 percent of all surveyed physicians (including family/general practitioners and internists) believed that pelvic examination is a useful screening test for gynecologic cancers. In a nationally representative survey of obstetrician-gynecologists, approximately 50 percent reported that performing a bimanual examination to detect ovarian cancer was very important, and approximately 20 and 25 percent thought it was very important for the detection of uterine and cervical cancer, respectively. For other gynecological conditions, more than 50 percent thought the bimanual examination was very important for detecting benign ovarian conditions, nearly 60 percent thought it was very important to detect benign uterine conditions, and about 30 percent reported it was very important in detecting both subclinical PID and uterine position. Almost all of the surveyed physicians indicated that they would perform a bimanual examination during a routine visit with an asymptomatic patient.
The Centers for Medicare & Medicaid Services cover screening pelvic examinations as a standalone billable service, without patient copayment, every 24 months for all asymptomatic women and yearly for high-risk women (i.e., women at high risk for vaginal or cervical cancer or women of childbearing age with a vaginal or cervical abnormality found on pelvic examination in the preceding 3 years).18

Unpublished data from 2008 through 2010 indicate that the majority of preventive care visits to obstetrician-gynecologists (76%) included a pelvic examination. In contrast, only a quarter of visits to family medicine physicians (25%) and even fewer to internal medicine physicians (14%) included a pelvic examination (E. Hing, personal communication, January 22, 2015).

**Recommendations of Other Groups**

Professional organizations vary in their recommendations regarding routine screening pelvic examinations (Table 2). The American College of Physicians recently released a guideline recommending that practitioners not perform screening pelvic examinations (except for cervical cancer screening by visual inspection of the cervix and cervical swabs) in asymptomatic, average-risk women for the purpose of screening for gynecologic cancers, PID, and other benign gynecologic conditions.9 The basis of this recommendation was a systematic review that identified no benefits of pelvic examination but some exposure to unnecessary and avoidable harms. The recommendation does not apply to women who present with symptoms (e.g., abnormal bleeding, pain), in which case the pelvic examination would be an appropriate diagnostic procedure to consider. The guideline is endorsed by the American Academy of Family Physicians.19

ACOG recommends an annual pelvic examination for women age 21 years or older but acknowledges there is no evidence in support of or against this recommendation.8 Furthermore, it notes that this examination is not necessary to prescribe hormonal contraception in healthy women or to screen for STIs. For females younger than age 21 years, ACOG recommends a pelvic examination if indicated by medical history. For symptomatic patients age 21 years or older, joint decisionmaking with the clinician and patient is advised to determine whether pelvic examination should be performed. In 2015, ACOG convened the Well-Woman Task Force (WWTF) and released recommendations for the well-woman visit.20 The WWTF recommended that external examinations may be performed annually in healthy patients age 21 years or older, but the inclusion of speculum and bimanual examination for asymptomatic women without specific indications (e.g., cervical cancer screening) should be a shared, informed decision between the patient and provider. The WWTF categorized its pelvic examination recommendation as “qualified,” meaning that it is based on expert opinion rather than clinical evidence.

Despite the inconsistent recommendations for screening pelvic examinations, the available guidelines from national organizations regarding screening for individual gynecologic conditions are similar (Table 3). Recommendations on the types and timing of tests that are effective for cervical cancer, gonorrhea, or chlamydia screening are consistent, as are recommendations against screening for ovarian cancer using currently available approaches (including the
bimanual pelvic examination). Recommendations for other gynecologic conditions are few.

**Previous Related USPSTF Topics**

The USPSTF has not made a prior recommendation regarding the routine use of screening pelvic examinations in unselected asymptomatic women, but it has issued several recommendations for screening for benign and malignant gynecologic conditions, including ovarian cancer, cervical cancer, herpes, chlamydia, gonorrhea, and BV (Table 4).
Chapter 2. Methods

Scope and Purpose

The USPSTF will use this evidence review to issue a new recommendation statement on the use of periodic screening pelvic examinations in unselected, asymptomatic, nonpregnant adult women. This review addresses the benefits and harms of screening with the pelvic examination for gynecologic cancers and conditions as well as the diagnostic accuracy of the pelvic examination in detecting individual gynecologic cancers and conditions.

While the pelvic examination is common for adolescent and pregnant women, these populations were specifically excluded from the scope of this review. The purpose of conducting the pelvic examination in unselected nonpregnant adult women may differ from that in special populations of adolescents (e.g., Tanner staging, congenital abnormality case-finding) or pregnant women (e.g., pregnancy dating, pelvic outlet adequacy, cervical dilation checks). Likewise, screening for congenital gynecological conditions was excluded because this review focuses on routine periodic screening, and many congenital conditions would be detected at the symptomatic stage, during pregnancy, at infertility workup, or incidentally during cervical cancer screening.

Further, the USPSTF previously determined that there is good evidence for primary screening approaches for cervical cancer, gonorrhea, and chlamydia. Since the pelvic examination alone is less accurate than the existing screening approaches for these conditions (i.e., NAATs for gonorrhea and chlamydia, Pap/HPV cotesting for cervical cancer), they were not included in the scope of this review.

Key Questions and Analytic Framework

The analytic framework is presented in Figure 1.

Key Questions

1. What is the direct evidence for the effectiveness of the pelvic examination in a) reducing all-cause mortality, b) reducing cancer- and disease-specific morbidity and mortality, and c) improving quality of life?
2. What are the test performance characteristics of the pelvic examination (sensitivity, specificity, and positive and negative predictive values) in screening for gynecologic cancers and other gynecologic conditions?
3. What are the adverse effects of screening by pelvic examination?
Data Sources and Searches

We searched MEDLINE, PubMed (publisher-supplied references only), and the Cochrane Central Register of Controlled Trials to locate relevant studies for all three key questions. We searched for articles from the earliest date indexed (1946 for MEDLINE) through January 13, 2016. We supplemented our database searches with experts’ suggestions and by reviewing reference lists from all other recent existing systematic reviews. We also searched selected sources of grey literature, including ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform, for ongoing trials. The National Cancer Institute provided previously unpublished 1 and 5 year followup data from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial from the subset of women receiving bimanual ovarian palpation and rectovaginal examination; the 5 year results were subsequently published.21

Study Selection

Two investigators independently reviewed 8,678 titles and abstracts by using an online platform (abstrackr22) and 316 articles (Appendix A Figure 1) with specified inclusion criteria (Appendix A Table 1). We resolved discrepancies through consensus and consultation with a third investigator. We excluded articles that did not meet inclusion criteria or those we rated as poor quality. To ensure that studies using the pelvic examination as a secondary screening test (e.g., ovarian cancer screening studies using the tumor marker cancer antigen 125 [CA-125] and ultrasound technology that also included a pelvic examination component) were not missed, we were more inclusive during the review of abstracts and titles. As a result, many studies were excluded at the full-text review. Appendix B lists all excluded trials.

Eligible studies included unselected adult women who were not symptomatic or pregnant. We excluded studies that were conducted solely in symptomatic populations.

For the greatest applicability to U.S. practice, we focused on studies conducted in developed countries, as defined by “very high” development according to the 2014 United Nations Human Development Index.23 We included only studies that published their results in English because of resource constraints.

Any study that examined the effects of pelvic examination on all-cause mortality, cancer- or disease-specific morbidity or mortality, or quality of life was eligible for inclusion in our review. Further, studies examining the screening accuracy of the pelvic examination in a single encounter or as a periodic program of screening were also eligible.

Quality Assessment and Data Abstraction

At least two reviewers critically appraised all articles that met the inclusion criteria using the Newcastle Ottawa Scale for cohort and case-control studies24 and Quality Assessment of Diagnostic Accuracy Studies I and II for studies of diagnostic accuracy,25, 26 adapted to align
with the USPSTF’s design-specific quality criteria (Appendix A Table 2). We rated articles as good, fair, or poor quality. In general, a good-quality study met all criteria, indicating low risk of bias. A fair-quality study did not meet, or it was unclear if it met, at least one criterion and also had no known important limitations that could invalidate its results. A poor-quality study had a single fatal flaw or multiple important limitations. We excluded poor-quality studies from this review. Disagreements about critical appraisal were resolved by consensus and, if needed, consultation with a third independent reviewer.

One reviewer extracted key elements of included studies into standardized evidence tables. A second reviewer checked the data for accuracy. Evidence tables were tailored for each key question and to specific study designs and/or specific screening tests. Tables generally included details on study quality, setting and population (e.g., country, inclusion criteria, age, sex, race/ethnicity, symptomatic), screening test and protocol (e.g., who administered, how it was administered, definition of a positive test), reference standard or comparator (if applicable), length of followup, and outcomes (e.g., mortality, sensitivity and specificity, harms).

**Data Synthesis and Analysis**

We synthesized results by key question and type of screening test. We used a standardized summary of evidence table to describe the overall strength of evidence for each key question. This table included the number and design of included studies, summary of results, consistency or precision of results, reporting bias, summary of study quality, limitations of the body of evidence, and applicability of findings.

The results are organized by key question. Since seven of the eight included studies reported outcomes for both key questions 2 (accuracy) and 3 (harms), we present the results for both key questions together for each disease condition. An exception was a single additional study on harms in a section entitled “Other Harms.”

**Key Question 1**

There were no studies found for key question 1.

**Key Questions 2 and 3**

This combined question focused on the one-time test performance of a single pelvic examination for a single condition. We organized our synthesis by condition and discuss the harms of examination due to false-positive results or further diagnostic workup. We calculated sensitivity and specificity in Stata version 13.1 (StataCorp, College Station, TX) using Jeffrey’s confidence intervals. We used 2×2 tables constructed from data reported in the primary studies. If the observed sensitivity or specificity was 100 percent, only the lower 95 percent confidence interval was calculated. In many cases, the data presented in our report differ slightly from the published paper because of these calculations. Since there was a limited number of studies for each condition, no pooled analyses were conducted.
Key Question 3

Due to the limited number of studies that were included only for key question 3, these results are summarized qualitatively.

Expert Review and Public Comment

The draft research plan was posted for public comment on the USPSTF Web site from December 4, 2014, through January 7, 2015. After that feedback was reviewed, the rectovaginal examination was included as a component of the pelvic examination. No other substantive changes were made. The full draft report was also reviewed by invited content experts and USPSTF federal partners and posted for public comment on the USPSTF Web site from July 5 to July 25, 2016. Comments received during any period were reviewed, considered, and addressed, as appropriate. No new substantive issues were identified that were not previously considered and no major changes were made to the text in the final report.

USPSTF Involvement

The authors worked with four USPSTF liaisons throughout the review process to develop and refine the analytic framework. These liaisons also helped to develop the key questions and to resolve scope issues for the final evidence synthesis.

This research was funded by the Agency for Healthcare Research and Quality under a contract to support the work of the USPSTF. Agency staff provided oversight for the project, reviewed the draft report, and assisted with the federal partner review of the draft report.
Chapter 3. Results

Key Question 1. What Is the Direct Evidence for the Effectiveness of the Pelvic Examination in a) Reducing All-Cause Mortality, b) Reducing Cancer- and Disease-Specific Morbidity and Mortality, and c) Improving Quality of Life?

We found no studies that assessed the effectiveness of pelvic examination in reducing all-cause mortality, reducing cancer- and disease-specific morbidity and mortality, or improving quality of life.

Key Questions 2 and 3. What Are the Test Performance Characteristics of the Pelvic Examination in Screening for Gynecologic Cancers and Other Gynecologic Conditions? What Are the Adverse Effects of Screening Using the Pelvic Examination?

We found four studies examining the accuracy of the pelvic examination to detect ovarian cancer, two studies for BV, and one study each for trichomoniasis and genital herpes. All of these accuracy studies (k=8) were included for harms (false-positive rates and resulting diagnostic workup), and one additional study was included for other harms.

Ovarian Cancer

Summary of Findings

Despite limitations of the evidence, the low prevalence of ovarian cancer in the general population consistently resulted in low positive predictive values (PPVs) for the screening pelvic examination in detecting ovarian cancer. Based on a large study of more than 20,000 women, sensitivity is low (<5%) for the detection of ovarian cancer. Considering all four included screening studies, we could not estimate accuracy with precision due to rarity of the disease, few studies, and short followup time in most studies. In the evidence we reviewed, surgery due to abnormal pelvic examination results ranged from 5 to 36 percent of women, depending on the study design and management protocols.

Study Characteristics

We identified one good-quality and three fair-quality studies (n=26,432) that examined the screening accuracy of pelvic examination in identifying ovarian cancer (P. Pinsky, written communication, May 2, 2016). One large, multicenter U.S. randomized, controlled trial, the PLCO trial, recruited average-risk women ages 55 to 74 years from the community with an
overall aim of examining the benefits and harms of ovarian cancer screening using a combination of three modalities: blood testing for CA-125, transvaginal ultrasound, and ovarian palpation/rectovaginal examination (ovarian palpation was dropped 5 years after trial recruitment began because no ovarian cancers were detected solely based on an abnormal ovarian palpation examination). The other three studies, conducted in Greece, Australia, and the United Kingdom, primarily recruited average-risk women ages 40 to 45 years and older from the community, but none excluded women with a family history of ovarian cancer, and one actively recruited younger women with a family history of ovarian cancer (Table 5). One study recruited all or nearly all postmenopausal women based on a lower age limit of 55 years and one study specifically recruited postmenopausal women older than age 45 years. The PLCO trial excluded women with prior ovarian, lung, or colon cancer. Two studies excluded women with a history of ovarian cancer, any malignancy, or bilateral oophorectomy. The fourth study recruited apparently healthy women without exclusions, and any women with a past history of cancer had to be in remission to participate. None of the studies excluded women with a family history of ovarian cancer. The number of participants who received a pelvic examination ranged from 1,010 to 20,872 across the four studies.

One study defined the test as bimanual ovarian palpation plus rectovaginal examination. One study clearly defined the index pelvic examination test as the bimanual and speculum examination, while the other two studies did not define the index test beyond “pelvic exam” or “vaginal exam.” Two studies specified that experienced gynecologists or examiners performed the examination, another stated that a single examiner performed all examinations, and the third stated that one of two physicians examined all women. Only one study specified the ultrasonography operator as a gynecologist specializing in ultrasonography. In the PLCO trial, all participants received the bimanual ovarian palpation and rectovaginal examination in addition to blood testing for CA-125 and transvaginal ultrasound, with screening tests repeated annually. Women who received the ovarian palpation/rectovaginal examination had a mean of 2.4 examinations (with no more than 4 examinations) over the initial 5 years of the trial period when this examination was a part of the screening intervention protocol. In three of the studies, all women had an examination with CA-125 blood testing, and only those with abnormal pelvic examination findings or abnormal CA-125 results had transvaginal or abdominal ultrasonography. These three studies used different thresholds for acceptable CA-125 levels (>30 U/mL, >35 U/mL, and ≥35 U/mL). Likewise, these three studies used differing thresholds to define an abnormal reference ultrasonography result (18 mL/8 mL in pre- and postmenopausal women, respectively; >6 cm; and >8.8 mL). For the PLCO trial, cancer incidence followup at 1 year and up to 5 years was captured in medical records and patient questionnaires. For the other three studies, followup at 1 year consisted of a postal questionnaire for all patients or for those with normal pelvic examination findings and CA-125 levels; a fourth study additionally measured CA-125 levels at 1 year for patients with normal baseline pelvic examination and CA-125 results. Any patients with abnormal CA-125 levels and normal ultrasonography findings were followed with serial CA-125 blood testing and/or ultrasonography every 3 to 6 months. For all four studies, any abnormal results were referred for further management.

Mean or median age ranged from 51 to 63 years. Forty-three and 65 percent of participants were postmenopausal in the two studies reporting menopausal status. In the other two studies, all
or nearly all participants were postmenopausal.28,31

**Yield and Accuracy**

Ovarian cancer prevalence was reported as 0.10 percent in three studies28,29 (P. Pinsky, written communication, May 2, 2016) and 0.04 percent in one study30; the longer followup from the PLCO trial (up to 5 years) reported a 0.30 percent prevalence of ovarian cancer (P. Pinsky, written communication, May 2, 2016)21(Table 6). Focusing on comparable 1-year data from the four studies, the proportion of participants with positive pelvic examination results ranged from 1.2 to 8.7 percent. Sensitivity was reported as 100 percent in two of the studies28,29 (up to two ovarian cancer cases were palpable on pelvic examination) and 0 percent in the study where the single case of ovarian cancer was not detected on the pelvic examination.30 The PLCO trial reported a sensitivity of 4.3 and 2.8 percent from the first screening examination at 1 year and at up to 5 years followup, respectively. In this trial, over the multiple rounds of screening (mean number of screenings, 2.4 [range, 1 to 4]), 91 cases of cancer were detected within 5 years of a screening examination and 88 cases (96.7%) were not detected by the palpation examination (P. Pinsky, written communication, May 2, 2016). Specificity ranged from 91 to 99 percent in the four studies. Calculated PPV ranged from 0 to 3.6 percent, and negative predictive value (NPV) was 99 percent or greater for all studies. Accuracy estimates had wide confidence intervals due to the very low event rate.

**Harms**

Because more than 98 percent of the women with abnormal pelvic examination findings were false positives, additional imaging and unnecessary surgical intervention are potential harms of pelvic examination screening for ovarian cancer. The prevalence of laparoscopy or laparotomy for patients with abnormal findings on pelvic examination ranged from 5 to 36 percent. In the Greek study, 17 percent of the women with abnormal pelvic examination results underwent surgery due to the examination results. Pathology findings revealed two cases of ovarian cancer (one was metastatic and the other was a stage Ia serous cystadenocarcinoma), four serous cystadenomas, three mucinous cystadenomas, five endometroid cysts of the ovary, 12 benign cysts, and three normal pathology results.29 In the Australian study, two women (5%) with abnormal findings on pelvic examination had surgery. The surgeries revealed that one patient had a fibroid uterus and one patient had a normal (negative) result; the single case of ovarian cancer was not detected on pelvic examination.30 In the U.K. study that recruited solely postmenopausal women, 36 percent of women with an abnormal pelvic examination result underwent surgery due to the examination results: one woman had ovarian cancer and nine women had benign conditions (six had benign ovarian cysts, one had a fimbrial cyst, and two had no identified pelvic pathology).28 In the PLCO trial, the surgery rate occurring within 1 year of an abnormal ovarian palpation examination was 11.2 percent (at the longest followup), with a complication rate (any complication: surgical, pulmonary, cardiovascular, infection, other) of 1.0 percent (Table 7). Further diagnostic procedures occurring subsequent to and within 1 year after an abnormal palpation examination in the trial are reported in Table 7 (P. Pinsky, written communication, May 2, 2016).21
Summary of Findings

No screening studies were conducted solely in asymptomatic primary care populations. Two studies with large proportions of symptomatic patients had substantial clinical and methodological heterogeneity (populations, personnel performing index test, description of results of index tests, reference standards) and statistical heterogeneity (disparate accuracy results). These limitations hindered conclusions regarding the accuracy of the pelvic examination as a screening test for BV. Both included studies should be considered exploratory, hypothesis-generating investigations that cannot be used to estimate the accuracy of the pelvic examination as a screening test for BV in primary care populations.

Study Characteristics

We identified two fair-quality U.S. studies (n=930) that assessed the accuracy of different approaches to diagnosing BV, including pelvic examination. Gutman et al recruited any woman undergoing a speculum examination from a hospital-based primary care, colposcopy, or research clinic, whereas Eschenbach et al recruited nonpregnant women ages 16 to 50 years from an STI clinic (Table 8). In the first study, personnel performing “routine pelvic examination” included second- through fourth-year obstetrician-gynecologists residents, research nurses, or an attending gynecologist, while the second study specified that a single “women’s health care specialist” performed a “standardized pelvic examination,” with specific attention to the appearance of the vulva, vagina, and cervix; characteristics of vaginal discharge; and cervical, uterine, and adnexal tenderness.

The Gutman study provided some details on patient characteristics and reported a mean age of 24.1 years, with 38 percent of patients being white, 30 percent black, and 27 percent Hispanic. Thirty-three percent of patients in the Gutman study were symptomatic, while the STI clinic study reported 59 percent presenting with some pelvic or abdominal symptom as a chief complaint. Risk factors for BV were not reported for either study.

Neither study had a primary aim of estimating the accuracy of the pelvic examination; instead, they explored different clinical signs and diagnostic criteria for BV measured against a gold standard. The aim of the Gutman study was to report the diagnostic accuracy of using any two of Amsel’s criteria compared with the traditional diagnostic criteria of three of the four Amsel’s criteria (thin, homogeneous discharge; vaginal pH >4.5; positive whiff test or release of amine odor with potassium hydroxide; and clue cells on saline wet preparation microscopy). The aim of the Eschenbach study was to compare the observed findings on clinical pelvic examination with the gold-standard diagnostic criteria for BV. In the Gutman study, the index test included Amsel’s criteria with a score of 3 or greater; the reference test was a Gram stain with a Nugent’s criteria score of 7 or greater. The Eschenbach study used the index test of a standard pelvic examination reporting the accuracy of individually observed findings from a physical examination (homogeneous, frothy, increased, or yellow vaginal discharge; ectopy; and adnexal tenderness) compared with the reference standard of pH level and Gram stain microscopy. BV was diagnosed if the Gram stain revealed *Gardenerella*, one or more other bacterial morphologic...
types, and *Lactobacillus* (at quantities of ≤4 per oil immersion field).

**Yield and Accuracy**

Both studies reported a high prevalence of BV (39% and 47%). Gutman et al. reported the sensitivity and specificity of thin, homogeneous discharge as 79 and 54 percent, respectively; PPV and NPV were calculated to be 52 and 80 percent, respectively (Table 9). Eschenbach et al. reported the sensitivity and specificity of homogeneous discharge as 69 and 97 percent, respectively (PPV, 95%; NPV, 79%). That study also reported sensitivity for other individual pelvic examination findings, which ranged from 2 percent (frothy discharge) to 51 percent (ectopy), and specificity, which ranged from 48 percent (ectopy) to 100 percent (frothy discharge). PPV ranged from 42 to 100 percent and NPV from 52 to 58 percent. Data reporting did not allow for calculations for strictly asymptomatic patients.

**Harms**

A possible harm of using pelvic examination to screen for BV is that the test could result in false-positive or false-negative results, leading to unnecessary diagnostic workup for some women, while others would not receive indicated treatment. In the Gutman study, the false-positive and false-negative rates for the pelvic examination finding of thin, homogeneous discharge were 46 and 21 percent, respectively. In the second study, the false-positive and false-negative rates for homogeneous discharge were 3 and 31 percent, respectively. Individual false-positive and false-negative rates for other signs ranged from 0 to 52 percent and 49 to 98 percent, respectively.

**Genital Herpes (Herpes Simplex Virus-1 or -2)**

**Summary of Findings**

No screening studies were conducted solely in asymptomatic primary care populations. The single available study on the accuracy of pelvic examination to detect genital herpes recruited women from an STI clinic who were at high risk for the condition. Nonetheless, even in this higher STI prevalence population, 78 percent of women who had contracted any type of genital herpes at any time had asymptomatic shedding or latent disease, which would not be detectable with a pelvic examination. In this single study, the pelvic examination finding of vulvar ulcerations had a sensitivity of 20 percent and specificity of 98 percent in detecting genital herpes simplex virus (HSV) at any stage.

**Study Characteristics**

One fair-quality trial by Koutsky et al. (n=779) assessed the accuracy of approaches to detect genital herpes infection and provided data on the accuracy of specific pelvic examination findings in detecting this condition. The study recruited nonpregnant women ages 16 to 50 years from the same population seen in the STI clinic for the BV study by Eschenbach and for the trichomoniasis study by Wolner-Hansson et al. (Table 8). All pelvic examinations were performed by one “women’s health care specialist.” Mean age was 24 years and 70 percent of
participants were white. Almost all patients were sexually active (98%). Seven percent used condoms as the primary method of contraception and 33 percent did not use any contraception. Ten percent were symptomatic.

The index test was a “genital examination” with colposcopy (our report did not include lesions detected by colposcopy). A positive pelvic examination result was defined as clinician-detected lesions, but we could use reported data for only vulvar ulcerations and tender inguinal nodes in the accuracy calculations. All patients received the reference test, which included cultures from urine, cervical swabs, anal swabs, and any lesion swabs (all HSV isolates were confirmed and typed by direct immunofluorescence with use of mouse monoclonal antibodies), as well as serum testing for HSV-1 or HSV-2 antibodies using the western blot. The authors clearly defined cases as first episode (bilateral painful multiple vesicles, pustules, or ulcers on external genitalia, perineum, perianal area, or vaginal walls; cervical necrosis; or unilateral lesions plus constitutional symptoms without history of similar episodes, plus culture positivity and HSV-2 antibody negativity), recurrent episode (unilateral painful lesions on external genitalia, perineum, or perianal area; or bilateral small lesions and similar history or cervical ulcers without associated constitutional symptoms, plus culture or HSV-2 antibody positivity), asymptomatic viral shedding (no signs or symptoms, plus culture or antibody positivity), or latent subclinical infection (no signs or symptoms, plus HSV-2 antibodies present).

**Yield and Accuracy**

Nearly half (48%) of all study participants were diagnosed with genital herpes at some stage of the disease: 6 percent were diagnosed at the first episode, 5 percent had symptomatic recurrence, 2 percent had asymptomatic shedding, and 35 percent had latent HSV-2 infection (Table 9). Among patients at any stage of genital herpes, 22 percent were symptomatic on examination. We calculated the specificity and sensitivity of specific individual clinical findings as reported in the study (i.e., vulvar ulcerations, tender inguinal nodes) but were unable to use data on cervical ulcers because these numbers were aggregated to include cervical ulcers detected grossly with speculum examination as well as those found using a colposcope.

The presence of vulvar ulcerations had a sensitivity of 20 percent and specificity of 98 percent in detecting genital HSV at any stage (PPV, 88%; NPV, 57%). Similarly, the presence of tender inguinal lymphadenopathy had a sensitivity of 14 percent and specificity of 97 percent (PPV, 82%; NPV, 55%). Data reporting did not allow calculations for strictly asymptomatic patients.

**Harms**

For the clinical finding of vulvar ulceration, the false-positive and false-negative rates were 2 and 80 percent, respectively, for any stage of genital herpes. For the clinical finding of tender lymphadenopathy, the false-positive and false-negative rates were 3 and 86 percent, respectively, for any stage of genital herpes.
Trichomoniasis

Summary of Findings

There were no screening studies conducted solely in asymptomatic primary care populations. One study of women with a high prevalence of symptoms (>50%) for *Trichomonas vaginalis* (trichomoniasis) who were recruited from an STI clinic provided an exploratory analysis of the accuracy of individual clinical examination findings, showing less than 60 percent sensitivity for detecting trichomoniasis.\(^{35}\)

Study Characteristics

The one study, a fair-quality trial by Wolner-Hanssen et al (n=779), aimed to analyze the clinical manifestations of trichomoniasis and determine the accuracy of specific clinical findings on pelvic examination in detecting trichomoniasis.\(^{35}\) The study recruited nonpregnant women ages 16 to 50 years from the same population seen in the STI clinic as the BV study by Eschenbach\(^ {33}\) and the HSV study by Koutsky (*Table 8*).\(^ {34}\) All pelvic examinations were performed by a single “women’s health care specialist.” Patient characteristics were described previously (Genital Herpes, Study Characteristics) from this random sample of STI clinic patients. At least half of the patients had vaginal symptoms: yellow discharge (23%), abnormal vaginal odor (36%), and vulvar itching (51%).

The index test was a standardized pelvic examination that included colposcopy. For the purposes of this report, we did not consider any findings from colposcopic examination. The definition of an abnormal finding was reported as increased vaginal fluid, and reported if the clinician noted moderate to markedly increased discharge compared with that seen in patients without genital infections. Saline preparation as well as Gram stain of vaginal samples were examined under a microscope at 100× and 400× magnification. All patients were specifically evaluated for colpitis macularis (“strawberry cervix,” defined as diffuse or patchy maculoerythematos lesions of the ectocervical epithelium). The reference test for identifying trichomoniasis was culture.

Yield and Accuracy

The prevalence of culture-confirmed trichomoniasis was 15 percent. For the most specific clinical sign, colpitis macularis (detected grossly, without a colposcope), we calculated the sensitivity as 2 percent and the specificity as 100 percent; PPV was calculated as 100 percent and NPV as 85 percent (*Table 9*). For other individual clinical findings, sensitivity ranged from 8 to 59 percent and specificity from 72 to 99 percent; PPV ranged from 19 to 62 percent and NPV from 86 to 91 percent (*Table 9*). Data reporting did not provide sufficient information to calculate the sensitivity and specificity for the presence of any one or more abnormal findings on pelvic examination. Data reporting did not allow calculations for strictly asymptomatic patients.

Harms

Pelvic examination screening for trichomoniasis could result in missed cases whereby women do not receive indicated treatment. For the clinical findings of colpitis macularis, purulent
discharge, frothy discharge, vulvar erythema, and vaginal erythema, the false-positive rate ranged from 0 percent (colpitis macularis) to 28 percent (vulvar erythema). The false-negative rate ranged from 41 percent (purulent discharge) to 98 percent (colpitis macularis). The degree of harm from false-positive results is expected to be minimal given that the diagnostic test is benign and confirmation is conducted in the clinic, without delay, between a positive screening result and confirmation.

**Other Harms**

**Summary of Findings**

Beyond the false-positive rates and missed cases (described above) from accuracy studies, we identified one additional small fair-quality cohort study investigating a possible association between the pelvic examination and subsequent development of urinary symptoms. Further research is needed, in larger studies with urine culture-confirmed urinary tract infection (UTI) as the outcome, to confirm or disprove this potential harm.

**Characteristics of Included Studies**

In addition to the studies of harms related to the sensitivity and specificity estimates reported in diagnostic accuracy studies, we identified one study that considered the possibility of genitourinary infection being caused by routine pelvic examinations.\(^3^6\) In this poor-to-fair quality 4-week prospective controlled cohort (n=150), sexually active women ages 18 to 40 years were seen in a university-based family medicine residency clinic. Subjects presented for a screening speculum examination, Pap test, and bimanual examination, while age-matched controls presented for other kinds of health maintenance visits. The study excluded women who had current or chronic urinary or vaginal symptoms, were being treated with anti-infectives, had diabetes, or were taking immunosuppressants. Outcomes were obtained through daily self-reported logs of urinary symptoms as well as medical chart review for a UTI diagnosis. Half of the enrolled patients (49%) dropped out before the end of the 4-week trial.

The average age of those completing the study was 26 years. The groups differed in two statistically significant ways: the control patients had intercourse more frequently and they used condoms more often than the group that had undergone pelvic examination.

**Results**

UTI was diagnosed clinically (without culture) in only one subject in the pelvic examination group. Dysuria (11/63 vs. 6/87; \(p<0.01\)) and urinary frequency (17/63 vs. 12/87; \(p<0.01\)) were more common in the pelvic examination group during the 4-week followup. This study was limited in that there was high loss to followup, it was underpowered to detect a difference in UTI diagnoses between groups, and there were significant between-group differences in intercourse frequency and barrier use. This exploratory study cannot be used to make conclusions about the causality between pelvic examinations and UTIs.
Chapter 4. Discussion

We identified no literature that assessed the overall value of the pelvic examination in improving health outcomes for any medical conditions. Despite the many medical conditions that are plausibly detectable or that physicians cite as a rationale for routine screening pelvic examinations, our review identified diagnostic accuracy studies for only one cancerous condition (ovarian cancer) and three infectious conditions (BV, trichomoniasis, and HSV) (Table 10).

Screening Accuracy

Ovarian Cancer

Our systematic review findings are consistent with other recent systematic reviews, but our review is the first to present unpublished data from the PLCO trial from more than 20,000 screened women (P. Pinsky, written communication, May 2, 2016). All four studies included in our review recruited average-risk women; unsurprisingly, the prevalence of ovarian cancer was quite low (0.04% to 0.10%) and the PPVs were also consistently low (<4%). These rare cases reflect the low incidence of ovarian cancer seen in U.S. women (0.01% [11/100,000 women]). The downstream consequences resulting from positive pelvic examination findings include surveillance with ultrasonography (one-time or repeated) or other imaging and, in some cases, surgery. In our included studies, at 1 year, the prevalence of abnormal pelvic examination results ranged from 1.2 to 8.7 percent, and the prevalence of surgery among those women with abnormal findings ranged from 5 to 36 percent (P. Pinsky, written communication, May 2, 2016). Due to advances in ultrasound technology since the publication of these studies, it could be hypothesized that the surgical intervention rate lies on the lower end of this range. In the only included U.S. study, the large PLCO trial of almost all postmenopausal women, the ovarian palpation protocol occurred in the early to mid-1990s and reported a surgery rate of 11.2 percent within 1 year of an abnormal pelvic examination after one to four rounds of screening. There is no more recent evidence available to estimate the risk of surgical intervention resulting from screening pelvic examinations.

The PLCO trial used an even more sensitive screening procedure (transvaginal ultrasonography and blood testing for CA-125) and found abnormal ultrasonography results in 5 percent of women (n=1,338) who had received a baseline examination (n=28,519). Between the ultrasonography results and CA-125 levels, 6 percent of women required further assessment (n=1,703) and 2 percent (n=570) underwent surgery (n=325 laparotomy; n=245 laparoscopy and/or vaginal approach). Ninety-eight percent of women with an initial abnormal result on any of the screening tests used in the PLCO trial, and 94 percent of those who underwent surgery following a positive result were not diagnosed with ovarian cancer. Furthermore, evidence suggests that complex ovarian cysts detected on ultrasound are not immediate precursors to ovarian cancer. As mentioned earlier, the PLCO trial originally included bimanual examination of the ovaries and rectovaginal examination in the screening protocol. The pelvic examination component was discontinued, however, because no ovarian cancers were detected solely with bimanual palpation. Notably, even the more sensitive screening tests (transvaginal...
ultrasonography and CA-125 blood testing) were not found to significantly reduce rates of ovarian cancer–related mortality.41

Cross-sectional studies showed that under the most optimal circumstances (patients preoperatively under anesthesia, all with some pelvic abnormality as indication for surgery, examination performed by attending physicians), the accuracy of the pelvic examination for detecting pelvic masses is low.42, 43 Moreover, the accuracy of this examination to detect pelvic masses has been shown to be lower when performed by inexperienced trainee examiners, with obese patients, or with patients with an enlarged uterus.42-45 Our narrative synthesis of the limited available evidence suggests poor performance of the screening pelvic examination for detecting ovarian cancer.

**Infectious Diseases**

There were no diagnostic accuracy studies for infectious conditions in solely asymptomatic, average-risk populations. We did include four studies reporting diagnostic accuracy in high-risk settings (STI clinics or populations with high rates of symptoms), acknowledging that these studies likely overestimate accuracy characteristics when pelvic examinations are administered to average-risk, asymptomatic primary care populations. Again, even in these high-risk populations, the reported sensitivities are well below what would be considered minimal thresholds for clinically useful screening instruments.

**Harms**

Our review identified few studies on harms that met the inclusion criteria. The studies were largely the same ones from key question 2 (accuracy) on estimating indirect harms from false-positive results and missed cases, with an additional small cohort study by Tiemstra and colleagues36 on possible associations between pelvic examination and subsequent urinary symptoms. The potential downstream harms resulting from the diagnostic workup vary widely by gynecologic condition. For example, a false-positive result on pelvic examination for adnexal mass could result in ultrasonographic surveillance with or without diagnostic laparoscopy, while a false-positive finding on pelvic examination for abnormal vaginal discharge may result in the additional cost of a laboratory Gram stain or NAAT or even unnecessary empiric antibiotic treatment. From a patient’s perspective, and in the absence of empiric treatment, false-positive results associated with screening for trichomoniasis might have a modest impact given that the diagnostic testing and immediate confirmation or disconfirmation does not require an invasive test and there is little delay between the screening and diagnostic test results.

Another recent systematic review38 included 14 cross-sectional surveys46-59 and one cohort study60 addressing harms associated with pelvic examination and women’s attitudes about the examination; these studies were not included in our review because they did not meet the inclusion criteria. Exclusions were primarily due to their lack of generalizability and applicability to the U.S. primary care setting. The authors of that systematic review concluded that the pelvic examination may lead to pain, fear, anxiety, discomfort, or embarrassment in some proportion of women (range, 10% to 80%), but those data were of low quality. Additional
cross-sectional literature suggests that certain populations of women—especially those with a history of sexual violence\textsuperscript{61-64} or abuse,\textsuperscript{65} chronic pelvic pain,\textsuperscript{66} or obesity\textsuperscript{67}—report more negative experiences from a pelvic examination and, as a result, may avoid seeking medical care.

**Clinical Implications**

Controversy surrounds the clinical implications of changes in screening periods for cervical cancer with respect to the role of annual screening pelvic examination.\textsuperscript{4, 68-72} There are concerns that scrutiny of the accuracy of screening pelvic examination is misdirected, as the benefit of other routinely provided physical examination components (e.g., heart and lung auscultation) in the context of screening is similarly lacking.\textsuperscript{69} Rather than viewing the examination as a screening test, this perspective suggests that it is a point of contact with patients with broader clinical purpose, including to facilitate discussion of sensitive topics that would otherwise not be brought up.\textsuperscript{72} These concerns may echo physicians’ attitudes about the annual physical examination in general; one survey reported that most primary care physicians believe that an annual physical examination provides counseling time for preventive services, improves detection of subclinical disease, improves therapeutic relationships, and is desired by patients.\textsuperscript{73, 74} Others are concerned that clinicians rely on ultrasonography so heavily that clinical acumen for the pelvic examination has declined; this concern has been cited as one reason for continued performance of routine pelvic examinations.\textsuperscript{7, 70, 75} On the other hand, the screening pelvic examination can cause anxiety and discomfort and could pose unnecessary barriers to care,\textsuperscript{52} especially in certain subpopulations of women (e.g., those with a history of sexual abuse\textsuperscript{62-64}). Notably, despite recommendations to the contrary,\textsuperscript{76, 77} a survey reported that 79 percent of obstetrician-gynecologists thought at least one component of the pelvic examination was of some importance for determining contraception eligibility,\textsuperscript{78} and a survey from 2008 to 2009 found that nearly a third of obstetrician-gynecologists and family medicine physicians required the patient to undergo a pelvic examination before being prescribed oral contraceptives.\textsuperscript{79}

The fragmentation of preventive services in women’s health care is a well-recognized problem.\textsuperscript{80, 81} Women’s preventive care is provided by clinicians with diverse training and professional orientations, and women seek care from different types of providers over the lifespan, resulting in variability in the comprehensiveness of primary care.\textsuperscript{82-84} It is uncertain if changes to routine screening pelvic examination practices will affect women’s patterns of health care use and their receipt of comprehensive primary and preventive health care at different stages of life.

**Limitations**

Our systematic review captured all of the English-language published literature on the screening accuracy of the pelvic examination in asymptomatic, average-risk populations. In our initial abstract review, we conservatively included studies that could possibly meet the inclusion criteria for full-text review. In almost all cases, these studies did not meet inclusion criteria on full-text review. We specifically excluded studies recruiting participants at high risk for ovarian cancer (e.g., those with symptoms, known masses, or a family history of ovarian cancer)\textsuperscript{85-88} or those reporting incomplete data regarding accuracy\textsuperscript{89, 90}; these studies are summarized
elsewhere. We included STI accuracy studies in settings outside of primary care, but these studies probably included some average-risk patients, thus providing data on accuracy which should be cautiously interpreted for average-risk primary care populations. In addition, our study design inclusion criteria excluded several qualitative studies and survey studies on the harms of pelvic examination (including discomfort and anxiety), but we do not believe that those studies added precision to the estimation of screening harms.

The aim of the conceptual framework presented in our report was to define the potential yield as well as the presence of well-established, evidence-based alternative screening tests for conditions like cervical cancer, gonorrhea, and chlamydia. We acknowledge that no studies examined the yield of the pelvic examination in detecting any treatable pelvic pathology. In other words, while the sensitivity of the pelvic examination was low for detecting a single condition (e.g., ovarian cancer or BV), on the basis of current evidence we could not estimate the value of the screening pelvic examination to detect any condition in the list of possible disease conditions prior to clinical presentation with symptoms. Epidemiologic evidence for estimating the burden of a single condition, biologic plausibility of early detection using the pelvic examination, treatability of the disease at earlier stages, and alternative, superior screening approaches are considerations for clinical guidance, as there is no literature available to estimate the potential cumulative benefits or harms of screening pelvic examination.

This systematic review did not evaluate the role of history taking in eliciting symptoms in patients who do not present with gynecologic-related chief complaints but do affirm one or more gynecological symptoms on review of body systems. This topic was not considered in the scope of this systematic review since it was aimed at determining the effectiveness of routine screening pelvic examinations in average-risk, asymptomatic women in primary care settings. We did perform a targeted search of this approach to case-finding and found no relevant literature. Furthermore, studies examining the accuracy of the pelvic examination as a diagnostic tool for symptomatic patients (e.g., patients presenting with symptoms of pelvic pain, vaginal discharge, or dyspareunia) were outside the scope of this review. Pelvic examination remains an important tool in diagnosing pelvic pathology for symptomatic patients.

**Research Gaps**

We found no studies examining the effect of the pelvic examination on morbidity, mortality, or quality of life. Further, we identified no in-progress studies examining the effectiveness or screening accuracy of the pelvic examination for any condition. Given the inconsistent guidelines and limited evidence, trials randomizing women to different pelvic screening examination protocols could provide estimates of the benefits and harms for women at different stages of life. In light of the limited evidence on the clinical benefits or harms of routine pelvic examination for preventive screening, research questions related to improvements to women’s primary health care delivery warrant consideration. In the face of changing clinical practice with respect to cervical cancer screening frequency and new recommendations from professional organizations, it remains unclear if altering the schedule for routine pelvic examinations will influence the uptake of other evidence-based preventive services (e.g., blood pressure or obesity screening). Similarly, there is no literature on the opportunity costs the pelvic examination could
pose for the receipt of other recommended preventive services. These changes and others would be important to study given the longstanding patterns of women’s primary health care delivery. Also needed are investigations that compare strategies for implementing patient-centered approaches to preventive screening along with provider- and patient-focused strategies to continue improving the delivery of evidence-based prevention.

Further research into the primary care gaps and coordination issues that women face in obtaining comprehensive primary care might help to clarify optimal patterns of health care use from the patient’s perspective as well as from public health and clinical standpoints. It remains unclear what components of physical examination in routine primary care visits are most important for maintaining health and whether women’s patterns of health care use would change if routine pelvic examinations were not recommended per the American College of Physicians’ guidelines or were based on shared decisionmaking conversations, as suggested in the recent WWTF guidelines. Changes to routine practices of pelvic examination could either improve or worsen the comprehensiveness and continuity of primary care for women; the effects could also depend on a woman’s age or overall health. Patients’ expectations and preferences for pelvic examination also warrant further investigation, as the current guidance on screening pelvic examinations from ACOG and WWTF suggest a shared decisionmaking paradigm wherein these personal perspectives would inform practices. Regardless of the need for targeted preventive screening services, some women may wish to have an annual gynecological visit.

Conclusions

No studies have provided evidence of the health benefits—and limited evidence on the accuracy—of the screening pelvic examination for gynecologic conditions that might be detected. Although lack of evidence is not conclusive evidence of no benefit, the existing evidence highlights the limited sensitivity of screening pelvic examination in detecting ovarian cancer and select infections. Asymptomatic conditions that might be detected during a routine pelvic examination and treated to improve a patient’s health have not been fully outlined and evaluated for test accuracy. Nor has there been sufficient evidence to fully evaluate the potential harms of the examination. Furthermore, no studies have examined the health outcomes for screened and unscreened populations to provide overarching evidence of the overall benefits and harms of the screening pelvic examination across all potential health conditions prevented.
References


68. Saraiya M. Evidence is lacking to support pelvic examinations as a screening tool for non-cervical cancers or other conditions. Evid Based Med. 2015;20(1):31. PMID: 25260357.


111. Sobel JD. Trichomoniasis. UpToDate. 2014.


Figure 1. Analytic Framework

Unselected Adult Females, Asymptomatic, Not Pregnant

Screening by Pelvic Exam

Abnormal Pelvic Exam

Evaluation and Early Detection

Gynecologic Cancers

Other Gynecologic Conditions

Treatment

Decreased All-Cause Mortality
Decreased Cancer- and Disease-Specific Morbidity and Mortality
Improved Quality of Life

Harms of Screening

Harms of Evaluation

Harms of Treatment

Screening With Pelvic Examination 30 Kaiser Permanente Research Affiliates EPC
### Table 1. Epidemiology of Gynecologic Cancers and Conditions*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Population Affected</th>
<th>Burden/Epidemiology</th>
<th>Typical Clinical Presentation</th>
<th>Expected Pelvic Examination Finding in Asymptomatic Women</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometrial</td>
<td>Primarily postmenopausal (mean age of diagnosis, 60 years)</td>
<td>Incidence rate: 25.1/100,000† Mortality rate: 4.4/100,000†39</td>
<td>Abnormal vaginal bleeding93</td>
<td>Enlarged uterus on bimanual exam, gross lesions on internal speculum exam (advanced disease)</td>
</tr>
<tr>
<td>Ovarian</td>
<td>All ages, most frequently those ages 55–64 years</td>
<td>Incidence rate: 11.3/100,000† Mortality rate: 7.4/100,000†39</td>
<td>Persistent, vague symptoms (usually after metastasizing)95</td>
<td>Enlarged adnexa, ascites (bimanual exam)</td>
</tr>
<tr>
<td>Vaginal</td>
<td>All ages, but usually those age ≥60 years</td>
<td>Incidence rate: 0.7/100,000† Mortality rate: 0.2/100,000†39</td>
<td>Vaginal discharge; abnormal bleeding; change in bathroom habits; pelvic or abdominal pain, dysuria, dyspareunia97</td>
<td>Gross vaginal lesions on internal speculum exam</td>
</tr>
<tr>
<td>Vulvar</td>
<td>All ages, but mostly women ages 75–84 years</td>
<td>Incidence rate: 2.6/100,000† Mortality rate: 0.5/100,000†39</td>
<td>Itching, burning, or bleeding on the vulva; changes in vulva skin color or appearance; sores, lumps, or ulcers on vulva; pelvic pain, dysuria, dyspareunia97</td>
<td>Gross vulvar lesions on external exam</td>
</tr>
<tr>
<td><strong>Infectious diseases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial vaginosis</td>
<td>All ages, most commonly women ages 15–44 years</td>
<td>Most common vaginal infection among females ages 15–44 years 99</td>
<td>Often asymptomatic99</td>
<td>Asymptomatic discharge on internal speculum exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prevalence: 29.2% (ages 14–49 years)100</td>
<td>Malodorous vaginal discharge</td>
<td></td>
</tr>
<tr>
<td>Candidiasis</td>
<td>All ages; but those who have diabetes, are pregnant, have long-term use of broad-spectrum antibiotics, or use corticosteroid medications are at higher risk101</td>
<td>Nearly 75% of adult women have had ≥1 candidiasis occurrence101</td>
<td>Symptomatic vaginal discharge, pruritus</td>
<td>Vaginal discharge detected on internal speculum exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Between 29% and 49% of premenopausal women had ≥1 lifetime episode102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genital warts (HPV)</td>
<td>All ages</td>
<td>120 incident cases/100,000 women each year103</td>
<td>Asymptomatic or symptomatic, depending on location and size of warts</td>
<td>Gross lesions on external or internal speculum exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lifetime history of anogenital warts: 7.2%103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herpes (HSV-1, HSV-2)</td>
<td>All ages</td>
<td>20.9% (HSV-2) among females ages 14–49 years104</td>
<td>Asymptomatic or primary/secondary disease with typical labial ulcerative lesions</td>
<td>Gross ulcerative lesions on external or internal speculum exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>52.3% (HSV-1) among women ages 20–29 years, 33.2% among women ages 14–19 years105</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Screening With Pelvic Examination* 31 Kaiser Permanente Research Affiliates EPC
<table>
<thead>
<tr>
<th>Condition</th>
<th>Population Affected</th>
<th>Burden/Epidemiology</th>
<th>Typical Clinical Presentation</th>
<th>Expected Pelvic Examination Finding in Asymptomatic Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pelvic inflammatory disease</td>
<td>Sexually active, especially with untreated STIs [^{106}]</td>
<td>Proportion of women (ages 15–44 years) ever treated for PID: 5.0% [^{107, 108}]</td>
<td>Abdominal or pelvic pain, discharge, abnormal vaginal bleeding, fever or chills</td>
<td>Tenderness on bimanual exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diagnosis rate (ages 15–44 years): 236.0/100,000 [^{109}]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trichomoniasis</td>
<td>All ages, but more commonly older women [^{110}]</td>
<td>Prevalence (ages 14–49 years): 3.1% [^{111}]</td>
<td>Often asymptomatic [^{111}]</td>
<td>Discharge, colpitis macularis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Purulent, malodorous, thin discharge associated with burning, pruritus, dysuria, frequency, lower abdominal pain, or dyspareunia [^{111}]</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Atrophic vaginitis</td>
<td>Prevalence from 4% in premenopausal women to 47% in postmenopausal women [^{112}]</td>
<td>Reported symptoms (dyspareunia, spotting, vaginal discharge, burning, soreness)</td>
<td>Atrophic changes on internal speculum exam</td>
</tr>
<tr>
<td>Cervical polyps</td>
<td>All ages; most commonly among parous women age ≥20 years [^{113}]</td>
<td>NR</td>
<td>Often asymptomatic; abnormal bleeding [^{113}]</td>
<td>Cervical polyp on internal speculum exam</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>All ages; most commonly among women ages 25–35 years</td>
<td>Prevalence in the general population is unknown</td>
<td>Dysmenorrhea, pelvic pain, dyspareunia, infertility, bowel upset, bowel pain, ovarian mass, dysuria, other urinary problems</td>
<td>Pelvic mass could be detected with bimanual exam on ovaries, uterus, peritoneum, and uterosacral ligaments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1% undergoing major surgery for any gynecologic indication [^{114}]</td>
<td>May occur asymptotically</td>
<td>Less commonly, internal speculum exam could detect an endometric lesion on the cervix or vaginal mucosa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1%–7% undergoing tubal sterilization [^{114}]</td>
<td>Most present symptomatically (chronic pelvic pain, 71%–87%) [^{115}]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12%–32% of reproductive age undergoing laparoscopy to determine the cause of pelvic pain [^{114}]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9%–50% undergoing laparoscopy for infertility [^{114}]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Population Affected</td>
<td>Burden/Epidemiology</td>
<td>Typical Clinical Presentation</td>
<td>Expected Pelvic Examination Finding in Asymptomatic Women</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Ovarian cysts</td>
<td>All ages</td>
<td>Simple cyst at initial visualization (ages 55–74 years): 15%**</td>
<td>Often asymptomatic; pelvic pain</td>
<td>Ovarian mass and/or tenderness on bimanual exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-year incidence of new simple cysts (ages 55–74 years): 8%**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urinary incontinence prevalence: 15.7%††118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fecal incontinence prevalence: 9.1%††118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pelvic organ prolapse prevalence: 2.9%††118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvic floor dysfunction/</td>
<td>Older, obese, hysterectomized,</td>
<td>≥1 pelvic floor disorder: 25.0%117</td>
<td>Asymptomatic in early stages; urinary obstruction or incontinence, bowel incontinence</td>
<td>Weak pelvic floor muscles prolapse on internal speculum and/or bimanual exam</td>
</tr>
<tr>
<td>Pelvic organ prolapse</td>
<td>pregnant, labored, or gave birth</td>
<td>Urinary incontinence prevalence: 15.7%††118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fecal incontinence prevalence: 9.1%††118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pelvic organ prolapse prevalence: 2.9%††118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uterine fibroids</td>
<td>Most women diagnosed at ages 30–50 years</td>
<td>Self-reported prevalence: 6.9%</td>
<td>Dysmenorrhea, menorrhagia, metromenorrhagia</td>
<td>Uterine mass on bimanual exam</td>
</tr>
<tr>
<td>Vulvar lichen sclerosis</td>
<td>All ages, but most commonly peri- or</td>
<td>True prevalence unknown</td>
<td>Vulvar pruritus, dyspareunia, dysuria, soreness, irritation are common symptoms; may occur asymptatically</td>
<td>Characteristic thin, white, atrophic skin and changes in vulvar architecture</td>
</tr>
<tr>
<td></td>
<td>postmenopausal women†20</td>
<td>Ranges from 1/30 (older women) to 1/59 (women in a general gynecologic practice)140</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Cervical cancer, gonorrhea, and chlamydia are not included in this table because there are existing strong recommendations for alternative screening methods from the USPSTF and other guideline groups.
† 2012 rates (per 100,000 women).
¶ Diagnosis rates for women ages 15–44 years enrolled in private insurance plans in 2005.
** 2010 projection.
†† From the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; only women ages 55–74 years were included.

**Abbreviations:** HPV=human papillomavirus; HSV=herpes simplex virus.
Table 2. Recommendations on the Periodic Pelvic Examination for Asymptomatic Adult Women

<table>
<thead>
<tr>
<th>Group or Professional Society</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAFP</td>
<td>Screening pelvic examination in asymptomatic, nonpregnant, adult women is not recommended. (Based on the ACP recommendation.)</td>
</tr>
<tr>
<td>ACOG</td>
<td>No evidence supports the routine internal examination of the healthy, asymptomatic patient before age 21 years. Recommends that a pelvic examination be performed in all patients age 21 years and older. No evidence supports or refutes the annual pelvic examination or speculum and bimanual examination for the asymptomatic, low-risk patient. The decision whether or not to perform a complete pelvic examination at the time of the periodic health examination for the asymptomatic patient should be a shared decision after a discussion between the patient and her health care provider.</td>
</tr>
<tr>
<td>ACP</td>
<td>Recommends against performing screening pelvic examination in asymptomatic, nonpregnant, adult women (strong recommendation, moderate-quality evidence).</td>
</tr>
<tr>
<td>WWTF</td>
<td>For patients age 21 years and older, recommends speculum and/or bimanual examination for asymptomatic patients with specific indications (e.g., intrauterine device placement, cervical cancer screening). External examinations may be performed annually in healthy patients. The inclusion of speculum, bimanual examination, or both in otherwise well women should evolve from informed decisionmaking between patient and provider. (Qualified) (Based on the ACOG recommendation.)</td>
</tr>
</tbody>
</table>

*Abbreviations:* AAFP=American Academy of Family Physicians; ACOG=American College of Obstetricians and Gynecologists; ACP=American College of Physicians; WWTF=Well-Woman Task Force.
Table 3. U.S. National Guidelines and Statements on Screening for Individual Gynecologic Conditions in Unselected Adult Women Who Are Asymptomatic and Not Pregnant*

<table>
<thead>
<tr>
<th>Condition</th>
<th>USPSTF</th>
<th>AAFP</th>
<th>ACOG</th>
<th>ACS</th>
<th>CDC</th>
<th>WWTF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometrial</td>
<td>-</td>
<td>-</td>
<td>No screening test†</td>
<td>No indication that screening is warranted for women with no identified risk factors‡</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ovarian</td>
<td>Recommend against screening (D)123</td>
<td>Recommend against screening (D)124</td>
<td>No effective strategy for screening125</td>
<td>Currently no reliable screening tests126</td>
<td>-</td>
<td>Screening for ovarian cancer is not recommended (Strong†)</td>
</tr>
<tr>
<td>Vaginal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Vulvar</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Infectious disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial vaginosis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Candidiasis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Genital warts</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(HPV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herpes (HSV-1, HSV-2)</td>
<td>Recommend against screening (D)128</td>
<td>Recommend against screening (D)129</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PID</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Trichomoniasis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrophic vaginitis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cervical polyps</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ovarian cysts</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pelvic organ prolapse</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Uterine fibroids</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Recommendations for cervical cancer, gonorrhea, and chlamydia are not included in this table because there are existing strong recommendations for screening from the USPSTF and other guideline groups.

† Based on evidence-based or evidence-informed guidelines.

Abbreviations: AAFP=American Academy of Family Physicians; ACOG=American College of Obstetricians and Gynecologists; ACS=American Cancer Society; CDC=Centers for Disease Control and Prevention; HPV=human papillomavirus; HSV=herpes simplex virus; USPSTF=U.S. Preventive Services Task Force; WWTF=Well-Woman Task Force.
<table>
<thead>
<tr>
<th>Screening Topic, Year</th>
<th>Grade</th>
<th>Recommendation/Statement</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial vaginosis in pregnancy, 2008&lt;sup&gt;130&lt;/sup&gt;</td>
<td>D</td>
<td>Do not screen pregnant women at low risk of preterm birth. Evidence is insufficient to make a recommendation for women at high risk of preterm birth.</td>
<td>--</td>
</tr>
<tr>
<td>Cervical cancer, 2012&lt;sup&gt;2&lt;/sup&gt;</td>
<td>A</td>
<td>Screen women ages 21–65 years using cytology and women ages 30–65 years using cytology and human papillomavirus testing. Do not screen women age &lt;21 years or &gt;65 years who have had adequate prior screening, are not at high risk of cervical cancer, had a hysterectomy with removal of the cervix, or have no history of high-grade precancerous lesion or cervical cancer.</td>
<td>Update in progress</td>
</tr>
<tr>
<td>Chlamydia, 2014&lt;sup&gt;91&lt;/sup&gt;</td>
<td>B</td>
<td>Screen sexually active women age &lt;25 years and older women at increased risk of chlamydia.</td>
<td>--</td>
</tr>
<tr>
<td>Gonorrhea, 2014&lt;sup&gt;91&lt;/sup&gt;</td>
<td>B</td>
<td>Screen sexually active women age &lt;25 years and older women at increased risk of gonorrhea.</td>
<td>--</td>
</tr>
<tr>
<td>Herpes simplex, genital, 2005&lt;sup&gt;128&lt;/sup&gt;</td>
<td>D</td>
<td>Do not screen asymptomatic pregnant women, adults, and adolescents.</td>
<td>Update in progress</td>
</tr>
<tr>
<td>Ovarian cancer, 2012&lt;sup&gt;123&lt;/sup&gt;</td>
<td>D</td>
<td>Do not screen asymptomatic women without known genetic mutations.</td>
<td>Update in progress</td>
</tr>
</tbody>
</table>

A = Strongly Recommended: The USPSTF strongly recommends that clinicians provide [the service] to eligible patients. The USPSTF found good evidence that [the service] improves important health outcomes and concludes that benefits substantially outweigh harms.

B = Recommended: The USPSTF recommends that clinicians provide [the service] to eligible patients. The USPSTF found at least fair evidence that [the service] improves important health outcomes and concludes that benefits outweigh harms.

C = No Recommendation: The USPSTF makes no recommendation for or against routine provision of [the service]. The USPSTF found at least fair evidence that [the service] can improve health outcomes but concludes that the balance of benefits and harms is too close to justify a general recommendation.

D = Not Recommended: The USPSTF recommends against routinely providing [the service] to asymptomatic patients. The USPSTF found at least fair evidence that [the service] is ineffective or that harms outweigh benefits.

I = Insufficient Evidence to Make a Recommendation: The USPSTF concludes that the evidence is insufficient to recommend for or against routinely providing [the service]. Evidence that the [service] is effective is lacking, of poor quality, or conflicting and the balance of benefits and harms.

**Abbreviation:** USPSTF=U.S. Preventive Services Task Force.
Table 5. Study Characteristics, Ovarian Cancer Screening

| Author, Year | Country/Recruitment Setting | Study Aim | Inclusion/Exclusion Criteria | N Screened | Age | Screening Test Description | Reference Standard Description | Yield of Ovarian Cancer, n/N (%) |
|--------------|-----------------------------|-----------|-----------------------------|-----------|-----|---------------------------|--------------------------------|--------------------------------|---------------------------------|
| PLCO, 2016*  | US Community                | To determine the effect of specific cancer screening tests on cause-specific mortality | **Inclusion:** Women ages 55–74 years **Exclusion:** Undergoing treatment for cancer (excluding basal cell and squamous cell skin cancer); known prior cancer of the lung, colon, rectum, or ovary; previous surgical removal of 1 lung or the entire colon; had a colonoscopy, sigmoidoscopy, or barium enema in the past 3 years; unable or unwilling to sign the consent form | 20,872 | 62.9 (mean) 55–74 (range) | Palpable ovarian mass or cul-de-sac nodularity (for obese patients with nonpalpable ovaries, the exam was considered negative) | Diagnosis of ovarian cancer within 1 to 5 years of exam based on medical records, patient questionnaires for ovarian cancer diagnoses (All women also received a TVU and CA-125 measurement. While these results were available to the provider, they are not being used as a reference standard.) | 23/20,872 (0.1) for 1-year followup 72/20,872 (0.3) for 1- to 5-year followup* |
| Adonakis, 1996* | Greece Community | Investigate effectiveness of pelvic exam and CA-125 followed by ultrasonography as a screening method | **Inclusion:** Age ≥45 years without any evidence of adnexal pathology **Exclusion:** History of ovarian cancer (familial or not) or any other malignancy; bilateral oophorectomy; with ascites | 2000 | 58.1 (mean) 45–80 (range) | Detection of palpable adnexal mass on pelvic exam | Transvaginal ultrasonography for those with abnormal pelvic exam result or serum CA-125 ≥35 U/mL 1-year followup visit to measure CA-125 levels for those with normal pelvic exam results and CA-125 levels | 2/2000 (0.1) |
| Grover, 1995* | Australia Community | Assess effectiveness of serum CA-125 plus vaginal exam as a screening test | **Inclusion:** Apparently healthy and age ≥40 years (younger females included if they had a family history of ovarian cancer) **Exclusion:** NR | 2550 | 51 (median) 21–92 (range) | Adnexal mass was palpable during bimanual exam in postmenopausal females, or if a larger than normal-size ovary was palpable in premenopausal females | Abdominal and/or vaginal ultrasonography for those with abnormal pelvic exam result or serum CA-125 >35 U/mL 1-year postal questionnaire for all patients | 1/2550 (0.04) |
Table 5. Study Characteristics, Ovarian Cancer Screening

<table>
<thead>
<tr>
<th>Author, Year Quality</th>
<th>Country Recruitment Setting</th>
<th>Study Aim</th>
<th>Inclusion/Exclusion Criteria</th>
<th>N Screened</th>
<th>Age</th>
<th>Screening Test Description</th>
<th>Reference Standard Description</th>
<th>Yield of Ovarian Cancer, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacobs, 1988** Fair</td>
<td>UK Community</td>
<td>Examine screening capabilities of vaginal exam, CA-125, and ultrasonography in various combinations</td>
<td>Inclusion: Age &gt;45 years; amenorrheic for &gt;12 months Exclusion: History of ovarian cancer or bilateral oophrectomy; being treated for any malignancy</td>
<td>1010</td>
<td>54.0 (median) 45–83 (range)</td>
<td>Palpable pelvic mass of any size that could be clinically distinguished as being separate from the uterus and gastrointestinal tract during vaginal exam</td>
<td>Abdominal ultrasonography for those with abnormal pelvic exam result or serum CA-125 &gt;30 U/mL 1-year postal questionnaire for those with normal pelvic exam results and CA-125 levels</td>
<td>1/1010 (0.1)</td>
</tr>
</tbody>
</table>

* Includes all participants also diagnosed within the first year (n=23).

**Abbreviations: CA-125=cancer antigen 125; N=number; n=number; NR=not reported; PLCO=Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial.
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Quality</th>
<th>Country</th>
<th>N</th>
<th>Age (mean)</th>
<th>Followup</th>
<th>Yield, n (%)</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>FPR (95% CI)</th>
<th>FNR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLCO, 2016**&lt;sup&gt;†&lt;/sup&gt; 3 (2016 personal communication with Dr. Paul Pinsky) Good US</td>
<td>20,872 63 years</td>
<td>1 year</td>
<td>23/20,872 (0.1)</td>
<td>1</td>
<td>22</td>
<td>242</td>
<td>20,607</td>
<td>4.3 (0.5 to 18.6)</td>
<td>98.8 (98.7 to 99.0)</td>
<td>0.4 (0.04 to 1.9)</td>
<td>99.9 (99.8 to 99.9)</td>
<td>1.2 (1.0 to 1.3)</td>
<td>95.7 (81.4 to 99.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 years</td>
<td>72/20,872 (0.3)</td>
<td>2</td>
<td>70</td>
<td>241</td>
<td>20,559</td>
<td>2.8 (0.6 to 8.6)</td>
<td>98.8 (98.7 to 99.0)</td>
<td>0.8 (0.2 to 2.6)</td>
<td>99.7 (99.6 to 99.7)</td>
<td>1.2 (1.0 to 1.3)</td>
<td>97.2 (91.4 to 99.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adonakis, 1996**&lt;sup&gt;‡&lt;/sup&gt; Fair Greece</td>
<td>2000 58 years</td>
<td>1 year</td>
<td>2 (0.1)</td>
<td>2</td>
<td>0</td>
<td>172</td>
<td>1826</td>
<td>100.0&lt;sup&gt;‡&lt;/sup&gt; (33.3 to 100.0)</td>
<td>91.4&lt;sup&gt;‡&lt;/sup&gt; (90.1 to 92.6)</td>
<td>1.2&lt;sup&gt;‡&lt;/sup&gt; (0.2 to 3.6)</td>
<td>100.0&lt;sup&gt;‡&lt;/sup&gt; (99.9 to 100.0)</td>
<td>8.6&lt;sup&gt;‡&lt;/sup&gt; (7.4 to 9.9)</td>
<td>0&lt;sup&gt;‡&lt;/sup&gt; (0 to 66.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grover, 1995**&lt;sup&gt;‡&lt;/sup&gt; Fair Australia</td>
<td>2550 51 years</td>
<td>1 year</td>
<td>1 (0.04)</td>
<td>0</td>
<td>1</td>
<td>40</td>
<td>2509</td>
<td>0 (0 to 85.3)</td>
<td>98.4 (97.9 to 98.9)</td>
<td>0 (0 to 6.0)</td>
<td>100.0 (99.8 to 100.0)</td>
<td>1.6 (1.1 to 2.1)</td>
<td>100.0 (14.7 to 100.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacobs, 1988**&lt;sup&gt;‡&lt;/sup&gt; Fair UK</td>
<td>1010 54 years</td>
<td>1 year</td>
<td>1 (0.1)</td>
<td>1</td>
<td>0</td>
<td>27</td>
<td>982</td>
<td>100.0 (14.7 to 100.0)</td>
<td>97.3 (96.2 to 98.2)</td>
<td>3.6 (0.4 to 15.5)</td>
<td>100.0 (99.7 to 100.0)</td>
<td>2.7 (1.8 to 3.8)</td>
<td>0 (0 to 85.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Length of followup for ovarian cancer diagnosis.

<sup>†</sup> Calculated, not study reported.

<sup>‡</sup> Adonakis et al**<sup>‡</sup> treated ambiguous pelvic examinations as negative screens, whereas we treated the ambiguous examinations as positive screens. The author-reported sensitivity was 97.2%, specificity 66.7%, and PPV 3.4%.

** Abbreviations:** CA-125=cancer antigen 125; CI=confidence interval; FN=false negative; FP=false positive; N=number; n=number; NPV=negative predictive value; PLCO=Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; PPV=positive predictive value; TN=true negative; TP=true positive.
Table 7. Diagnostic Procedures Within 1 Year of a Positive Palpation Examination in Women Without an Ovarian Cancer Diagnosis: PLCO Trial Only

<table>
<thead>
<tr>
<th>Diagnostic procedure</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional CA-125†</td>
<td>10 (4.1)</td>
<td>26 (5.5)</td>
</tr>
<tr>
<td>Additional TVU§</td>
<td>47 (19.4)</td>
<td>87 (18.3)</td>
</tr>
<tr>
<td>Abdominal CT</td>
<td>7 (2.9)</td>
<td>11 (2.3)</td>
</tr>
<tr>
<td>Surgery</td>
<td>31 (12.8)</td>
<td>53 (11.2)</td>
</tr>
<tr>
<td>Any complication††</td>
<td>4 (1.7)</td>
<td>5 (1.0)</td>
</tr>
<tr>
<td>Surgical complication</td>
<td>1 (0.4)</td>
<td>1 (0.2)</td>
</tr>
</tbody>
</table>

* Due to rolling recruitment and early termination of the palpation component in the screening intervention arm of the trial, the number of palpation screening visits women completed was variable. On average, women received 2.4 palpation screening visits (28.0% received 1, 24.8% received 2, 24.4% received 3, and 22.8% received 4).
† Data from a 2016 personal communication with Dr. Paul Pinsky, National Cancer Institute.
‡ Data from Doroudi et al.21
§ All women received CA-125 and TVU as part of the PLCO trial protocol for ovarian cancer screening and these results were available to each woman’s provider. These two diagnostic procedures in the table indicate additional CA-125 and TVU were conducted with a woman’s provider within 1 year.
|| Any complication includes: surgical, infection, cardiovascular, pulmonary or other complication.

**Abbreviations:** CA-125=cancer antigen 125; n=number; PLCO=Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; TVU=transvaginal ultrasound.
### Table 8. Study Characteristics, Infectious Diseases

<table>
<thead>
<tr>
<th>Author, Year Quality</th>
<th>Country Recruitment Setting</th>
<th>Study Aim</th>
<th>Inclusion/Exclusion Criteria</th>
<th>N Screened</th>
<th>Mean Age</th>
<th>Prevalence of Symptoms</th>
<th>Screening Test</th>
<th>Reference Standard Description</th>
<th>Yield of Disease, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial Vaginosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gutman, 2005 Fair</td>
<td>US Hospital</td>
<td>Determine whether current clinical criteria for diagnosing BV can be simplified by using 2 clinical criteria rather than the standard (3 of 4 Amsel’s criteria)</td>
<td>Inclusion: any woman undergoing a speculum exam, Exclusion: large amount of vaginal bleeding on exam</td>
<td>269</td>
<td>24.1</td>
<td>Any symptoms: 32.7%</td>
<td>Thin, homogeneous vaginal discharge</td>
<td>Nugent’s criteria: a score of ≥7 defined a diagnosis of BV*</td>
<td>104/269 (38.7)</td>
</tr>
<tr>
<td>Eschenbach, 1988 Fair</td>
<td>US STI clinic</td>
<td>Compare accuracy of Gram stain criteria for BV with composite clinical criteria for diagnosing BV</td>
<td>Inclusion: age 16–50 years, English speaking, Exclusion: pregnant, used oral antibiotics or vaginal medication in previous 14 days, hysterectomized, severely mentally or physically incapacitated, <em>Trichomonas vaginalis</em> (by culture), no evaluable Gram stain</td>
<td>661</td>
<td>NR</td>
<td>Any pelvic or abdominal symptom as a chief complaint: 59%</td>
<td>Standardized pelvic exam, with attention to appearance of vulva, vagina, and cervix (erythema, friability of cervix, color of cervical mucus), characteristics of vaginal discharge (amount, color, other characteristics), and tenderness (cervical, uterine, adnexal)</td>
<td>pH of vaginal contents, clue and epithelial cells present on microscopy, fishy amine-like odor</td>
<td>311/661 (47.0)</td>
</tr>
</tbody>
</table>
Table 8. Study Characteristics, Infectious Diseases

<table>
<thead>
<tr>
<th>Author, Year Quality</th>
<th>Country Recruitment Setting</th>
<th>Study Aim</th>
<th>Inclusion/Exclusion Criteria</th>
<th>N Screened</th>
<th>Mean Age</th>
<th>Prevalence of Symptoms</th>
<th>Screening Test</th>
<th>Reference Standard Description</th>
<th>Yield of Disease, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genital Herpes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koutsky, 1992† Fair</td>
<td>US STI clinic</td>
<td>Assess relative merits of different approaches to detecting genital HSV infection, including the approach of clinical exam and viral isolation</td>
<td>Inclusion: age 16–50 years; English-speaking</td>
<td>779</td>
<td>24 years</td>
<td>22% of women with evidence of herpes presented symptomatically</td>
<td>Genital exam, looking for vulva ulcereations and tender inguinal nodes</td>
<td>For HSV isolation: collection of urine, specimens from cervix and anal canal, swabs from external genital lesions; serum Western blot for antibodies</td>
<td>372/779 (47.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70% white</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trichomoniasis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolner-Hansson, 1989† Fair</td>
<td>US STI clinic</td>
<td>Identify relationships of specific genital microbial pathogens to clinical manifestations</td>
<td>Inclusion: ages 16–50 years, English-speaking</td>
<td>779</td>
<td>24 years</td>
<td>Yellow discharge: 23%</td>
<td>Standardized pelvic exam with specific attention to appearance of vulva, vagina, and cervix; abnormal results included colitis macularis (&quot;strawberry cervix&quot;), purulent discharge, frothy discharge, vulvar or vaginal erythema</td>
<td>Cultures from 2 vaginal specimens examined for growth of <em>Trichomonas vaginalis</em> (identified by characteristic morphology and motility in unstained wet mounts)</td>
<td>118/778 (15.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70% white</td>
<td>Abnormal vaginal odor: 36%</td>
<td>Vulvar itching: 51%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 0- to 10-point score describing numbers of Lactobacilli, Gardnerella, and curved Gram-negative bacilli in 100× microscopy field of Gram stain sample.
† These three studies include the same sample of women.

**Abbreviations:** BV=bacterial vaginosis; HSV=herpes simplex virus; N=number; n=number; STI=sexually transmitted infection.
<table>
<thead>
<tr>
<th>Author, Year Quality Country</th>
<th>N Mean Age Race/Ethnicity</th>
<th>Yield, n (%)</th>
<th>Screening Test</th>
<th>TP*</th>
<th>FN*</th>
<th>FP*</th>
<th>TN*</th>
<th>Sensitivity (95% CI)²</th>
<th>Specificity (95% CI)²</th>
<th>PPV (95% CI)²</th>
<th>NPV (95% CI)²</th>
<th>FPR (95% CI)²</th>
<th>FNR (95% CI)²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial Vaginosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gutman, 2005 Fair US</td>
<td>269 24.1 years 38% white</td>
<td>104 (38.7)</td>
<td>Thin, homogeneous discharge on pelvic exam</td>
<td>82</td>
<td>22</td>
<td>76</td>
<td>89</td>
<td>78.8 (70.3 to 85.8)</td>
<td>53.9 (46.3 to 61.4)</td>
<td>51.9 (44.1 to 59.6)</td>
<td>80.2 (72.0 to 86.8)</td>
<td>46.1 (38.6 to 53.7)</td>
<td>21.2 (14.2 to 29.7)</td>
</tr>
<tr>
<td>Eschenbach, 1988 Fair US†</td>
<td>661 NR NR (47.0)</td>
<td>311 (47.0)</td>
<td>Homogeneous discharge</td>
<td>184</td>
<td>82</td>
<td>9</td>
<td>309</td>
<td>69.2 (63.4 to 74.5)</td>
<td>97.2 (94.9 to 98.6)</td>
<td>95.3 (91.7 to 97.7)</td>
<td>79.0 (74.8 to 82.8)</td>
<td>2.8 (1.4 to 5.1)</td>
<td>30.8 (25.5 to 36.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Frothy discharge</td>
<td>6</td>
<td>260</td>
<td>0</td>
<td>318</td>
<td>2.3 (0.9 to 45.9)</td>
<td>100.0 (99.2 to 100.0)</td>
<td>100 (67.0 to 100.0)</td>
<td>55.0 (50.9 to 59.0)</td>
<td>0 (0 to 7.9)</td>
<td>97.7 (95.4 to 99.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Increased discharge</td>
<td>25</td>
<td>241</td>
<td>14</td>
<td>304</td>
<td>9.4 (6.3 to 13.3)</td>
<td>95.6 (92.9 to 97.5)</td>
<td>64.1 (48.5 to 77.7)</td>
<td>55.8 (51.6 to 59.9)</td>
<td>4.4 (2.5 to 7.1)</td>
<td>90.6 (86.7 to 93.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yellow discharge</td>
<td>83</td>
<td>179</td>
<td>53</td>
<td>246</td>
<td>31.7 (26.3 to 37.5)</td>
<td>82.3 (77.6 to 86.3)</td>
<td>61.0 (52.7 to 68.9)</td>
<td>57.9 (53.1 to 62.5)</td>
<td>17.7 (13.7 to 22.4)</td>
<td>68.3 (62.5 to 73.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ectopy (any)</td>
<td>155</td>
<td>151</td>
<td>176</td>
<td>165</td>
<td>50.7 (45.1 to 56.2)</td>
<td>48.4 (43.1 to 53.7)</td>
<td>46.8 (41.5 to 52.2)</td>
<td>52.2 (46.7 to 57.7)</td>
<td>51.6 (46.3 to 56.9)</td>
<td>49.3 (43.8 to 54.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ectopy (50%)</td>
<td>20</td>
<td>286</td>
<td>28</td>
<td>313</td>
<td>6.5 (4.2 to 9.7)</td>
<td>91.8 (88.5 to 94.4)</td>
<td>41.7 (28.5 to 55.8)</td>
<td>52.3 (48.3 to 56.2)</td>
<td>8.2 (5.6 to 11.5)</td>
<td>93.5 (90.3 to 95.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adnexal tenderness</td>
<td>11</td>
<td>282</td>
<td>1</td>
<td>331</td>
<td>3.8 (2.0 to 6.4)</td>
<td>99.7 (98.6 to 100.0)</td>
<td>91.7 (67.2 to 99.1)</td>
<td>54.0 (50.0 to 57.9)</td>
<td>0.3 (0.03 to 1.4)</td>
<td>96.2 (93.6 to 98.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Uterine tenderness</td>
<td>11</td>
<td>297</td>
<td>5</td>
<td>343</td>
<td>3.6 (1.9 to 6.1)</td>
<td>98.6 (96.9 to 99.4)</td>
<td>68.8 (44.4 to 86.9)</td>
<td>53.6 (49.7 to 57.4)</td>
<td>1.4 (0.6 to 3.1)</td>
<td>96.4 (93.9 to 98.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cervical motion tenderness</td>
<td>8</td>
<td>301</td>
<td>2</td>
<td>346</td>
<td>2.6 (1.2 to 4.8)</td>
<td>99.4 (98.2 to 99.9)</td>
<td>80.0 (49.7 to 95.6)</td>
<td>53.5 (49.6 to 57.3)</td>
<td>0.6 (0.1 to 1.8)</td>
<td>97.4 (95.2 to 98.8)</td>
</tr>
</tbody>
</table>
Table 9. Summary of Diagnostic Accuracy of Pelvic Examination for Infectious Disease

<table>
<thead>
<tr>
<th>Author, Year Quality</th>
<th>Mean Age</th>
<th>Race/ Ethnicity</th>
<th>Screening Test</th>
<th>N</th>
<th>Yield, n (%)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)†</th>
<th>NPV (95% CI)†</th>
<th>FPR (95% CI)</th>
<th>FNR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genital Herpes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koutsky, 1992†‡§</td>
<td>779</td>
<td>24 years</td>
<td>Vulvar ulcerations</td>
<td>73</td>
<td>299 10</td>
<td>19.6 (15.8 to 23.9)</td>
<td>97.5 (95.7 to 98.7)</td>
<td>88.0 (79.7 to 93.6)</td>
<td>57.0 (53.3 to 60.7)</td>
<td>2.5 (1.3 to 4.3)</td>
<td>80.4 (76.1 to 84.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70% white</td>
<td>Tender inguinal nodes</td>
<td>53</td>
<td>319 12</td>
<td>14.2 (11.0 to 18.1)</td>
<td>97.1 (95.1 to 98.4)</td>
<td>81.5 (70.8 to 89.5)</td>
<td>55.3 (51.7 to 58.9)</td>
<td>2.9 (1.6 to 4.9)</td>
<td>85.8 (81.9 to 89.0)</td>
</tr>
<tr>
<td><strong>Trichomoniiasis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolner-Hansson, 1989§</td>
<td>779</td>
<td>24 years</td>
<td>Colpitis macularis</td>
<td>2</td>
<td>116 0</td>
<td>1.7 (0.4 to 5.3)</td>
<td>100.0 (99.6 to 100.0)</td>
<td>100.0 (33.3 to 100.0)</td>
<td>85.1 (82.4 to 87.4)</td>
<td>0 (0 to 0.4)</td>
<td>98.3 (94.7 to 99.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70% white</td>
<td>Purulent discharge</td>
<td>58</td>
<td>40 134</td>
<td>59.2 (49.3 to 68.5)</td>
<td>76.1 (72.4 to 79.5)</td>
<td>30.2 (24.0 to 37.0)</td>
<td>91.4 (88.6 to 93.7)</td>
<td>23.9 (20.5 to 27.6)</td>
<td>40.8 (31.5 to 50.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Frothy discharge</td>
<td>8</td>
<td>92 5</td>
<td>8.0 (3.8 to 14.5)</td>
<td>99.1 (98.1 to 99.7)</td>
<td>61.5 (35.0 to 83.5)</td>
<td>86.3 (83.5 to 88.7)</td>
<td>0.9 (0.3 to 1.9)</td>
<td>92.0 (85.5 to 96.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vulvar erythema</td>
<td>44</td>
<td>74 185</td>
<td>37.3 (29.0 to 46.2)</td>
<td>72.0 (68.5 to 75.3)</td>
<td>19.2 (14.5 to 24.7)</td>
<td>86.5 (83.5 to 89.2)</td>
<td>28.0 (24.7 to 31.5)</td>
<td>62.7 (53.8 to 71.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vaginal erythema</td>
<td>23</td>
<td>95 46</td>
<td>19.5 (13.1 to 27.3)</td>
<td>93.0 (90.9 to 94.8)</td>
<td>33.3 (23.1 to 45.0)</td>
<td>86.6 (84.0 to 89.0)</td>
<td>7.0 (5.2 to 9.1)</td>
<td>80.5 (72.7 to 86.9)</td>
</tr>
</tbody>
</table>

* Note that these numbers do not always add up to the number of people screened.
† Calculated; not reported in the study.
‡ These three studies include the same sample of women.
§ STI clinic; population at higher risk and more symptomatic than average.

**Abbreviations:** BV=bacterial vaginosis; CI=confidence interval; FN=false negative; FP=false positive; HSV=herpes simplex virus; N=number; n=number; NPV=negative predictive value; NR=not reported; PPV=positive predictive value; TN=true negative; TP=true positive.
### Table 10. Summary of Evidence by Key Question and Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th># Studies (k)</th>
<th>Sample Size (n)</th>
<th>Summary of Findings*</th>
<th>Body of Evidence Limitations†</th>
<th>Quality</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ1. Direct screening effectiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>No evidence</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KQ2. Screening accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>k=4</td>
<td>n=26,432</td>
<td>Sensitivity was reported as 100% in 2 studies where 1 or 2 ovarian cancer cases were palpable on pelvic exam; sensitivity was 0% in 1 study where the single case of ovarian cancer was not detected on pelvic exam. The large PLCO screening trial reported a sensitivity of 4.3% at 1 year. Specificity ranged from 91% to nearly 99%.</td>
<td>Rare ovarian cancer events; accuracy estimates had wide confidence intervals due to the very low event rate.</td>
<td>Fair</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td>n=26,432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 prospective diagnostic accuracy studies and 1 RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial vaginosis</td>
<td>k=2</td>
<td>n=930</td>
<td>In 1 study, sensitivity and specificity of thin, homogeneous discharge was 79% and 54%, respectively; the second study reported these values as 69% and 97%, respectively.</td>
<td>No screening studies conducted solely in asymptomatic primary care populations; the studies had large proportions of symptomatic patients and substantial clinical heterogeneity (populations, personnel performing index test, description of results of index tests, reference standards) and statistical heterogeneity (disparate accuracy results).</td>
<td>Fair</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>n=930</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prospective diagnostic accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genital herpes</td>
<td>k=1</td>
<td>n=779</td>
<td>Pelvic exam finding of vulvar ulcerations had a sensitivity of 20% and specificity of 98% in detecting genital herpes simplex virus at any stage.</td>
<td>No screening studies conducted solely in asymptomatic primary care populations; 78% of women with any genital herpes had asymptomatic shedding or latent disease which could not be detected by pelvic exam.</td>
<td>Fair</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>n=779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prospective diagnostic accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trichomoniasis</td>
<td>k=1</td>
<td>n=779</td>
<td>For colpitis macularis, sensitivity was 2% and specificity 100%; for other individual clinical findings, sensitivity ranged from 8% to 59% and specificity from 72% to 99%.</td>
<td>No screening studies conducted solely in asymptomatic primary care populations; data reporting did not provide sufficient information to calculate the sensitivity and specificity for the presence of one or more abnormal finding on pelvic exam.</td>
<td>Fair</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>n=779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prospective diagnostic accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td># Studies (k)</td>
<td>Sample Size (n)</td>
<td>Design</td>
<td>Summary of Findings*</td>
<td>Body of Evidence Limitations†</td>
<td>Quality</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>----------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Pelvic examination</td>
<td>k=1</td>
<td>n=150</td>
<td>Prospective cohort</td>
<td>Urinary tract infection was diagnosed clinically (without culture) for 1 subject in the pelvic exam group; dysuria (11/63 vs. 6/87; p&lt;0.01) and urinary frequency (17/63 vs. 12/87; p&lt;0.01) were more common in the pelvic exam group during the 4-week followup.</td>
<td>High loss to followup rate, underpowered to detect differences in urinary tract infections, and between-group differences in intercourse and barrier use; cannot be used to make conclusions about causality between pelvic exams and urinary tract infections.</td>
<td>Poor to Fair</td>
</tr>
<tr>
<td>KQ3. Ovarian cancer</td>
<td>k=4</td>
<td>n=26,432</td>
<td>3 Prospective diagnostic accuracy studies and 1 RCT</td>
<td>Percentage of patients with positive pelvic exam results who underwent surgery ranged from 5 to 36, depending on the study design and management protocols. Single study reported complication rate of 1% at 5 years downstream from abnormal pelvic exam.</td>
<td>Limited number of studies, different ultrasound techniques and threshold positivity, and rare ovarian cancer events.</td>
<td>Fair</td>
</tr>
<tr>
<td>Bacterial vaginosis</td>
<td>k=2</td>
<td>n=930</td>
<td>Prospective diagnostic accuracy</td>
<td>In 1 study, false-positive and false-negative rates for a pelvic exam finding of thin, homogeneous discharge were 46% and 21%, respectively; in a second study, the rates were 3% and 31%, respectively.</td>
<td>High-risk, symptomatic population</td>
<td>Fair</td>
</tr>
<tr>
<td>Genital herpes</td>
<td>k=1</td>
<td>n=779</td>
<td>Prospective diagnostic accuracy</td>
<td>For the clinical finding of vulvar ulceration, the false-negative and false-positive rates were 80% and 2% for any stage of genital herpes; for clinical finding of tender lymphadenopathy, the rates were 86% and 3%, respectively, for any stage of genital herpes</td>
<td>High-risk, symptomatic population</td>
<td>Fair</td>
</tr>
<tr>
<td>Trichomoniasis</td>
<td>k=1</td>
<td>n=779</td>
<td>Prospective diagnostic accuracy</td>
<td>For clinical findings of colpitis macularis, purulent discharge, frothy discharge, vulvar erythema, and vaginal erythema, the false-negative rate ranged from 41% to 92% and the false-positive rate ranged from 0% to 28%; the degree of harm from false-positive results is expected to be minimal given that the diagnostic test is benign and confirmation is conducted in the clinic without delay between a positive screening finding and confirmation.</td>
<td>High-risk, symptomatic population</td>
<td>Fair</td>
</tr>
</tbody>
</table>

* Includes consistency and precision.
† Includes reporting bias.

**Abbreviations**: k=number of studies; KQ=key question; n=number.
Appendix A. Detailed Methods

Literature Search Strategies for Primary Literature

Key:
/ = MeSH subject heading
$ = truncation
* = truncation
ab = word in abstract
ae = adverse effects
adj# = adjacent within x number of words
kw=keyword
mo=mortality
nm = name of substance
pt = publication type
ti = word in title

Cochrane Central Register of Controlled Trials (via Wiley)

#1 ((gynecolog* or gynaecolog* or genital* or pelvis or pelvic or uterus or uterine or ovary
or ovaries or ovarian or (fallopian next tube*) or cervix or cervical or vagina* or vulva* or
rectovaginal or bimanual or speculum or well-woman or "well woman" or prolapse*) near/5
(exam* or palpate* or palpation* or assess* or screen* or measur*)):ti
#2 (pelvic or bimanual or gynecolog* or gynaecolog* or rectovaginal):ab,kw next
exam*:ab,kw 319
#3 "cervical spine":ti,ab,kw
#4 (#1 or #2) not #3 in Trials

Ovid MEDLINE search strategy

1 Physical Examination/
2 Diagnostic Tests, Routine/
3 Digital Rectal Examination/
4 Palpation/
5 Mass screening/
6 Early detection of cancer/
7 1 or 2 or 3 or 4 or 5 or 6
8 Genitalia, Female/
9 Pelvis/
10 Adnexa Uteri/
11 Fallopian Tubes/
12 Uterus/
13 Cervix Uteri/
14 Ovaries/
15 Vagina/
16 Vulva/
17 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18 7 and 17
Appendix A. Detailed Methods

19  Gynecological Examination/
20   ((gynecologic$ or genital$ or pelvic or pelvic or uterus or uterine or ovary or ovaries or ovarian or fallopian tube$ or cervix or cervical or vagina$ or vulva$ or rectovaginal or bimanual or speculum or well-woman or prolapse$) adj5 (exam$ or palpate$ or palpation$ or assess$ or screen$)).ti.
21   pelvic exam$.ab.
22   bimanual exam$.ab.
23   gynecologic exam$.ab.
24   rectovaginal exam$.ab.
25   ((gynecologic$ or genital$ or pelvic or pelvic or uterus or uterine or ovary or ovaries or ovarian or fallopian tube$ or cervix or cervical or vagina$ or vulva$ or rectovaginal or bimanual or speculum or well-woman or prolapse$) adj5 (exam$ or palpate$ or palpation$ or assess$ or screen$)).ti,ab.
26   limit 25 to ("in data review" or in process or "pubmed not medline")
27   18 or 19 or 20 or 21 or 22 or 23 or 24 or 26
28   cervical spine.ti,ab.
29   27 not 28
30   Male/ not (Female/ and Male/)
31   Animal/ not (Animal/ and Human/)
32   29 not (30 or 31)
33   clinical trials as topic/ or controlled clinical trials as topic/ or randomized controlled trials as topic/ or meta-analysis as topic/
34   (clinical trial or controlled clinical trial or meta analysis or randomized controlled trial).pt.
35   Random$.ti,ab.
36   control groups/ or double-blind method/ or single-blind method/
37   clinical trial$.ti,ab.
38   controlled trial$.ti,ab.
39   meta analy$.ti,ab.
40   33 or 34 or 35 or 36 or 37 or 38 or 39
41   32 and 40
42   "Sensitivity and Specificity"/
43   "Predictive Value of Tests"/
44   ROC Curve/
45   False Negative Reactions/
46   False Positive Reactions/
47   Diagnostic Errors/
48   "Reproducibility of Results"/
49   Reference Values/
50   Reference Standards/
51   Observer Variation/
52   Receiver operat$.ti,ab.
53   ROC curve$.ti,ab.
54   sensitivit$.ti,ab.
55   specificit$.ti,ab.
56   predictive value.ti,ab.
57   accuracy.ti,ab.

Screening With Pelvic Examination

Kaiser Permanente Research Affiliates EPC
Appendix A. Detailed Methods

58 false positive$.ti,ab.
59 false negative$.ti,ab.
60 miss rate$.ti,ab.
61 error rate$.ti,ab.
62 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61
63 32 and 62
64 Mortality/
65 safety.ti,ab.
66 harm$.ti,ab.
67 mortality.ti,ab.
68 complication$.ti,ab.
69 (adverse adj2 (interaction$ or response$ or effect$ or event$ or reaction$ or outcome$)).ti,ab.
70 adverse effects.fs.
71 mortality.fs.
72 Pain/
73 Acute Pain/
74 pain$.ti,ab.
75 discomfort.ti,ab.
76 uncomfortable.ti,ab.
77 Stress, Psychological/
78 Anxiety/
79 ((psychological or mental) adj3 distress).ti,ab.
80 (anxiety or anxious).ti,ab.
81 embarrass$.ti,ab.
82 fear$.ti,ab.
83 Unnecessary Procedures/
84 (unnecessary or unneeded) adj5 (diagnostic or treat$ or workup or work up or procedure$)).ti,ab.
85 overtreat$.ti,ab.
86 overdiagnos$.ti,ab.
87 (false adj (assurance or reassurance)).ti,ab.
88 or/64-87
89 32 and 88
90 41 or 63 or 89
91 limit 90 to english language
Appendix A. Detailed Methods

PubMed search strategy (publisher-supplied)

#4 Search #3 AND publisher[sb] AND English[Language]
#3 Search #1 AND #2
#2 Search (exam*[title] OR palpate*[title] OR palpation*[title] OR assess*[title] OR screen*[title])
#1 Search (gynecolog*[title] OR gynaecolog*[title] OR genital*[title] OR pelvis*[title] OR pelvic*[title]
   OR bimanual*[title] OR speculum*[title] OR well-woman*[title] OR prolapse*[title])
Appendix A Figure 1. Literature Flow Diagram

# of unique records identified through database searching: 8657
# of unique records identified from Ebell or Bloomfield: 17
# of unique records identified through other sources: 4

# of records screened: 8678
# of records excluded: 8362

# of full-text articles assessed for eligibility: 316

Articles reviewed for Key Question 1: 36

Articles excluded for Key Question 1:
- Relevance: 0
- Design: 17
- Setting: 2
- Population: 0
- Outcomes: 0
- Intervention: 16
- Quality: 0
- Condition: 1

Articles included for Key Question 1: 0 (0 studies)

Articles reviewed for Key Question 2: 243

Articles excluded for Key Question 2:
- Relevance: 1
- Design: 41
- Setting: 17
- Population: 58
- Outcomes: 13
- Intervention: 95
- Quality: 2
- Condition: 6

Articles included for Key Question 2: 10* (8 studies)

Articles reviewed for Key Question 3: 114

Articles excluded for Key Question 3:
- Relevance: 6
- Design: 39
- Setting: 5
- Population: 5
- Outcomes: 14
- Intervention: 26
- Quality: 3
- Condition: 5

Articles included for Key Question 3: 11* (9 studies)

* 8 studies (in 10 articles) are included for both KQ2 and KQ3
### Appendix A Table 1. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Topic</th>
<th>Key Question</th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Populations</td>
<td>1–3</td>
<td>Age ≥18 years, general unselected females, asymptomatic, not pregnant, women with or without hysterectomy, postmenopausal women</td>
<td>Children and adolescents, age &lt;18 years, pregnant adolescents and women</td>
</tr>
<tr>
<td>Settings</td>
<td>1–3</td>
<td>Developed countries (“very high” development per the Human Development Index*), primary care outpatient setting (or similar settings applicable to primary care)</td>
<td>Settings not applicable to primary care</td>
</tr>
<tr>
<td>Conditions</td>
<td>1–3</td>
<td>Gynecologic cancers (e.g., ovarian, vulvar, vaginal, endometrial) and other gynecologic conditions (e.g., candidiasis, human papilloma virus, herpes simplex virus, trichomoniasis, bacterial vaginosis, atrophic vaginitis, fibroids, pelvic organ prolapse, pelvic floor dysfunction, pelvic inflammatory disease, cervical polyps, ovarian cysts, uterine fibroids, endometriosis) not listed in exclusion</td>
<td>Cervical cancer, gonorrhea, chlamydia, any nongynecologic cancer (e.g., colorectal cancer) or nongynecologic condition (e.g., hemorrhoids)</td>
</tr>
<tr>
<td>Interventions</td>
<td>1–3</td>
<td>Pelvic examination (external inspection, internal speculum examination, bimanual examination, rectovaginal examination) for screening; entire pelvic examination or components of pelvic examination</td>
<td>Pelvic examination for diagnosis, digital rectal exam, Papanicolaou test, human papillomavirus test</td>
</tr>
<tr>
<td>Comparisons</td>
<td>1</td>
<td>No pelvic examination</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Reference standard</td>
<td>No reference standard, or reference standard applied to a nonrandom subset</td>
</tr>
<tr>
<td>Outcomes</td>
<td>1</td>
<td>All-cause mortality, cancer-specific mortality or morbidity for included cancers, disease-specific morbidity for included conditions (may include abnormal bleeding, pelvic pain, incontinence, infertility), quality of life</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Sensitivity, specificity, likelihood ratios, positive predictive values, negative predictive values</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Unnecessary diagnostic workup, unnecessary treatment, physical pain/discomfort, barrier to obtaining hormonal contraception, psychological harms</td>
<td>Psychological measures that do not use validated scales of pain/discomfort or other harms</td>
</tr>
<tr>
<td>Study Designs</td>
<td>1</td>
<td>Systematic reviews, randomized controlled trials</td>
<td>Narrative reviews, editorials, case series, case reports, statistical models that extrapolate beyond direct clinical evidence, cross-sectional surveys with limited generalizability to current U.S. practice</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Systematic reviews of diagnostic accuracy studies, diagnostic accuracy studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Observational studies not listed in exclusion, randomized controlled trials, controlled clinical trials</td>
<td>Poor-quality studies</td>
</tr>
</tbody>
</table>

* Very high United Nations Human Development Index (or equivalent), 2014: Andorra, Argentina, Australia, Austria, Bahrain, Belgium, Brunei Darussalam, Canada, Chile, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Kuwait, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, New Zealand, Norway, Poland, Portugal, Qatar, Saudi Arabia, Singapore, Slovakia, Slovenia, Spain, Sweden, Switzerland, Taiwan, United Arab Emirates, United Kingdom, United States.
## Appendix A Table 2. Quality Assessment Criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Adapted Quality Criteria</th>
<th>USPSTF Ratings</th>
</tr>
</thead>
</table>
| Observational studies (e.g., prospective cohort studies), adapted from the Newcastle-Ottawa Scale (NOS)\(^2\) | • Was there representativeness of the exposed cohort?  
• Was the nonexposed cohort systematically selected?  
• Was the ascertainment of exposure reported?  
• Was the outcome of interest not present at baseline?  
• Were measurements equal, valid, and reliable?  
• Were outcome assessors blinded?  
• Was followup long enough for the outcome to occur?  
• Was there acceptable followup? | **Good:** Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis.  
**Fair:** Studies will be graded “fair” if any or all of the following problems occur, without the fatal flaws noted in the “poor” category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for.  
**Poor:** Studies will be graded “poor” if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. |
### Appendix A Table 2. Quality Assessment Criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Adapted Quality Criteria</th>
<th>USPSTF Ratings(^{27})</th>
</tr>
</thead>
</table>
| Diagnostic accuracy studies, adapted from the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) I\(^{25}\) and II\(^{26}\) instrument | - Could the selection of patients have introduced bias?  
  o Was the spectrum of patients representative of the patients who will receive the test in PC?  
  o Was the selection process clearly defined?  
  o Are there concerns that the included patients and setting do not match the review question?  
- Could the conduct or interpretation of the index test have introduced bias?  
  o Was the index test interpreted without knowledge of the reference standard results?  
  o If a threshold was used, was it prespecified?  
  o Are there concerns that the index test, its conduct, or its interpretation differ from the review question?  
- Could the conduct or interpretation of the reference standard have introduced bias?  
  o Is the reference standard likely to correctly classify the target condition?  
  o Was the reference standard interpreted without knowledge of the index test results?  
  o Are there concerns that the target condition as defined by the reference standard does not match the review question?  
  o Did the whole or partial selection of patients receive the reference standard?  
- Could the patient flow have introduced bias?  
  o Was there an appropriate interval between the index test and reference standard?  
  o Did all patients receive the same reference standard?  
  o Were all patients included in the analysis? | **Good:** Evaluates relevant available screening test; uses a credible reference standard; interprets reference standard independently of screening test; reliability of test assessed; has few or handles indeterminate results in a reasonable manner; includes large number (more than 100) broad-spectrum patients with and without disease.  
**Fair:** Evaluates relevant available screening test; uses reasonable although not best standard; interprets reference standard independent of screening test; moderate sample size (50 to 100 subjects) and a "medium" spectrum of patients.  
**Poor:** Has fatal flaw such as: Uses inappropriate reference standard; screening test improperly administered; biased ascertainment of reference standard; very small sample size or very narrow selected spectrum of patients. |

Note: All studies were classified as good, fair, or poor according to the USPSTF Procedure Manual.\(^{27}\)
Appendix B. Excluded Studies

<table>
<thead>
<tr>
<th>Reason for Exclusion</th>
<th>Study</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1. Irrelevant study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2. Not an included study design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2a. No use of reference standard (or reference standard applied to a nonrandom subset)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2b. Cross-sectional surveys with limited generalizability to current U.S. practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3. Not an included setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3a. Not a country with a very high United Nations Human Development Index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3b. Not generalizable to primary care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4. Not an included population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4a. High-risk or symptomatic patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5. No relevant outcomes, or incomplete outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5a. No additional relevant data (primary article included)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5b. Measure using an unvalidated scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6. Not an included intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E7. Poor-quality study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8. Not an included gynecologic condition (cervical cancer, gonorrhea, chlamydia)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Appendix B. Excluded Studies


Appendix B. Excluded Studies


63. Cook C. 'About as comfortable as a stranger putting their finger up your nose': speculation about the (extra)ordinary in gynaecological examinations. Cult Health Sex. 2011;13(7):767-80. PMID: 21656407. KQ3E2.


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


207. Partridge EE, Kreimer AR, Buys SS, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial: Results from 4 years of annual screening in a randomized trial. Gynecol Oncol. 2007;104(3 Suppl 1):S14-5, Abstract 0. PMID: None. KQ1E6, KQ2E6.


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


Appendix B. Excluded Studies


