Diagnostic accuracy of tests for lymph node status in primary cervical cancer: a systematic review and meta-analysis

Tara J. Selman, Christopher Mann MD, Javier Zamora PhD, Tracy-Louise Appleyard MBBS, Khalid Khan MSc

∞   See related article page 867

ABSTRACT

Background: Lymph node status is the key to determining the prognosis and treatment of cervical cancer. However, it cannot be assessed clinically, and testing for nodal metastasis is controversial. We sought to systematically review the diagnostic accuracy literature on sentinel node biopsy, positron emission tomography, magnetic resonance imaging and computed tomography to evaluate the accuracy of each index test in determining lymph node status in patients with cervical cancer.

Methods: We searched MEDLINE (1966–2006), EMBASE (1980–2006), Medion (1980–2006) and the Cochrane library (Issue 2, 2006) for relevant articles. We also manually searched the reference lists from primary articles and reviews, and we contacted experts in the field for conference abstracts and unpublished studies. We performed random-effects meta-analysis of accuracy indices, and we performed meta-regression analysis to test the effect of study quality on diagnostic accuracy and to identify other sources of heterogeneity.

Results: We included 72 relevant primary studies, involving a total of 5042 women, in our analysis. We found that, in determining lymph node status, sentinel node biopsy had a pooled positive likelihood ratio of 40.8 (95% confidence interval [CI] 24.6–67.6) and a pooled negative likelihood ratio of 0.18 (95% CI 0.14–0.24). The pooled positive likelihood ratios (and 95% CI) were 15.3 (7.9–29.6) for positron emission tomography, 6.4 (4.9–8.3) for magnetic resonance imaging and 4.3 (3.0–6.2) for computed tomography. The pooled negative likelihood ratios (and 95% CIs) were 0.27 (0.11–0.66) for positron emission tomography, 0.50 (0.39–0.64) for magnetic resonance imaging and 0.58 (0.48–0.70) for computed tomography. Using a 27% pretest probability of lymph node metastasis among all cases (regardless of stage), we found that a positive sentinel node biopsy result increased post-test probability to 94% (95% CI 90%–96%), whereas a positive finding on positron emission tomography increased it to 85% (75%–92%).

Interpretation: Sentinel node biopsy has greater accuracy in determining lymph node status among women with primary cervical cancer than current commonly used imaging methods.

Une version française de ce résumé est disponible à l'adresse www.cmaj.ca/cgi/content/full/178/7/855/DC1

I n the United States and European Union it is estimated that cervical cancer is diagnosed in about 42 000 women each year.[1,2] The International Federation of Gynecology and Obstetrics criteria currently used to stage cervical cancer do not account for lymph node involvement, but the lymphatic spread of the disease is key to determining prognosis and appropriate treatment.

The primary treatment options for early cervical cancer (stage 1B1 or less advanced) are surgery and chemoradiotherapy, which have similar survival rates.[3] Surgical treatment offers some degree of fertility preservation and may not have the long-term complications associated with chemoradiotherapy. However, because it is not possible to clinically detect pelvic and para-aortic lymph node metastasis, surgery typically includes lymphadenectomy, which may reveal metastatic spread. In such cases, chemoradiotherapy is required, which would make the initial surgical procedure unnecessary in retrospect. There has been considerable interest in using a preoperative, noninvasive test to determine lymph node status in order to select the most appropriate treatment option. Such a test could avoid unnecessary surgical intervention, reduce morbidity and correctly direct choice of treatment.

In recent years, the use of magnetic resonance imaging and computed tomography to determine lymph node status

From the Department of Reproductive and Child Health (Selman, Mann, Khan), University of Birmingham, Birmingham Women's Hospital, Birmingham, United Kingdom; the Clinical Biostatistics Unit (Zamora), Hospital Ramón y Cajal, Madrid, Spain; and the Department of Obstetrics and Gynaecology (Appleyard), St. Michael's Hospital, Bristol, United Kingdom

has increased; however, neither method has been formally included as part of International Federation of Gynecology and Obstetrics staging of cervical cancer. Sentinel node biopsy and positron emission tomography have emerged as competitors to magnetic resonance imaging and computed tomography. The accuracy of magnetic resonance imaging, computed tomography and positron emission tomography has been assessed previously in reviews,[4–6] but updates are required because recent studies have reported on their diagnostic accuracy[7] and on quality assessment in diagnostic reviews.[8] Given this background, we performed a systematic review of the literature to compare the accuracy of these 4 methods in determining lymph node status in patients with cervical cancer.

Methods

Using a prospective protocol based on widely recommended methods[9–13] we carried out a systematic review of the literature.

Literature search
We attempted to capture in our search all studies that reported the diagnostic accuracy of magnetic resonance imaging, computed tomography, positron emission tomography and sentinel node biopsy in the detection of lymphatic spread of cervical cancer. We did not search for studies on ultrasonography or pedal lymphoscintigraphy because it is generally accepted that these techniques are not suitable for assessing lymph node status in cervical cancer. To identify relevant studies, we searched MEDLINE (1966–2006), EMBASE (1980–2006), the Cochrane Library (Issue 2, 2006 and Medion (1980–2006). We combined National Library of Medicine medical subject headings, keywords and word variants for cervical cancer with terms for each of the 4 index tests and lymphadenopathy. We also manually searched the reference lists from primary articles and other reviews that we found on those databases to identify studies that may have been missed in electronic searching. In addition, we contacted experts in the field to obtain unpublished studies and conference abstracts, and we reviewed these in an attempt to identify grey literature. Details of the search strategy are provided in Appendix 1 (available online at www.cmaj.ca/cgi/content/full/178/7/855/DC2).

Study selection
Our study selection was a 2-stage process. First, from the electronic searches we retrieved the full-text articles of potentially relevant citations and evaluated them based on predefined selection criteria. We selected studies if they reported on the accuracy of the index tests compared with histological examination of the lymph nodes (reference standard) in women with a primary presentation of cervical cancer of any histological type or stage, and if their data could be used to create 2 × 2 tables. We excluded studies if they involved fewer than 10 participants. No language restrictions were applied. In cases of duplicate publication of manuscripts, we selected the most recent version. Of the studies that met our initial in-

clusion criteria, we made our final selection after examining the full-text article. Two of us (T.J.S. and T.L.A.) independently reviewed 10% of the manuscripts; one of us (T.J.S.) reviewed the rest of the studies. A third reviewer (K.S.K.) resolved any disagreements.

Data abstraction
We collected information on study characteristics, quality and accuracy results from each of the selected articles using a standardized data-collection form. The study characteristics that we extracted were the stage of disease, the index test, the reference standard methodology, and the setting and year of the study. We recorded accuracy data from the studies in 2 × 2 tables. For the purpose of analysis, when a manuscript reported the accuracy of more than one index test, we reported on the tests separately. We excluded nondiagnostic test results and failure to complete the test (e.g., inability to detect the sentinel node, inadequate histology) from the 2 × 2 tables; however, we recorded these occurrences and the results from the reference standard, if provided, in each case.

Quality assessment
We defined methodologic quality as the confidence that the study design, conduct and analysis minimized biases in the estimation of test accuracy. We used the existing QUADAS (quality assessment of diagnostic accuracy studies) criteria and the STARD (standards for reporting of diagnostic accuracy) criteria to generate the quality-assessment criteria for our evaluation of the studies in our review.[8,14–16]

For the study population, we considered consecutive or random recruitment of eligible women to be ideal; convenience sampling, such as arbitrary recruitment or nonconsecutive recruitment, was deemed inadequate. In addition, we considered prospective recruitment of patients to be potentially associated with a lesser degree of bias than retrospective recruitment. We considered the description of the population to be ideal if the study clarified the stage of disease and recorded the patients' body mass index, which can affect imaging techniques. We recorded the stage of disease as early (stage 1B1 or less advanced), or late (more advanced than stage 1B1) in accordance with International Federation of Gynecology and Obstetrics staging criteria.

We considered the reporting of the index test to be ideal if there was sufficient detail to allow other researchers to replicate the test. It was also important for the time interval between the index test and the reference standard to be described; we considered an interval of 4 weeks or less to be suitable.[16] For the reference standard itself, a description of the method of histological verification was important, and we deemed it preferable for the readers of the reference standard to be blinded to the index test results.

We examined partial and differential verification by comparing information in the articles on the number of women recruited into the study and the number of women for whom outcome data were known. We considered verification to be ideal if all women originally enrolled into the study without legitimate exclusions were included in the

data analysis. We examined whether withdrawals from the study were explained and whether uninterpretable results were reported.

We evaluated the main strengths and weaknesses against the quality-assessment criteria for all studies included in our systematic review. We did not attempt to collapse our assessment of quality into a score because suggested methods for such an approach have been found to have poor validity, and the collapsed scores may obscure the strengths and weaknesses of a study rather than clarify them. We did, however, perform a meta-regression analysis (described in the next section), and we used our findings to categorize studies as high, medium or low quality. We felt that the quality-assessment criteria were essential for all primary studies, regardless of the type of index test used.

Data synthesis and statistical analysis

We computed the sensitivity, specificity and likelihood ratios for each index test. When 2 × 2 tables contained cells for which the value was 0, we added 0.5 to those cells to allow for the calculation of variances.[17] We examined a threshold effect by plotting sensitivity against reverse specificity in a receiver-operating-characteristic analysis and by calculating Spearman correlation coefficients.[13]

We examined heterogeneity visually, using forest plots of sensitivity, specificity and likelihood ratios, and statistically, using the Cochran Q test.[17] We explored the reasons for heterogeneity using meta-regression and subgroup analyses, planned a priori in keeping with published recommendations.[18,19]

To explore the effect of study quality, we first used all the 2 × 2 tables to assess whether the quality-assessment criteria produced variation in the log of the diagnostic odds ratio. We performed univariable meta-regression analysis to select the quality-assessment criteria that had a statistically significant effect on diagnostic performance. We then performed a multivariable analysis to identify those criteria that had the most effect in our data set, which allowed us to categorize studies into high-, medium- or low-quality subgroups. High-quality studies met all of the quality-assessment criteria found to have a statistically significant effect in the multivariable analysis; medium-quality studies met at least one criterion; and low-quality studies did not meet any of the criteria. Assuming that high-quality studies provide the most valid assessment of test accuracy, we used high-quality studies as the reference category to determine whether medium- and low-quality studies had biased estimates of accuracy.

We investigated the heterogeneity resulting from population or test characteristics by exploring the effects of predefined variables.[19] For the population, we examined the effect of disease stage (early v. late or mixed) and the lymph node groups (pelvic v. para-aortic). For test characteristics, we considered the type of index test, the method of sentinel node identification, the surgical approach (open or laparoscopic) used for the sentinel node biopsy and the histological method used in the reference standard (immunohistochemistry, or hematoxylin and eosin staining). We initially performed univariable analysis and then progressed to multivariable analy-

sis. To assess the effects of the type of index test, we used magnetic resonance imaging as the reference category, because, where available, it is the most frequently used index test to determine lymph node status in cervical cancer. We adjusted the models produced by multivariable analysis for the effect of study quality. To explore the potential effect of multiple counting of the same patients, we performed a sensitivity analysis in which we excluded duplicate data and compared results with those obtained from the whole data set.

We summarized sensitivity, specificity and likelihood ratios with 95% confidence intervals (CIs) for each index test separately.[20–22] We pooled individual results weighted in inverse proportion to variance using a random-effects model[12] in light of unexplained heterogeneity. We tested the robustness of these meta-analytical summaries with those generated using a bivariable method.[23]

We used pooled likelihood ratios to determine post-test probability for positive and negative index test results. We
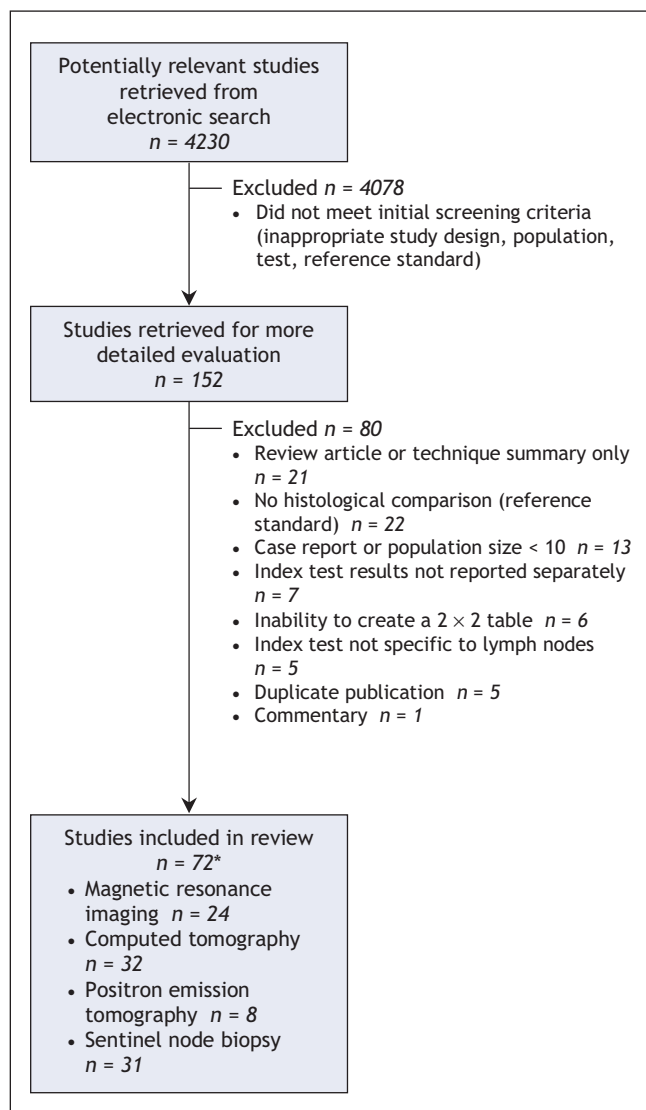
Figure 1: Search and selection of studies for systematic review. *Some studies reported on more than one index test.

computed the range of uncertainty in estimations of post-test probability using the upper and lower bounds of the CIs of likelihood ratios for each test. We pooled the failure rates of sentinel node biopsy using a random-effects model, weighting each proportion by the inverse of its variance. We investigated the effects of the different techniques, (blue dye alone, and technetium 99m colloidal albumin with or without blue dye) on successful identification of the sentinel node. We explored the possibility of publication and related biases using funnel plots of log diagnostic odds ratio versus the inverse of variance.[24,25]

Results

Of the 4230 articles identified through the electronic database searches, we retrieved the full text of 152 articles for more de-tailed evaluation. A total of 72 studies (citations are included in Appendix 1, available online at www.cmaj.ca/cgi/content /full/178/7/855/DC2) involving 5042 women with cervical cancer met the selection criteria for the review (Figure 1). From the data in those studies, we created 95 2 × 2 tables, each evaluating 1 of 4 index tests. A proportion of the study participants (13.8%, 695/5042) were included more than once in 17 of the 2 × 2 tables. The studies had a number of methodological deficiencies. Appendices 2 and 3 (available online at www.cmaj.ca/cgi/content/full/178/7/855/DC2) sum-marize the salient features and quality of each of the studies.

Explanation of heterogeneity
We found that heterogeneity was apparent among the stud-ies, but was statistically nonsignificant (Appendix 4, avail-able online at www.cmaj.ca/cgi/content/full/178/7/855/DC2).

Table 1: Factors affecting estimations of accuracy for sentinel node biopsy, positron emission tomography, magnetic resonance imaging and computed tomography in determining lymph node status in cervical cancer

Factor	Univariable analysis		Multivariable analysis	
	OR* (95% CI)	p value	OR* (95% CI)	p value
Index test type				
Sentinel node biopsy v. magnetic resonance imaging	20.68 (8.80-48.59)	< 0.01¶	18.49 (3.59-95.17)	< 0.01**
Positron emission tomography v. magnetic resonance imaging	4.22 (1.41-12.66)	0.01¶	3.84 (1.22-12.12)	0.02**
Computed tomography v. magnetic resonance imaging	0.61 (0.36-1.03)	0.06¶	0.63 (0.36-1.12)	0.11**
Quality-assessment criteria†				
Data collection (prospective v. retrospective)	0.49 (0.25-0.95)	0.03	0.81 (0.42-1.54)	0.51
Verification (whole, random sample v. incomplete)	0.47 (0.25-0.96)	0.02	0.61 (0.34-1.10)	0.10
Adequate v. inadequate description of index test	0.19 (0.09-0.38)	< 0.01	0.27 (0.14-0.53)	< 0.01
Adequate v. inadequate description of reference standard	0.38 (0.17-0.87)	0.02	0.68 (0.32-1.43)	0.30
Reporting of study withdrawals (reported v. nonreported)	0.33 (0.17-0.65)	< 0.01	0.44 (0.23-0.86)	0.02
Time period between index test and reference standard (≤ 4 weeks v. > 4 weeks)	0.42 (0.23-0.77)	< 0.01	0.44 (0.26-0.76)	< 0.01
Study quality‡				
High v. medium	0.07 (0.03-0.16)	< 0.01¶	0.48 (0.16-1.41)	0.18**
High v. low	0.06 (0.02-0.21)	< 0.01¶	0.48 (0.12-1.19)	0.29**
Index test characteristics				
Sentinel node biopsy (technetium 99m v. blue dye)	1.84 (0.54-6.23)	0.32		—
Sentinel node biopsy (laparoscopic v. open)	1.08 (0.37-3.20)	0.78		—
Other study characteristics				
Reference standard (immunohistochemistry v. hematoxylin and eosin staining)§	0.05 (0.02-0.15)	0.05	1.68 (0.33-8.64)	0.53**
Stage of disease (early v. advanced or mixed)	0.41 (0.15-1.07)	0.07	1.39 (0.62-3.07)	0.42**
Lymph node type (pelvic v. para-aortic or mixed)	1.20 (0.62-2.27)	0.57	0.94 (0.52-1.69)	0.84**

Note: CI = confidence interval, OR = odds ratio.
*Relative diagnostic odds ratio comparing the diagnostic odds ratios described in studies of magnetic resonance imaging (reference category) with those described in studies of another index test. A relative diagnostic odds ratio > 1 indicates an index test with greater diagnostic accuracy than the reference category.
†Only study quality items that were found in the univariable analysis to be statistically significant are shown.
‡Study quality was calculated using the 3 quality-assessment criteria that were found in the multivariable analysis to be statistically significant. Studies that met all 3 criteria were considered to be high quality, those that met 1 criterion were considered to be medium quality, and those that did not meet the criteria were considered to be low quality. This quality grading was used in subsequent multivariable analysis.
§For the purpose of analysis, for studies that did not state the reference standard type, we presumed that hematoxylin and eosin staining was used because it is the more common method.
¶Univariable analysis using dummy variables to set the reference category as medium and low study quality in each case.
**Multivariable analysis including quality grade, index test type, reference standard type, stage of disease and lymph node type as explanatory variable. See details in methods section.

We explored whether the quality-assessment criteria had an effect on accuracy and found that 6 criteria (data collection, verification, adequate description of the index test, adequate description of the reference standard, reporting of study withdrawals, and time period between index test and reference standard < 4 weeks) had a significant influence in univariable analysis. Three criteria (adequate description of the index test, reporting of study withdrawals, and time period between index test and reference standard < 4 weeks) remained significant after multivariable analysis (Table 1) We used these 3 criteria to grade the quality of studies: 32 were high quality, 52 were medium quality and 11 were low quality. Results of our univariable analysis indicated that estimates of diagnostic accuracy, as measured by diagnostic odds ratios (OR), were more conservative in high-quality studies than in medium-quality studies (relative diagnostic OR 0.07, 95% CI 0.03–0.16) or low-quality studies (relative diagnostic OR 0.06, 95% CI 0.02–0.21). However, these differences were no longer significant in the multivariable analysis (Table 1). Univariable analysis showed that the type of index test and the reference standard could explain the heterogeneity, and that the other clinical variables we defined previously did not affect the accuracy of the test. However, the type of index test remained the only significant explanation for overall heterogeneity after adjustment for the effects of other predefined quality-assessment criteria (Table 1). Given the low proportion of duplicated data, sensitivity analyses excluding these studies showed practically the same results (data not shown).

Diagnostic test characteristics

Individual and summary results for each of the 4 index tests are shown in Appendix 3 (available online at www.cmaj.ca /cgi/content/full/178/7/855/DC2) and Figure 2. For each of the index tests, variation in sensitivity was much greater than variation in specificity, but there was no visual or statistical correlation between sensitivity and specificity (Table 2). We found that sentinel node biopsy was the most accurate index test in determining lymph node status, with a positive likelihood ratio of 40.8 (95% CI 24.6–67.6 and a negative likelihood ratio of 0.18 (95% CI 0.14–0.24) (Table 2). The failure rate (and 95% CI) for the detection of the sentinel node was 10.9% (1.5%–27.4%). The failure rates (and 95% CIs) were 8.4% (3.3%–15.5%) for using blue dye alone and 4.4% (2.0%–7.7%) for use of a combined technique using blue dye and technetium 99m colloidal albumin. We did not find a difference in accuracy between these techniques or between open and laparoscopic surgery (Table 1).

When we adjusted for the effects of study quality we found that positron emission tomography and sentinel node biopsy were significantly better methods for determining lymph node status than were magnetic resonance imaging and computed tomography (Table 1). In Table 3, we show the post-test probabilities associated with the various index tests in ruling out lymph node metastasis in cervical cancer patients. Some heterogeneity remained among studies of each test, but this could not be explained by variables that we defined a priori. Subgroup analysis limiting

the comparison of index tests to high- and medium-quality studies did not alter our conclusions. The bivariable analysis showed results and comparisons to be similar to those described above (Figure 2). Funnel plots (not shown) for all the index tests were symmetrical, showing an absence of publication bias.

Interpretation

Our review showed that sentinel node biopsy provided a more accurate assessment of lymph node metastasis in cervical cancer than noninvasive imaging tests. We observed that positron emission tomography was more accurate than magnetic resonance imaging or computed tomography, although the current summary of the accuracy of positron emission tomography may be imprecise owing to the relatively small number of studies evaluating the accuracy of that method.

Our review, which complied with the current criteria for diagnostic reviews,[7,13] provides a robust summary of the available evidence to date. We performed an extensive search for studies, used well-developed methods for quality assessment and investigated potential sources of heterogeneity with advanced statistical techniques planned a priori. The deficiencies in methodologic quality that we found in the studies in
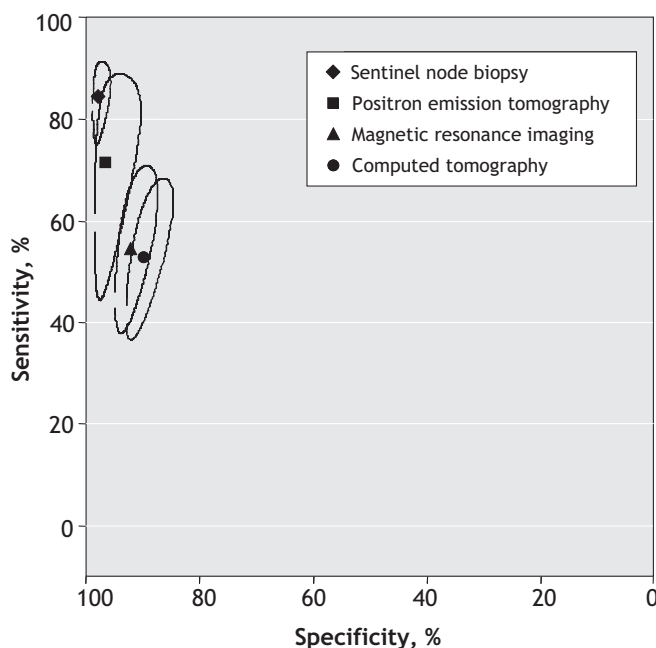
Figure 2: Bivariable analysis of the accuracy of sentinel node biopsy, positron emission tomography, magnetic resonance imaging and computed tomography in determining lymph node status in patients with cervical cancer. The bivariable analysis produces mean estimates of sensitivity and specificity along with the 95% confidence intervals of each index test. Each ellipsis represents the region containing likely combinations of the mean value of sensitivity and specificity. The closer the index values are to the upper-left corner, the greater the accuracy of that index test. Note: the x axis shows reversed specificity.

our review should help improve further research in this area. However, these deficiencies potentially threaten the validity of our findings. A potential limitation of our review was that the reviewers were not blinded to authors of the studies included in our analysis or to the journals in which the studies were published. However, one reviewer was not an expert in this field, which would limit bias due to author recognition. Another limitation of our approach may be that, in light of the unexplained heterogeneity in the results, meta-analysis should perhaps have been avoided.

We believe that our inferences concerning the value of tests are robust because our multivariable analysis exploring reasons for heterogeneity (in accordance with the recommended guidelines[11,12]) showed that the differences in accuracy among studies could not be explained by the vari-

ation in study quality, stage of disease, or site of lymph node (pelvic or para-aortic). In addition, estimates of test accuracy that we observed in high-quality studies were consistent with overall results. Homogeneity is one of the desired prerequisites for meta-analysis, but it is not an absolute requirement. There is an onus on the reviewer to thoroughly investigate the potential causes of heterogeneity, which we did. In the presence of unexplained and unavoidable heterogeneity, a random-effects model provides the most useful estimate for informing practice. Triangulating different methods to explore heterogeneity, to pool results of the meta-analysis and to compare index tests supported our main findings.

Based on our findings, women with cervical cancer, particularly younger women who wish to preserve reproductive

Table 2: Pooled and single estimates for index test prediction of lymph node status in patients with cervical cancer

Index test	No. of studies	No. of women	Sensitivity (95% CI), %	Specificity (95% CI), %	Pooled positive LR (95% CI)	Pooled negative LR (95% CI)
Sentinel node biopsy	31	1140	91.4 (87.1-94.6)	100 (99.6-100)	40.8 (24.6-67.6)	0.18 (0.14-0.24)
Positron emission tomography	8	445	74.7 (63.3-84.0)	97.6 (95.4-98.9)	15.3 (7.9-29.6)	0.27 (0.11-0.66)
Magnetic resonance imaging	24	1206	55.5 (49.2-61.7)	93.2 (91.4-94.0)	6.4 (4.9-8.3)	0.50 (0.39-0.64)
Computed tomography	32	2640	57.5 (53.5-61.4)	92.3 (91.1-93.5)	4.3 (3.0-6.2)	0.58 (0.48-0.70)

Note: CI = confidence interval, LR = likelihod ratio.

Table 3: Post-test probabilities of lymph node metastasis in patients with cervical cancer, by index test

Index test; cancer stage*	Pretest probability of lymph node metastasis	Post-test probability for positive test result (95% CI)	Post-test probability for negative test result (95% CI)
Sentinel node biopsy			
Early disease	0.19	0.90 (0.85-0.94)	0.04 (0.03-0.05)
All stages	0.27	0.94 (0.90-0.96)	0.06 (0.05-0.08)
Advanced disease	0.44	0.97 (0.95-0.98)	0.12 (0.10-0.16)
Positron emission tomography			
Early disease	0.19	0.78 (0.64-0.87)	0.06 (0.02-0.13)
All stages	0.27	0.85 (0.75-0.92)	0.09 (0.04-0.02)
Advanced disease	0.44	0.92 (0.83-0.96)	0.18 (0.08-0.34)
Magnetic resonance imaging			
Early disease	0.19	0.59 (0.53-0.66)	0.10 (0.08-0.13)
All stages	0.27	0.70 (0.64-0.75)	0.16 (0.13-0.20)
Advanced disease	0.44	0.83 (0.79-0.87)	0.28 (0.24-0.34)
Computed tomography			
Early disease	0.19	0.50 (0.40-0.59)	0.12 (0.10-0.14)
All stages	0.27	0.61 (0.52-0.70)	0.18 (0.15-0.21)
Advanced disease	0.44	0.77 (0.70-0.83)	0.32 (0.28-0.36)

Note: CI = confidence interval, LR = likelihood ratio.
*Early disease is defined as stages 1B1 or less advanced and advanced disease is defined as stages more advanced than 1B1, according to International Federation of Gynecology and Obstetrics staging criteria.

potential, may judiciously be able to avoid radical surgery or chemoradiotherapy if sentinel node biopsy is first used to determine the most appropriate treatment option. On the odds ratio scale, sentinel node biopsy was 20 times more accurate than magnetic resonance imaging, whereas positron emission tomography was 4 times more accurate. In breast cancer, sentinel node biopsy is now recommended for routine use in selected groups, and long-term survival data for patients with negative sentinel node biopsy results have shown no adverse outcomes.[26,27] Similarly, sentinel node biopsy is offered to patients with early melanoma,[28] and has shown potential in vulva cancer.[29] Armed with our findings on post-test probabilities (Table 3), patients and clinicians will be in a better position to individualize treatment.

A number of remaining issues need to be examined before translation into clinical practice can be considered. In this review we have calculated an average failure rate of 4.4% to detect the sentinel node. However, this does not take into account whether a sentinel node was detected bilaterally. Because the cervix is a central organ, we expect its drainage would be bilateral and that sentinel nodes, therefore, would be detected bilaterally. Appendix 2 (available online at www.cmaj.ca/cgi/content/full/178/7/855/DC2) clearly shows that not all authors of the studies we reviewed detected sentinel nodes bilaterally, or even reported if this was the case. The failure rate for node detection would increase greatly when taking this into consideration. There are concerns that sentinel node biopsy is more invasive than the other index tests. Although it is less invasive than laparoscopic and open surgery while still providing the same level of diagnostic accuracy, sentinel node biopsy still requires a general anesthetic, unlike magnetic resonance imaging, computed tomography and positron emmisson tomography. On a more positive note, none of the studies reviewed reported any serious side effects from sentinel node biopsy.

Magnetic resonance imaging and computed tomography are widely available, but we found these methods to be inferior to positron emission tomography, which brings into question their role in current practice. Technological improvements in magnetic resonance imaging, such as the introduction of phase array coils, do not appear to improve the diagnostic accuracy of this method.[30] Positron emission tomography is a relatively new technique with limited availability. Its diagnostic accuracy may improve with alterations in the techniques used — especially in patient hydration and bladder irrigation, without which the tracer can accumulate in the renal tract, giving false-positive results. Positron emission tomography does not rely on the size of the lymph node to determine its status, which allows it to detect metastasis much earlier. However, the method is subject to false results; for example, false-negative results may occur because of tumour necrosis altering the metabolism and tracer uptake, and false-positive results may occur because of inflammation and the increased metabolism associated with macrophage activity.[31] With improvement in technique and accessibility, positron emission tomography could provide a noninvasive accurate test.[32]

Accurate assessment of lymph node status in the staging of cervical cancer is important to direct treatment and reduce morbidity. Our review suggests that the imaging methods currently used to detect lymph node status may be inaccurate, that positron emission tomography may have a potential role and that sentinel node biopsy may be a minimally invasive, accurate assessment of lymph node status. Further research is needed to assess the practicality of using these techniques and the effect of implementing such tests on patient outcomes and health service costs. In addition, well-designed trials are required to assess the long-term survival of patients for whom treatment was directed following sentinel node biopsy or any of the other diagnostic tests.

This article has been peer reviewed.

REFERENCES

1. National Cancer Institute. Cervical cancer. Bethesda (MD): The Institute. Available: www.cancer.gov/cancertopics/types/cervical (accessed 2008 Feb 21).
2. Arbyn M, Raifu AO, Ferlay J. Burden of cervical cancer in Europe. *Ann Oncol* 2007;18:1708-15.
3. Hatch K. Cervical cancer. In: Berek JS, Hacker NF, editors. *Practical gynecologic oncology.* 2nd ed. Baltimore: Williams & Wilkins; 1994. p. 243-85.
4. Scheidler J, Hricak H, Yu KK, et al. Radiological evaluation of lymph node metastases in patients with cervical cancer. A meta-analysis. *JAMA* 1997;278:1096-101.
5. Bipat S, Glas AS. van d, V, Zwinderman AH, Bossuyt PM, Stoker J. Computed tomography and magnetic resonance imaging in staging of uterine cervical carcinoma: a systematic review. *Gynecol Oncol* 2003;91:59-66.
6. Havrilesky LJ, Kulasingam SL, Matchar DB, et al. FGD-PET for management of cervical and ovarian cancer. *Gynecol Oncol* 2005;97:183-91.
7. Selman TJ, Khan KS, Mann CH. An evidenced-based approach to test accuracy studies in gynecologic oncology: the 'STARD' checklist. *Gynecol Oncol* 2005;96:575-8.
8. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
9. Mallett S, Deeks JJ, Halligan S, et al. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. *BMJ* 2006;333:413-6.
10. Khan KS, Dinnes J, Kleijnen J. Systematic reviews to evaluate diagnostic tests. *Eur J Obstet Gynecol Reprod Biol* 2001;95:6-11.
11. Deeks J, Khan K, Song F, et al. Data synthesis. In: Khan K, Ter Riet G, Glandville J, et al, editors. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews.* 2nd ed. York (UK): University of York; 2001.
12. Deville W, Buntinx F. Guidelines for conducting systematic reviews of studies evaluating the accuracy of diagnostic studies. In: Knottnerus JA, editor. *The evidence base of clinical diagnosis.* London (UK): BMJ Publishing Group; 2002. p. 145-65.
13. Irwig L, Tostesen A, Gatsonis C. Guidelines for meta-analysis evaluating diagnostic tests. *Ann Intern Med* 1994;120:667-76.
14. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
15. Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469-76.
16. Whiting P, Westwood M, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.
17. Sankey S, Weissfeils L, Fine M, et al. An assessment of the use of the continuity correction for sparse data in meta-analysis. *Commun Stat Simulation Comput* 1996;25:1031-56.
18. Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;6:31.
19. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525-37.
20. Deeks J, Morris J. Evaluation diagnostic tests. *Baillieres Clin Obstet Gynaecol* 1996;10:613-30.

21. Jaeschke R, Guyatt G, Sackett D, et al. Users' guide to the medical literature, III: how to use an article about a diagnostic test, B. *JAMA* 1994;271:703-7.
22. Greenhalgh T. How to read a paper. papers that report diagnostic or screening reviews. *BMJ* 1997;315:540-3.
23. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.
24. Song F, Khan KS, Dinnes J, et al. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol* 2002;31:88-95.
25. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58:882-93.
26. Keshtgar MR, Chicken DW, Tobias JS. New approaches in breast cancer management: sentinel node biopsy and intraoperative radiotherapy. *Int J Fertil Womens Med* 2005;50:218-26.
27. Lyman GH, Giuliano AE, Somerfield MR, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol* 2005;23:7703-20.
28. Cochran AJ, Roberts A, Wen DR, et al. Update on lymphatic mapping and sentinel node biopsy in the management of patients with melanocytic tumours. *Pathology* 2004;36:478-84.
29. Selman TJ, Acheson N, Luesley DM, et al. A systematic review of diagnostic tests of inguinal lymph node status in squamous cell carcinoma of the vulva. *Gynecol Oncol* 2005;99:206-214.
30. Yu KK, Hricak H, Subak LL, et al. Preoperative staging of cervical carcinoma: phased array coil fast spin-echo versus body coil spin-echo T2-weighted MR imaging. *AJR Am J Roentgenol* 1998;171:707-11.
31. Reinhardt MJ, Ehritt-Braun C, Vogelgesang D, et al. Metastatic lymph nodes in patients with cervical cancer: detection with MR imaging and FDG PET. *Radiology* 2001;218:776-82.
32. Lord SJ, Irwing L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144:850-5.

Correspondence to: *Tara Selman, Birmingham Women's Hospital NHS Trust, Birmingham B15 2TG, UK; fax 0121 4141576; taraselman@blueyonder.co.uk*