A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer

Mark H. Ebell, MD, MS, MaryBeth B. Culp, MPH, Taylor J. Radke, MPH

Context: Ovarian cancer is common and has significant morbidity and mortality, partly because it is often diagnosed at a late stage. This study sought to determine the accuracy of individual symptoms and combinations of symptoms for the diagnosis of ovarian cancer.

Evidence acquisition: MEDLINE was searched, identifying 2,492 abstracts, reviewing 71 articles in full, and ultimately identifying 17 studies published between 2001 and 2014 that met the inclusion criteria. Data were abstracted by two researchers, and quality was assessed using the QUADAS-2 criteria adapted to the study question. Bivariate random effects meta-analysis was used where possible, and heterogeneity and threshold effects were explored using receiver operating characteristic curves. Data were analyzed in 2015.

Evidence synthesis: Most studies were at high risk of bias, primarily because of case-control design or differential verification bias. The highest positive likelihood ratios (LRs+) were found for presence of abdominal mass (LR+, 30.0); abdominal distension or increased girth (LR+, 16.0); abdominal or pelvic pain (LR+, 10.4); abdominal or pelvic bloating (LR+, 9.3); loss of appetite (LR+, 9.2); and a family history of ovarian cancer (LR+, 7.5). No symptoms were helpful at ruling out ovarian cancer when absent. The Ovarian Cancer Symptom Index was validated in five studies and (after excluding one outlier with different inclusion criteria) was 63% sensitive and 95% specific (LR+, 12.6; LR–, 0.39). Two other symptom scores had not been validated prospectively.

Conclusions: Several individual signs and symptoms significantly increase the likelihood of ovarian cancer when present. More work is needed to validate decision rules and develop new decision support tools integrating risk factors, symptoms, and possibly biomarkers to identify women at increased ovarian cancer risk.

(Am J Prev Med 2016;50(3):384–394)

Context

Ovarian cancer is a common cause of morbidity and mortality in women, with approximately 14,000 women dying in the U.S. each year from the disease. The lifetime incidence is 1.4% and lifetime mortality is 1.04% in the U.S.[1] Risk factors include increasing age, nulliparity, and early menarche or late menopause (all associated with an increasing number of ovulations), as well as the *BRCA1* or *BRCA2* mutations.[2,3] Unfortunately, because of the anatomic location of the ovaries deep in the pelvis, malignancies are often large and advanced at the time of diagnosis. Only 25% of women with ovarian cancer are diagnosed at Stage I or II, whereas 58% are stage III and 17% are stage IV. The latter stages have 10-year survival rates of 21% and less than 5%, respectively.[1]

Studies of screening in asymptomatic women, including the large Prostate Lung Colon Ovary RCT, have not found a benefit to screening for ovarian cancer using a combination of ultrasound and the CA-125 blood test.[4] The symptoms of ovarian cancer are often vague and ill defined and overlap with symptoms of much more common disorders such as dyspepsia, irritable bowel syndrome, menstruation, and menopause. This makes early diagnosis a challenge as well.

The literature searches of previous systematic reviews of symptoms of the clinical diagnosis of ovarian cancer, published in 2005 and 2009, are now at least 7 years old.[5,6] Bankhead and colleagues[5] performed a careful

From the Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, Georgia

Address correspondence to: Mark H. Ebell, MD, MS, Professor of Epidemiology, 233 Miller Hall, University of Georgia Health Sciences Campus, Athens GA 30602. E-mail: ebell@uga.edu.

0749-3797/$36.00

http://dx.doi.org/10.1016/j.amepre.2015.09.023

literature search, but included many case series and reported ORs rather than sensitivity and specificity. A second systematic review[6] also performed a thorough search but was a qualitative review of the literature and did not abstract quantitative data regarding the accuracy of symptoms. Neither systematic review used current standards for the assessment of diagnostic study quality,[7] the synthesis of data such as bivariate meta-analysis, or the reporting of CIs and predictive intervals around summary estimates.

A previous study[8] by this research team performed a systematic review of the bimanual pelvic exam for the diagnosis of ovarian cancer, finding that it lacked sensitivity. In the current systematic review, the results of a meta-analysis of individual symptoms and combinations of symptoms for the diagnosis of ovarian cancer are reported.

Evidence Acquisition

This study followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for presentation of the results of a systematic review. The PRISMA flow diagram is shown in Appendix A (available online), and the PRISMA checklist is shown in Appendix B (available online).

Inclusion Criteria

Cohort or case-control studies that provided data regarding symptoms of ovarian cancer, combinations of symptoms, or elements of the medical history present prior to the diagnosis of ovarian cancer were included. All studies had to provide enough information to calculate sensitivity and specificity for at least one symptom or symptom score. Studies in children, case reports and case series, studies with <20 patients, studies using something other than histopathology to confirm the diagnosis of ovarian cancer, and studies of patients with known ovarian cancer were excluded.

Search Strategy

The following search strategy was used in PubMed:

(*ovarian cancer*[tiab] OR *ovarian neoplasms*[MeSH Terms]) AND (*symptom*[tiab] OR *symptoms*[tiab] OR *abdominal pain*[tiab] OR *bloating*[tiab])

The search strategy was deliberately broad. The bibliographies of two previous systematic reviews[5,6] were also reviewed to identify additional studies that may have been missed by the initial PubMed search.

Data Abstraction

Each abstract identified by the initial PubMed search was reviewed by two of the investigators, and any article identified as possibly meeting the inclusion criteria was reviewed in full by all three investigators. Decisions regarding inclusion of a study, subsequent data abstraction, and study quality assessment were always performed in parallel by two investigators, with any discrepancies discussed and resolved following discussion or after consultation with the third investigator. Data regarding study characteristics, study quality, and information needed to calculate sensitivity and specificity were abstracted in parallel. Any discrepancies or conflicts were resolved by the first author.

Some studies reported data regarding accuracy for several time periods, such as 6 months, 1 year, or 2 years of follow-up. When a study included information from the same patients about symptoms during different time periods, the most recent data were used, as they are most likely to reflect the accuracy of this test in a clinical setting where a patient presents with vague or undifferentiated symptoms. Also, longer intervals from symptom evaluation to diagnosis may decrease sensitivity (as a symptom may not appear till later in the course of disease). For example, one study[9] asked about symptoms occurring within 6 months of diagnosis and also at any time before diagnosis; only the former data were used. Another[10] reported symptoms that occurred during the previous year (including pre-existing symptoms) as well as only those that began during the previous year; only the latter were used. These decisions were made to avoid duplication of data and to reflect clinical practice, which would generally value recent symptoms of new onset over older or pre-existing symptoms.

Symptoms judged to be evaluating the same thing (e.g., "increased abdominal girth" and "abdominal distension" or "abnormal vaginal bleeding," "irregular vaginal bleeding," and "postmenopausal bleeding") were combined.

Quality Assessment

The QUADAS-2 framework was adapted for this study.[7] That adaptation is shown in Appendix C (available online). For each of the 16 items, the definition was adapted to studies of the medical history for the diagnosis of ovarian cancer. Studies at low risk of bias in each of the four domains (patient selection, index text, reference standard, and flow and timing) were assessed to be at low risk of bias overall. Those with a high risk of bias in a single domain were assessed to be at moderate risk of bias, and those with a high risk of bias in two or more domains were assessed to be at high risk of bias overall.

Statistical Analysis

For each individual sign and symptom in each study, the sensitivity; specificity; likelihood ratios (LRs); and their associated CIs were calculated. If the value of any cell in the 2 X 2 table was zero, a continuity correction of 0.5 was added to each cell to avoid division by zero when calculating LRs. Bivariate meta-analysis used the "midas" procedure in Stata, version 13.1, which calculates summary estimates of sensitivity, specificity, LRs, and the area under the receiver operating characteristic (ROC) curve along with CIs for each value. If bivariate meta-analysis was not possible using this procedure owing to a small number of studies or unstable initial estimates, the "mada" procedure in R, version 3.1.1, was used to perform bivariate meta-analysis. This procedure calculates summary estimates for sensitivity and specificity along with their CIs, but does not calculate summary estimates for LRs. When using this procedure, the summary point estimates for sensitivity and specificity were used to calculate positive and negative LRs (LR+ and LR–, respectively). Where possible, results are reported separately for cohort and case-control studies.

Evaluation of heterogeneity is particularly challenging for diagnostic test studies. For case-control and especially cohort designs, there are often far fewer cases than controls, leading to less precise estimates and greater heterogeneity for sensitivity than specificity. Measures useful for meta-analysis of clinical trials such as the I^2 statistic[11] are not useful for diagnostic tests, because they fail to account for threshold effects.[12] That is, variation in the threshold used to define a positive test (either implicit or explicit) leads to joint variation in sensitivity and specificity along the ROC curve. Heterogeneity was therefore evaluated visually by inspections of plots of studies in ROC space. The area under the ROC curve was not calculated because of the small number of studies for most symptoms; this often led to clustering of a few studies in one corner of ROC space that made determination of this area unstable and unhelpful. CIs and prediction intervals (which describe the interval where one can expect to find the next data point measured) are shown for the summary estimates of sensitivity and specificity in ROC space for key variables and were created using the "mada" procedure in R. Analysis was performed in 2015.

Evidence Synthesis

Search Results

The initial search returned 2,336 articles; an updated search on February 20, 2015, identified 156 additional studies. A total of 71 studies met the initial screening and were reviewed in full. Two studies[13,14] used the same data set, but one reported individual symptoms and the other reported a novel symptom score, so both were included in the analysis. A pair of case-control studies[15,16] by the same authors were published in successive issues of the same journal, and appeared to have used largely the same data set. Both were excluded, as they did not describe how or when symptoms were evaluated. Two large cohort studies used administrative databases to identify women who had sought care for abdominal or pelvic symptoms during the 2 years before a diagnosis of ovarian cancer. One[17] used the QRESEARCH database and the other[18] used The Health Improvement Network database. These results are presented separately from those of cohort studies that directly interviewed patients. The review of the bibliographies of previous systematic reviews identified no new studies that met the inclusion criteria. Overall, 17 studies[9,10,13,14,17–29] published between 2001 and 2014 met the inclusion and exclusion criteria and were included in the systematic review. This process is summarized in the PRISMA flow diagram (Appendix A, available online).

Study Characteristics

The characteristics of included studies are shown in Table 1. All but two took place in the U.S. or United Kingdom. The case-control studies ranged in size from 200 women[19] to 4,554 women.[20] Three of the four cohort studies were larger than any case-control study, and the two[17,18] using administrative data had between 608,682 and 1,054,818 women in them. Most studies enrolled women who were at least 40 years old; the exception was a single case-control study[19] that only included women aged 15–35 years.

Study Quality

The assessment of study quality is summarized in the first column of Table 1 and in Appendix C (available online). Only one study[21] was judged to be at low risk of bias. Three cohort studies and two case-control studies were judged to be at moderate risk of bias. One[22] of the cohort studies only included women who had an initial abnormal screening test and the other two[17,18] used administrative data rather than direct patient interview. All but two case-control studies were assessed to be at high risk of bias, primarily because of patient selection and partial verification bias.

Accuracy of Individual Symptoms and Elements From the Medical History

The test characteristics for symptoms reported by at least three studies are shown in Table 2. Symptoms that had the highest LRs+, and therefore most increased the likelihood of ovarian cancer when present, were abdominal mass (LR+, 30.0); abdominal distension or increased girth (LR+, 16.0); abdominal or pelvic pain (LR+, 10.4); abdominal or pelvic bloating (LR+, 9.3); loss of appetite (LR+, 9.2); and a family history of ovarian cancer (LR+, 7.5). These symptoms generally had excellent specificity but only modest sensitivity for ovarian cancer.

Symptoms that were moderately helpful, with LRs+ between 3 and 7, included diarrhea (LR+, 6.2); isolated abdominal pain (LR+, 4.3); weight loss (LR+, 4.3); change in bowel habits (LR+, 4.2); constipation (LR+, 4.0); urinary frequency or urgency (LR+, 4.0); dyspepsia (LR+, 3.3); and abnormal vaginal bleeding (LR+, 3.6). Age >50 years, rectal bleeding, nausea, fatigue, back or flank pain, and the presence of any gastrointestinal or urinary tract symptom did not increase the likelihood of ovarian cancer in a clinically significant manner.

The absence of any symptom did not reduce the likelihood of ovarian cancer in a clinically significant way; the symptoms with the lowest LRs– were abdominal or pelvic pain (LR–, 0.55); abdominal pain (LR–, 0.57); and abdominal or pelvic bloating (LR–, 0.6). Oral contraceptive use, as found in previous studies (2), was protective, with an LR of 0.68.

The accuracy of key individual symptoms is summarized in ROC space in Figures 1A–1E. In general, there was more variation in estimates of sensitivity than specificity, owing to the smaller number of cases than

Table 1. Characteristics of Included Studies

Study design/author (year) risk of bias	Patient population	Age (years)	Ref	Country	Years of study
Cohort					
Andersen (2014)[21] Low	Women from primary care clinics at the University of Washington (*n*=5,012), of whom 8 were ultimately diagnosed with ovarian cancer	Range=40 or older	Symptomatic: CA-125, ultrasound, and follow-up as needed; asymptomatic: not in SEER database at least 1 year later	U.S.	2008–2011
Pavlik (2009)[22] Medium	Women (*n*=272) enrolled in the University of Kentucky Ovarian Cancer Screening Project who had an abnormal screening TVUS and underwent surgery, of whom 30 were diagnosed with ovarian cancer	M=58, range=32–89	Surgery with histologic confirmation	U.S.	1987–2008
Cohort studies administrative data					
Collins (2012)[18] Medium	Women (*n*=1,054,818) whose medical records were part of a large primary care registry (The Health Improvement Network), including 735 diagnosed with ovarian cancer	Median=49, IQ range=38–63	Cases: surgery with histologic confirmation; controls: no diagnosis of cancer in EHR at 2 years.	UK	2000–2008
Hippisley-Cox (2011)[17] Medium	Primary care patients (*n*=608,682) from a large primary care registry (QResearch), including 538 diagnosed with ovarian cancer	M=51, SD=15.4	Cases: surgery with histologic confirmation; controls: no cancer in EHR at 2 years	UK	2000–2010
Case-control					
Andersen (2008)[25] Medium	Cases (*n*=75): women undergoing surgery for pelvic masses who were later diagnosed with ovarian cancer; controls (*n*=254): women enrolled in a high-risk screening program	NR	Cases: CA125, surgery with histologic confirmation; controls: CA125	U.S.	2008
Andersen (2010)[24] Medium	Cases (*n*=74): women with ovarian cancer; controls (*n*=137): women at high risk for ovarian cancer participating in screening program	NR	Cases: CA125, HE4, surgery with histologic confirmation; controls: CA125, HE4, ultrasound	U.S.	2010
Behtash (2008)[19] High	Cases (*n*=100): Iranian women aged 15–35 years with ovarian cancer; controls (*n*=100): females in primary healthcare units for preventive care, matched by age and site	M=25, range=15–35	Cases: surgery with histologic confirmation; controls: no ref	Iran	1995–2005
Friedman (2005)[9] High	Cases (*n*=102): women with ovarian cancer who subscribed to Kaiser Permanente Medical Care Program; controls (*n*=102): women subscribers matched on age and site of care	M=58, range=29–87	Cases: surgery with histologic confirmation; controls: no ref	U.S.	2001
Goff (2004)[26] High	Cases (*n*=44): women with surgically confirmed ovarian cancer; controls (*n*=1,709): women visiting 2 primary care clinics at the University of Washington	Cases: M=55; controls: median=45, range=15–90	Cases: surgery with histologic confirmation; controls: no ref	U.S.	2001–2002

(continued on next page)

Table 1. Characteristics of Included Studies *(continued)*

Study design/author (year) risk of bias	Patient population	Age (years)	Ref	Country	Years of study
Grewal (2013)[14] High	Cases (n=212): women aged ≥40 years with primary ovarian cancer diagnosed between 2000 and 2007 in general practices; controls (n=1,060): 5 age-matched women from the same practice for each case	Range=40–80	Cases: surgery with histologic confirmation; controls: no ref	UK	2000–2007
Hamilton (2009)[13] High	Cases (n=212): women aged ≥40 years with primary ovarian cancer; controls (n=1,060): randomly selected women matched by age and general practice	Women aged ≥40 years	Cases: surgery with histologic confirmation; controls: no ref	UK	2000–2007
Lim (2012)[23] High	Cases (n=194): patients with primary ovarian cancer recruited before definitive diagnosis/treatment; controls (n=268): age-matched women who attended an ovarian cancer screening clinic	M=65, range=50–79	Cases: surgery with histologic confirmation; controls: no ref	UK	2006–2008
Lurie (2009)[29] High	Cases (n=432): Hawaiian women aged 19–88 years, who were diagnosed with histologically confirmed primary invasive ovarian carcinoma; controls (n=491): randomly selected women aged ≥18 years who were residents in Hawaii for a minimum of a year, had no prior history of ovarian cancer, and had at least one intact ovary	Cases: M=56; controls: M=57	Cases: surgery with histologic confirmation; controls: no ref	U.S.	1993–2007
Olson (2001)[28] High	Cases (n=168): women aged ≥18 years with ovarian cancer; controls (n=251): healthy women from the community recruited by random-digit dialing	Women aged ≥18 years	Cases: surgery with histologic confirmation; controls: no ref	U.S.	1994–1997
Riman (2002)[20] High	Cases (n=655): women diagnosed with an incident, histologically confirmed endothelial ovarian cancer; controls (n=3,899): women randomly selected from a continuously updated population register covering all residents of Sweden and sampled simultaneously with the cases	Cases: M=62, range=50–74; controls: M=63, range=50–74	Cases: surgery with histologic confirmation; controls: no ref	Sweden	1993–1995
Rossing (2010)[10] High	Cases (n=812): patients with epithelial ovarian cancer; controls (n=1,313): population-based control subjects matched by age and urban versus rural	Range=35–74	Cases: surgery with histologic confirmation; controls: no ref	U.S.	2002–2005
Vine (2003)[27] High	Cases (n=267): women with ovarian cancer identified by area hospitals on a rapid case ascertainment basis; controls (n=317): population-based women with at least one intact ovary identified using random-digit dialing and phone lists	Cases: M=54, range=20–74; controls: M=55.1, range=20–74	Cases: surgery with histologic confirmation; controls: no ref	U.S.	1999–2001

Note: See Appendix C (available online) for detailed information regarding risk of bias assessment.
EHR, electronic health record; NR, not reported; SEER, Surveillance, Epidemiology, and End Results; TVUS, transvaginal ultrasound.

Table 2. Accuracy of Individual Signs and Symptoms for the Diagnosis of Ovarian Cancer Based on Bivariate Meta-Analysis

Finding/study	Sensitivity	Specificity	LR+	LR−
Abdominal distension or increased girth				
Case-control studies[a]	0.46 (0.34, 0.58)	0.97 (0.90, 0.99)	13.3 (5.60, 31.4)	0.56 (0.46, 0.69)
Case-control and cohort studies[b]	0.32 (0.16, 0.53)	0.98 (0.94, 0.99)	16.0	0.68
Abdominal mass				
Case-control studies[b]	0.15 (0.04, 0.39)	0.995 (0.98, 1.0)	30.0	0.85
Abdominal or pelvic bloating				
Case-control studies[a]	0.43 (0.28, 0.60)	0.95 (0.89, 0.98)	9.31 (4.92, 17.61)	0.60 (0.46, 0.77)
Abdominal pain				
Case-control studies[a]	0.50 (0.40, 0.60)	0.89 (0.73, 0.96)	4.65 (2.04, 10.60)	0.56 (0.51, 0.62)
Case-control and cohort studies[a]	0.50 (0.44, 0.56)	0.88 (0.79, 0.94)	4.26 (2.62, 6.91)	0.57 (0.54, 0.60)
Abdominal or pelvic pain				
Case-control studies[b]	0.47 (0.27, 0.67)	0.95 (0.9, 0.98)	10.4	0.55
Abnormal vaginal bleeding				
Case-control studies[a]	0.13 (0.10, 0.18)	0.96 (0.91, 0.98)	3.00 (1.53, 5.88)	0.91 (0.87, 0.95)
Case-control and cohort studies[a]	0.13 (0.10, 0.17)	0.96 (0.93, 0.98)	3.60 (2.10, 6.30)	0.90 (0.88, 0.93)
Age ≥50 years				
Case-control studies[a]	0.73 (0.62, 0.82)	0.40 (0.24, 0.58)	1.20 (1.00, 1.50)	0.68 (0.50, 0.92)
Any gastrointestinal symptom				
Case-control studies[a]			1.25 (0.87, 1.78)	0.90 (0.73, 1.11)
Appetite loss				
Case-control studies[a]	0.16 (0.11, 0.23)	0.99 (0.96, 1.00)	21.1 (5.50, 81.3)	0.85 (0.79, 0.90)
Case-control and cohort studies[a]	0.07 (0.03, 0.17)	0.99 (0.99, 1.00)	9.20 (5.64, 15.0)	0.93 (0.87, 1.00)
Back or flank pain				
Case-control studies[a]	0.22 (0.13, 0.35)	0.84 (0.49, 0.97)	1.40 (0.38, 5.14)	0.89 (0.82, 0.97)
Change in bowel habits or IBS				
Case-control studies[b]	0.08 (0.02, 0.30)	0.98 (0.89, 1.0)	4.21	0.92
Constipation				
Case-control studies[a]	0.19 (0.11, 0.30)	0.95 (0.88, 0.98)	3.97 (2.28, 6.89)	0.85 (0.79, 0.91)
Diarrhea				
Case-control studies[a]	0.11 (0.04, 0.27)	0.98 (0.83, 1.00)	6.24 (1.19, 32.70)	0.90 (0.83, 0.98)
Family history ovarian cancer				
Case-control and cohort studies[a]	0.04 (0.03, 0.06)	0.99 (0.98, 1.00)	7.51 (3.34, 16.9)	0.97 (0.95, 0.98)
Fatigue				
Case-control studies[a]	0.35 (0.21, 0.53)	0.88 (0.71, 0.96)	2.55 (1.55, 4.19)	0.78 (0.68, 0.89)
Indigestion or dyspepsia				
Case-control studies[a]	0.30 (0.18, 0.46)	0.91 (0.73, 0.97)	3.28 (1.21, 8.85)	0.77 (0.65, 0.91)

(continued on next page)

Table 2. Accuracy of Individual Signs and Symptoms for the Diagnosis of Ovarian Cancer Based on Bivariate Meta-Analysis *(continued)*

Finding/study	Sensitivity	Specificity	LR+	LR−
Nausea				
Case-control studies[a]	0.13 (0.10, 0.17)	0.93 (0.85, 0.97)	1.95 (1.00, 3.80)	0.93 (0.89, 0.97)
Oral contraceptive use				
Case-control studies[b]	0.39 (0.31, 0.48)	0.43 (0.24, 0.65)	0.68	1.48
Rectal bleeding				
Case-control and cohort studies[b]	0.04 (0.02, 0.09)	0.98 (0.97, 0.99)	2.0	0.96
Urinary frequency or urgency				
Case-control studies[b]	0.27 (0.15, 0.43)	0.93 (0.85, 0.97)	3.97	0.78
Urinary tract, any symptom				
Case-control studies[a]	0.33 (0.24, 0.44)	0.82 (0.72, 0.89)	1.87 (1.40, 2.50)	0.81 (0.74, 0.89)
Weight loss				
Case-control and cohort studies[a]	0.07 (0.04, 0.12)	0.98 (0.98, 0.99)	4.27 (2.90, 6.30)	0.95 (0.91, 0.98)

Note: Appendix D (available online) is a version of this table that includes individual study results.
[a]Bivariate meta-analysis using Stata.
[b]Bivariate meta-analysis using R. This calculates summary estimates for sensitivity and specificity, which were used to calculate point estimates for LR+ and LR−.
LR+, positive likelihood ratio; LR−, negative likelihood ratio.

controls, especially in cohort studies. The area under the ROC curves is generally not informative because the points are typically clustered in one corner of the curve, making extrapolation of the curve imprecise. Threshold effects, characterized by decreasing specificity as sensitivity increases, were apparent for abdominal distension or increased girth, abdominal or pelvic pain, and abdominal or pelvic bloating. These may be caused by differing thresholds (often implicit rather than explicit for symptoms) to define the presence or absence of a symptom.

Combinations of Symptoms

Three teams of researchers developed combinations of symptoms for the diagnosis of ovarian cancer, which were reported in included studies. These results are summarized in Table 3. Goff and colleagues[30] proposed a clinical decision rule (Ovarian Cancer Symptom Index) that was positive if a woman had any of six symptoms (abdominal or pelvic pain, increased abdominal size or bloating, and feeling full or difficulty eating) for <1 year and occurring >12 times per month. The summary estimates in two cohort and three case-control validation studies were 53% for sensitivity and 94% for specificity (LR+, 9.0; LR−, 0.5) (Table 3). The summary ROC curve (Figure 1F) shows that one study[22] was an outlier, with much lower sensitivity, although this may be because this study was limited to patients with an initial abnormal ultrasound screening test. Excluding that study, the summary estimates are 63% for sensitivity and 95% for specificity (LR+, 12.6; LR−, 0.39). Although the LR− for this score is lower than that for any individual symptom, the LR+ is very good and is similar to several of the individual symptoms discussed above.

Two other clinical decision rules were identified in the literature. Grewal et al.[14] developed a multivariate model using data from a case-control study, then simplified it into a point score and added a variable for age >50 years. Two points were assigned for bloating, urinary frequency, rectal bleeding, and postmenopausal bleeding; 3 points for loss of appetite or abdominal pain; 5 points for abdominal distension, and 1 point for age ≥ 50 years. In the derivation population, a cutoff of ≥ 4 points was 73% sensitive and 91% specific (LR+, 8.4; LR−, 0.3). Finally, Lim and colleagues[23] proposed two symptom indices, one with six items and the other with five. In each case, four items overlapped with the Ovarian Symptom Index. Though the scores proposed by Grewal et al.[14] and Lim and colleagues[23] are promising, neither has been prospectively validated.

Discussion

A total of 17 case-control or cohort studies reported the sensitivity and specificity of at least one symptom of

Table 3. Accuracy of Clinical Decision Rules for the Diagnosis of Ovarian Cancer

Study	Sensitivity	Specificity	LR+	LR−
Goff Ovarian Cancer Symptom Index				
Cohort studies				
Andersen (2014)[21]	0.50 (0.01, 0.99)	0.96 (0.95, 0.96)	12.50 (3.11, 50.35)	0.52 (0.13, 2.08)
Pavlik (2009)[22]	0.20 (0.08, 0.39)	0.91 (0.87, 0.95)	2.30 (1.01, 5.25)	0.88 (0.73, 1.05)
Case-control studies				
Andersen (2008)[25]	0.64 (0.52, 0.75)	0.88 (0.84, 0.92)	5.42 (3.72, 7.90)	0.41 (0.30, 0.55)
Andersen (2010)[24]	0.64 (0.52, 0.74)	0.88 (0.82, 0.93)	5.44 (3.33, 8.89)	0.41 (0.30, 0.56)
Lim (2012)[23]	0.62 (0.54, 0.69)	0.99 (0.97, 1.00)	55.2 (17.8, 171)	0.39 (0.32, 0.46)
Combined all studies[a]	0.53 (0.36, 0.70)	0.94 (0.88, 0.97)	9.00 (4.07, 19.90)	0.50 (0.34, 0.72)
Combined without Pavlik[b]	0.63 (0.58, 0.68)	0.95 (0.86, 0.98)	12.6	0.39
Grewal (2013)[14] symptom score	0.73 (0.66, 0.78)	0.91 (0.89, 0.93)	8.37 (6.77, 10.3)	0.3 (0.24, 0.37)
Lim (2012),[23] Index 1	0.83 (0.77, 0.88)	0.91 (0.87, 0.94)	9.2 (6.3, 13.6)	0.19 (0.14, 0.26)
Lim (2012),[23] Index 2	0.78 (0.72, 0.84)	0.94 (0.90, 0.97)	13.1 (8.1, 21.1)	0.23 (0.18, 0.31)

[a]Bivariate meta-analysis using Stata.
[b]Bivariate meta-analysis using R. This calculates summary estimates for sensitivity and specificity, which were used to calculate point estimates for LR+ and LR−.
LR+, positive likelihood ratio; LR−, negative likelihood ratio.

ovarian cancer. Almost all had significant methodologic limitations: a case-control design; reliance on physician diagnostic codes in administrative data sets; spectrum bias (when the studied population did not include a spectrum of symptoms and disease typical of that in usual clinical practice); and verification bias (different reference standards). A single study[21] was felt to be at low risk of bias, but it only reported data for a single symptom score (the Ovarian Cancer Symptom Index).

Symptoms commonly associated with ovarian cancer such as abdominal distension, abdominal or pelvic bloating, abdominal mass, loss of appetite, and abdominal or pelvic pain significantly increased the likelihood of ovarian cancer when present, with LRs+ ranging from 9.3 to 30.0. Symptoms that were somewhat helpful included diarrhea, isolated abdominal pain, weight loss, change in bowel habits, constipation, urinary frequency or urgency, dyspepsia, and abnormal vaginal bleeding.

No symptom had an LR lower than 0.5, and most had LRs– between 0.8 and 1.0, indicating very little value when absent in ruling out ovarian cancer. This is disappointing, as it would be helpful if there were symptoms whose absence largely ruled out ovarian cancer.

Threshold effects were observed for several variables, such as abdominal distension or increased girth, abdominal or pelvic pain, and abdominal or pelvic bloating. Although threshold effects are easily understood with blood tests, when there is explicit variation in the cut offs defining normal and abnormal tests, they may also occur with signs and symptoms. The above symptoms are fairly subjective, and one can imagine that how the question is asked, when data are gathered in the progression of illness, or even cultural factors could affect the likelihood of a positive or negative response regarding each symptom. This implicit variation in test thresholds is an important source of variation in diagnostic accuracy, and suggests the need for more careful definition of terms for patients and physicians when studying diagnosis.

Combinations of symptoms are promising. This approach has been effective for diagnosis of strep throat, appendicitis, pulmonary embolism, and many other conditions.[31–33] Five validation studies of the Ovarian Cancer Symptom Index were found, but no validation studies of two other promising scores.[14,23] In general, combinations of symptoms had higher sensitivity than individual symptoms, but their specificity was still inadequate to be useful as a screening tool in a low prevalence population.

Limitations

The present analysis was primarily limited by the poor quality of most of the available studies. In addition, some studies appeared to have gathered relevant data, but they

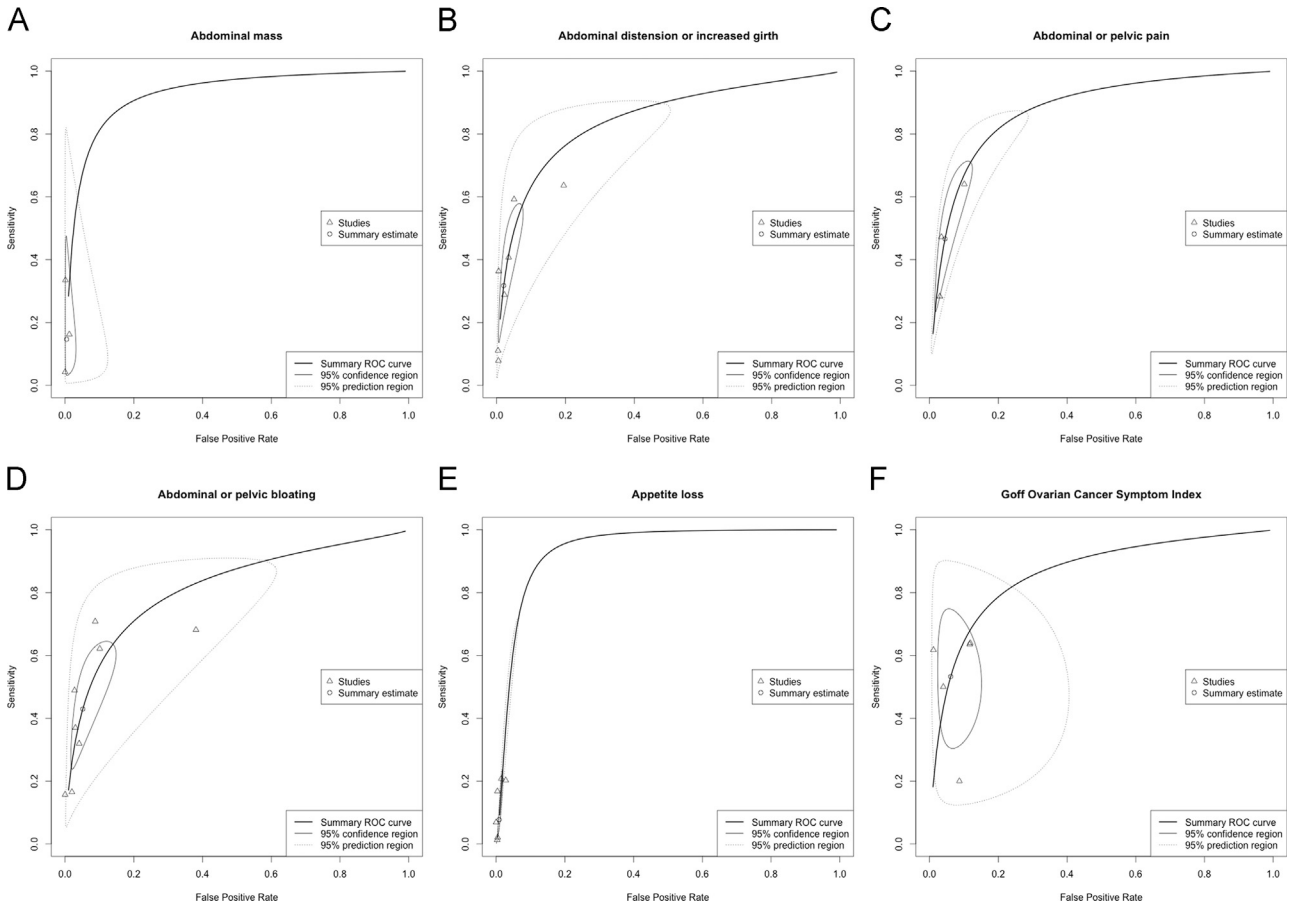

Figure 1. Summary ROC curves of the accuracy of key symptoms. A: abdominal mass; B: abdominal distension or increased girth; C: abdominal or pelvic pain; D: abdominal or pelvic bloating; E: loss of appetite; and F: Goff clinical decision rule.

were not reported in a way that data could be abstracted for the calculation of sensitivity and specificity. The small number of studies for most symptoms meant that in some cases it was not possible or appropriate to calculate the area under the ROC curve.

Suggestions for Future Research

Two promising symptom scores have not been validated prospectively, an important step before considering their integration into clinical practice.[14,23] The integration of risk factors such as menstrual or family history with symptoms has not been fully evaluated, and may be a promising avenue. Similarly, the integration of physical examination findings with symptoms in a decision rule has not been explored. Future studies of signs and symptoms for the diagnosis of ovarian and other cancers should fully report results, including reporting the number of true positive, false positive, false negative, and true negative results for each sign or symptom to facilitate future meta-analyses. Though a cohort design may not always be feasible owing to resource limitations, the controls should be selected to resemble cases as closely as possible other than the diagnosis of ovarian cancer. In addition, three[21,24,25] of the five validation studies of the Ovarian Cancer Symptom Index were performed by the group that developed the score. To ensure that they are truly generalizable, it and other symptoms scores should ideally be validated by groups other than the designers and in the primary care setting where such a tool would be used. Because of their promising sensitivity, more work is needed to validate decision rules and develop new decision support tools that integrate risk factors, symptoms, novel biomarkers, and even imaging to identify women at increased risk for ovarian cancer.

Implications for Clinical Practice

The annual incidence of ovarian cancer is low, and evaluation of suspected ovarian cancer may lead to invasive tests and even unnecessary surgery. In the Prostate Lung Colon Ovary RCT, approximately five cases of ovarian cancer were detected per 10,000 person-years of follow-up.[4] Given the point estimates of 63% sensitivity and 95% specificity for the Ovarian Cancer

Symptom Index, in a group of 10,000 women, it would detect three of five cancers, but 500 women would have a false positive result, resulting in worry, at a minimum an ultrasound of the ovaries, and possibly a surgical biopsy. The positive predictive value would be only 3 of 503 or 0.6%. Thus, there is currently inadequate evidence to recommend the Ovarian Cancer Symptom Index for implementation as a cancer screening tool.

Mark H. Ebell MD, MS, is a member of the U.S. Preventive Services Task Force (USPSTF). This article does not necessarily represent the views and policies of the USPSTF.

No financial disclosures were reported by the authors of this paper.

References

1. Jelovac D, Armstrong DK. Recent progress in the diagnosis and treatment of ovarian cancer. *CA Cancer J Clin*. 2011;61(3):183–203. http://dx.doi.org/10.3322/caac.20113.
2. Salehi F, Dunfield L, Phillips KP, Krewski D, Vanderhyden BC. Risk factors for ovarian cancer: an overview with emphasis on hormonal factors. *J Toxicol Environ Health B Crit Rev*. 2008;11(3-4):301–321. http://dx.doi.org/10.1080/10937400701876095.
3. Brody LC, Biesacker BB. Breast cancer susceptibility genes. *BRCA1 and BRCA2. Medicine (Baltimore)*. 1998;77(3):208–226. http://dx.doi.org/10.1097/00005792-199805000-00006.
4. Buys SS, Partridge E, Black A, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Trial. *JAMA*. 2011;305(22):2295–2303. http://dx.doi.org/10.1001/jama.2011.766.
5. Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. *BJOG*. 2005;112(7):857–865. http://dx.doi.org/10.1111/j.1471-0528.2005.00572.x.
6. Jayde V, White K, Blomfield P. Symptoms and diagnostic delay in ovarian cancer: a summary of the literature. *Contemp Nurse*. 2009;34(1):55–65. http://dx.doi.org/10.5172/conu.2009.34.1.055.
7. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536. http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009.
8. Ebell MH, Culp M, Lastinger K, Dasigi T. A systematic review of the bimanual examination as a test for ovarian cancer. *Am J Prev Med*. 2015;48(3):350–356. http://dx.doi.org/10.1016/j.amepre.2014.10.007.
9. Friedman GD, Skilling JS, Udaltsova NV, Smith LH. Early symptoms of ovarian cancer: a case-control study without recall bias. *Fam Pract*. 2005;22(5):548–553. http://dx.doi.org/10.1093/fampra/cmi044.
10. Rossing MA, Wicklund KG, Cushing-Haugen KL, Weiss NS. Predictive value of symptoms for early detection of ovarian cancer. *J Natl Cancer Inst*. 2010;102(4):222–229. http://dx.doi.org/10.1093/jnci/djp500.
11. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–1558. http://dx.doi.org/10.1002/sim.1186.
12. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration; 2010: Chapter 10. http://srdta.cochrane.org/.
13. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ*. 2009;339:b2998. http://dx.doi.org/10.1136/bmj.b2998.
14. Grewal K, Hamilton W, Sharp D. Ovarian cancer prediction: development of a scoring system for primary care. *BJOG*. 2013;120(8):1016–1019. http://dx.doi.org/10.1111/1471-0528.12200.
15. Macuks R, Baidekalna I, Donina S. Comparison of different ovarian cancer detection algorithms. *Eur J Gynaecol Oncol*. 2011;32(4):408–410.
16. Macuks R, Baidekalna I, Donina S. Diagnostic test for ovarian cancer composed of ovarian cancer symptom index, menopausal status and ovarian cancer antigen CA125. *Eur J Gynaecol Oncol*. 2011;32(3):286–288.
17. Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *BMJ*. 2011;344:d8009. http://dx.doi.org/10.1136/bmj.d8009.
18. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: independent and external validation of QCancer® (Ovarian) prediction model. *Eur J Cancer Care (Engl)*. 2013;22(4):423–429. http://dx.doi.org/10.1111/ecc.12015.
19. Behtash N, Ghayouri Azar E, Fakhrejahani F. Symptoms of ovarian cancer in young patients 2 years before diagnosis, a case-control study. *Eur J Cancer Care (Engl)*. 2008;17(5):483–487. 10.1111/j.1365-2354.2007.00890.x.
20. Riman T, Dickman PW, Nilsson S, et al. Risk factors for invasive epithelial ovarian cancer: results from a Swedish case-control study. *Am J Epidemiol*. 2002;156(4):363–373. http://dx.doi.org/10.1093/aje/kwf048.
21. Andersen MR, Lowe KA, Goff BA. Value of symptom-triggered diagnostic evaluation for ovarian cancer. *Obstet Gynecol*. 2014;123(1):73–79. http://dx.doi.org/10.1097/AOG.0000000000000051.
22. Pavlik EJ, Saunders BA, Doran S, et al. The search for meaning—Symptoms and transvaginal sonography screening for ovarian cancer: predicting malignancy. *Cancer*. 2009;115(16):3689–3698. http://dx.doi.org/10.1002/cncr.24407.
23. Lim AW, Mesher D, Gentry-Maharaj A, et al. Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst*. 2012;104(2):114–124. http://dx.doi.org/10.1093/jnci/djr486.
24. Andersen MR, Goff BA, Lowe KA, et al. Use of a Symptom Index, CA125, and HE4 to predict ovarian cancer. *Gynecol Oncol*. 2010;116(3):378–383. http://dx.doi.org/10.1016/j.ygyno.2009.10.087.
25. Andersen MR, Goff BA, Lowe KA, et al. Combining a symptoms index with CA 125 to improve detection of ovarian cancer. *Cancer*. 2008;113(3):484–489. http://dx.doi.org/10.1002/cncr.23577.
26. Goff BA, Mandel LS, Melancon CH, Muntz HG. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA*. 2004;291(22):2705–2712. http://dx.doi.org/10.1001/jama.291.22.2705.
27. Vine MF, Calingaert B, Berchuck A, Schildkraut JM. Characterization of prediagnostic symptoms among primary epithelial ovarian cancer cases and controls. *Gynecol Oncol*. 2003;90(1):75–82. http://dx.doi.org/10.1016/S0090-8258(03)00175-6.
28. Olson SH, Mignone L, Nakraseive C, Caputo TA, Barakat RR, Harlap S. Symptoms of ovarian cancer. *Obstet Gynecol*. 2001;98(2):212–217. http://dx.doi.org/10.1016/S0029-7844(01)01457-0.
29. Lurie G, Thompson PJ, McDuffie KE, Carney ME, Goodman MT. Prediagnostic symptoms of ovarian carcinoma: a case-control study. *Gynecol Oncol*. 2009;114(2):231–236. http://dx.doi.org/10.1016/j.ygyno.2009.05.001.
30. Goff BA, Mandel LS, Drescher CW, et al. Development of an ovarian cancer symptom index: possibilities for earlier detection. *Cancer*. 2007;109(2):221–227. http://dx.doi.org/10.1002/cncr.22371.
31. McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. *CMAJ*. 2000;163(7):811–815.
32. Ebell MH, Shinholser J. What are the most clinically useful cutoffs for the Alvarado and pediatric appendicitis scores? A systematic review. *Ann Emerg Med*. 2014;64(4):365–372. http://dx.doi.org/10.1016/j.annemergmed.2014.02.025.

33. Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. *Ann Intern Med*. 2001;135(2):98–107. http://dx.doi.org/10.7326/0003-4819-135-2-200107170-00010.

Appendix

Supporting data

Supplementary data associated with this article can be found at http://dx.doi.org/10.1016/j.amepre.2015.09.023.