possible, when what they mean is that the probability (P value) is less than .0001. They do not state in the table to what test or null hypothesis the P value refers, so it is not possible for the reader to interpret the statistical significance.

Second, Marra and colleagues¹ report that remediastinoscopy is associated with 100% specificity. That in itself is to be expected, simply because for the vast majority of tests (endobronchial ultrasonography included) performed with diagnostic biopsy, when tumor is seen in a biopsy specimen, it is invariably confirmed on the reference test (there would be serious problems if tumor diagnosed on the index test proved to be normal on the reference test). Therefore specificity (ability to rule in disease) is not the major focus for surgeons when evaluating remediastinoscopy for restaging the mediastinum. The same specificity will be obtained with any other tests combined with tissue biopsy, such as endobronchial ultrasonography (as referenced by Marra and colleagues¹).

Third, what is critical is sensitivity (the ability to rule out disease), so that patients with residual N2 disease are not brought to operation mistakenly, and the reported sensitivity for remediastinoscopy was 61%. It is important for readers to bear in mind that sensitivity truly starts at 50% (a test with no sensitivity, or equipoise), not 0%. Marra and colleagues¹ do not report a confidence interval for this estimate; when one is calculated, it ranges from 42% to 77%. That is, it crosses 50%. This low value of sensitivity is entirely in keeping with the previous literature on remediastinoscopy. De Leyn and associates² attributed this to the technical difficulty in obtaining adequate biopsy specimens from the areas of previous lymph node involvement.

Fourth, Marra and colleagues¹ have used their own arbitrary definition of "diagnostic accuracy," which includes (see footnote in their Table 3) in the denominator "no biopsy." This does not

comply with current Standards for Reporting of Diagnostic Accuracy (STARD).³ The importance of STARD was previously highlighted by Eugene Blackstone⁴ in this Journal in 2004. We believe that compliance with STARD is crucial for readers (which is why it is part of this Journal's policy), and we are surprised that a non-STARD compliant article was highlighted for continuing medical education credit.

On the basis of these findings, we believe that the conclusion that "remediastinoscopy provides a histologic proof of mediastinal downstaging with high diagnostic accuracy" cannot be sustained by the results presented by Marra and colleagues.¹ The results do indicate that residual disease can be ruled in with repeated mediastinoscopy, but the procedure is unable to rule out residual mediastinal disease with any degree of certainty (sensitivity of 61%, 95% confidence interval 42%–77%). Futile thoracotomy thus would not be avoided for a large proportion of patients were this technique used to restage the mediastinum after induction chemotherapy.

Eric Lim, MB, ChB, MD, MSc,
(Biostatistics), FRCS (C-Th)
Michael Dusmet, MD, FMH
The Royal Brompton Hospital
London, UK

References

1. Marra A, Hillejan L, Fechner S, Stamatis G. Remediastinoscopy in restaging of lung cancer after induction therapy. *J Thorac Cardiovasc Surg* 2008;135:843-9.
2. De Leyn P, Stroobants S, De Wever W, Lerut T, Coosemans W, Decker G, et al. Prospective comparative study of integrated positron emission tomography-computed tomography scan compared with remediastinoscopy in the assessment of residual mediastinal lymph node disease after induction chemotherapy for mediastinoscopy-proven stage IIIA-N2 non-small-cell lung cancer: a Leuven Lung Cancer Group study. *J Clin Oncol* 2006;24:3333-9.
3. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. *Am J Clin Pathol* 2003;119:18-22.
4. Blackstone EH, Lauer MS. Caveat emptor: the treachery of work-up bias. *J Thorac Cardiovasc Surg* 2004;128:341-4.

doi:10.1016/j.jtcvs.2008.04.032

Reply to the Editor:

We thank Drs Lim and Dusmet for their comments. We would like to answer some issues they addressed in the statistical analysis of our article.

As concerns the need to report specificity, competing methods for mediastinal restaging are of noninvasive (positron emission tomography, integrated positron emission tomography and computed tomography), or invasive (endoesophageal ultrasonography with fine-needle aspiration biopsy, endobronchial ultrasonography with transbronchial needle aspiration, remediastinoscopy) nature. Noninvasive methods do not provide histologic proof of their positive findings, so specificity varies to a high of 90% (and thus is lower than 100%). Invasive methods are able to provide a histologic or cytologic biopsy specimen of lymph nodes. In the first case, specificity is as high as 100%; in the second, however, cytologic examination may yield false-positive findings. In a review of the literature on cytologic study in pulmonary medicine, Policarpo-Nicolas and Wick¹ reported overall sensitivity and specificity values ranging from 60% to 96%, even in the most conscientious laboratories and considering the increased confidence of pathologists in cytomorphologic diagnosis.

The reported value of sensitivity of 61% in our series is consistent with the previous literature on remediastinoscopy (70%–83%), if slightly lower, with the exception of the study of De Leyn and colleagues,² who reported a disappointing 29% sensitivity in a small series of patients. That report was heavily biased by the inadequate reexploration of the subcarinal station in two thirds of cases (in which persistent nodal disease was found in 50% of cases at thoracotomy) and thus cannot be cited as a standard reference for the diagnostic value of remediastinoscopy. In most cases, sensitivity values of remediastinoscopy are comparable to or higher than those of the other methods for mediastinal restaging.

The greatest problem for the correct evaluation of sensitivity of mediastinoscopy lies in the small sizes of the series reported in the literature. Small sample size determines wide confidence intervals, which reflect a lack of precision in the inference of the diagnostic value to the general population.

Furthermore, discussion about confidence intervals is needed to correct some confusing statements of Drs Lim and Dusmet. Because the confidence level represents the likelihood that another sample will provide the same results (in our case, diagnostic sensitivity), a confidence level of 95% means that 95 out of 100 times, the sample percentages will fall within the confidence intervals. When applied to our study, this means that sensitivity of mediastinoscopy in 95 out of 100 measurements will have a value between 42% and 77%. In our measurement—that is, mediastinoscopy done by our surgical team and pathologic examinations done by our pathologists—sensitivity was exactly 61%. To say that this value “crosses 50%” makes no sense at all. The confidence interval is a way of putting a probability on what future samples (or repeats of the earlier sampling procedure with different users) would yield. In other words, it is a way of saying how repeatable the results are. It does not state the definitive value of the result (sensitivity) in the general population (patients with non-small cell lung cancer after induction therapy).

Throughout the phases of design, data collection, analysis, and report of the study we kept in mind all STARD items and checked the appropriateness of our statements against the STARD guidelines. Including “not biopsied nodal stations” in the formula to calculate the diagnostic accuracy is obvious, compliant with the STARD guidelines, and consistent with the methods of previous literature.³ For example, Drs Lim and Dusmet cited the article of De Leyn and colleagues,² in which sensitivity was calculated by including a considerable number of cases with inadequate

exploration of the subcarinal station (“no biopsy”) in which residual metastases were discovered at further thoracotomy (“false-negative” result of mediastinoscopy). In the evaluation of diagnostic yield of mediastinoscopy, inadequate biopsy of a metastatic lymph node station with a false-negative finding and no biopsy of the station are two faces of the same coin, the inability of the method to rule in disease, regardless of what factor leads to the false-negative result (fibrotic alteration of nodal tissue after previous mediastinoscopy, dense adhesions with anatomic inaccessibility, and so on).

Finally, we remind the readers that the accuracy value of mediastinoscopy was as high as 88%. The ability to rule in residual disease (specificity) was present in 100% of cases. The ability to rule out residual disease (sensitivity) was present in 61%; that is, unnecessary thoracotomy could be avoided for about two thirds of patients.

On the basis of these considerations, we reaffirm our conclusion that mediastinoscopy provides histologic proof of mediastinal staging with high diagnostic accuracy.

Alessandro Marra, MD, PhD^a

Ludger Hillejan, MD^b

Georgios Stamatis, MD^a

Department of Thoracic Surgery and
Endoscopy

Ruhrlandklinik

Essen, Germany^a

Department of Thoracic Surgery

Klinikum St Georg

Ostercappeln, Germany^b

References

1. Polcarpio-Nicolas ML, Wick MR. False-positive interpretations in respiratory cytopathology: exemplary cases and literature review. *Diagn Cytopathol* 2008;36:13-9.
2. De Leyn P, Stroobants S, De Wever W, Lerut T, Coosemans W, Decker G, et al. Prospective comparative study of integrated positron emission tomography-computed tomography scan compared with mediastinoscopy in the assessment of residual mediastinal lymph node disease after induction chemotherapy for mediastinoscopy-proven stage IIIA-N2 non-small-cell lung cancer: a Leuven Lung Cancer Group Study. *J Clin Oncol* 2006;24:3333-9.
3. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.

doi:10.1016/j.jtcvs.2008.08.053

REFERENT VALUES AND EQUIPOISE: EDITORS' NOTES To the Editor:

Lim and Dusmet,¹ in a letter to the editors critiquing an article by Marra and colleagues,² make 4 statements about referent values (accuracy, sensitivity, specificity, positive predictive value, and negative predictive value)³ that require clarification.

Statement 1: “Specificity (ability to rule in disease) is not the major focus for surgeons.”

Referent value specificity is the measure of the ability of a test to rule out disease. It is defined as the number of patients without the disease who are identified by the test (true-negative results) divided by total number of patients without disease (true-negative plus false-positive results).

Statement 2: “What is critical is sensitivity (the ability to rule out disease).”

Referent value sensitivity is a measure of the ability of a test to rule in disease. It is defined as the number of patients with disease who are identified by the test (true-positive results) divided by the total number of patients with the disease (true-positive and false-negative results).

Statement 3: “Mediastinoscopy is associated with 100% specificity. That in itself is expected ... when tumor is seen in a biopsy it is invariably seen on the reference test.”

We believe these statements indicate the authors are confusing specificity with sensitivity. In general, screening tests have low sensitivity but high specificity. As work-up continues and the population is increasingly “enriched” with disease, sensitivity increases and specificity decreases.

However, for the surgeon dealing with this “enriched” population and