Imaging–Histologic Discordance at Percutaneous Biopsy of the Lung

Andrea S. Oh, MD, Corey W. Arnold, PhD, Sitaram Vangala, MS, W. Dean Wallace, MD, Scott J. Genshaft, MD, Karthik Sarma, BS, Denise R. Aberle, MD

Rationale and Objectives: The purpose of this study was to quantify the degree of imaging–histologic discordance in a cohort of patients undergoing computed tomography (CT)–guided lung biopsy for focal lung disease.

Materials and Methods: A retrospective review was performed of 186 patients who underwent percutaneous lung biopsy of a parenchymal lesion at our institution between January and December 2009. Diagnostic radiology reports of CT or positron emission tomography–CTs performed before biopsy were used to classify the lesion as malignant or benign by five readers. Pathology reports of the biopsied lesions were classified by three readers. Inter-reader agreement and imaging–histologic concordance were quantified using kappa statistics. Discordant benign cases were then revisited to determine downstream effects.

Results: Inter-reader agreement on report content was substantial or almost perfect with kappas >0.783. Kappas for concordance were as follows: malignant (0.448), primary lung cancer (0.517), metastatic disease to lung (0.449), benign (0.510), and overall agreement (0.381). Of the twelve discordant benign cases that were revisited, four were found to be false negatives, resulting in a delay in diagnosis.

Conclusions: Our study of imaging–histologic discordance in percutaneous biopsy of lung lesions supports the need for imaging report standardization and improved integration and communication between the fields of radiology and pathology.

Key Words: Radiologic–pathologic correlation; standardized reporting; imaging–histologic discordance.

Radiology and pathology play central roles in cancer diagnosis but typically report findings independently of one another. Independent reporting can increase radiologic–pathologic discordance, defined as a discrepancy between imaging interpretation and histologic findings (1). Radiologic–pathologic correlation has been studied in various imaging specialties to gauge interpretive performance and accuracy, and to identify radiographic features corresponding to histologic findings (2–7). However, few studies have attempted to assess the utility of integrated radiologic–pathologic correlation for establishing imaging–histologic concordance or discordance as a method to prospectively identify missed carcinomas due to biopsy sampling error (8).

Radiologic–pathologic discordance may be categorized as either discordant malignant or discordant benign. The former refers to a lesion that appears radiologically benign, but is malignant on histology; the latter refers to a lesion suspicious for malignancy on imaging but benign histologically (9). In mammography, with the adoption of the Breast Imaging-Reporting and Data System (BI-RADS), this notion of discordance fits naturally due to strict, unambiguous radiologic guidelines governing diagnostic conclusions. BI-RADS provides a framework that allows instances of discordance to receive special consideration such that ostensibly negative pathology in cases of high radiographic suspicion warrant prompt repeat biopsy (10–14).

In contrast to breast imaging, there exists no standardized set of reporting guidelines for thoracic imaging (15). This fact makes the study of discordance challenging as radiology reports can contain more than one diagnosis for a lung lesion (eg, organizing pneumonia vs primary neoplasm) or no diagnoses at all. Discordance resulting from such ambiguity can be confusing to the referring physician because it may obscure the likelihood of malignancy (16). Furthermore, in instances of high suspicion of carcinoma by imaging, a nonspecific benign histologic diagnosis resulting from inadequate tissue sampling could lead to delayed diagnosis of a missed cancer by the referring clinician (17).

By nature, lung cancer imaging is relatively more complex than breast cancer imaging, with greater anatomic and pathologic diversity. It is therefore understandable that lung cancer imaging reports reflect this complexity through differential diagnoses, which may naturally conflict. However, it is nonetheless important to correlate radiology and pathology diagnoses to appraise accuracy in imaging interpretation,

Acad Radiol 2015; 22:481–487

From the Department of Radiology, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Suite 420, Los Angeles, CA 90024 (A.S.O., C.W.A., S.J.G., D.R.A.); Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California (S.V.); Department of Pathology, David Geffen School of Medicine at UCLA, Los Angeles, California (W.D.W.); and David Geffen School of Medicine at UCLA, Los Angeles, California (K.S.). Received June 12, 2014; accepted November 25, 2014. **Address correspondence to:** A.S.O. e-mail: aoh@mednet.ucla.edu

http://dx.doi.org/10.1016/j.acra.2014.11.009

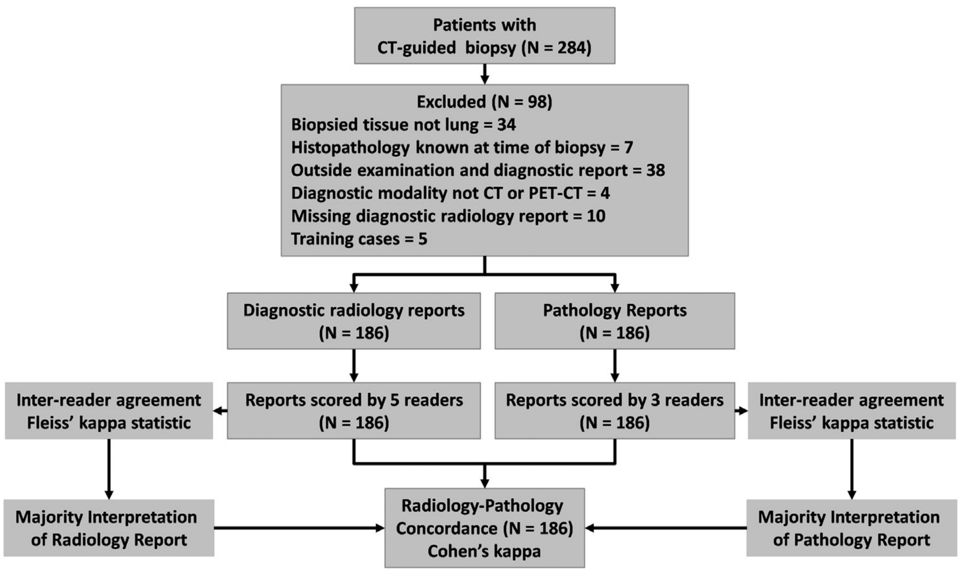

Figure 1. Experimental design. CT, computed tomography; PET, positron emission tomography.

identify potential causes of discordance, and make efforts to resolve these disparities.

To the best of our knowledge, there have been no previous studies to measure discordance in thoracic oncology. In this article, we sought to develop and apply a reliable scale for categorizing radiologic and pathologic findings to quantify the degree of imaging–histologic discordance in a cohort of patients undergoing computed tomography (CT)–guided lung biopsy for focal lung disease.

MATERIALS AND METHODS

With institutional review board approval, 186 patients undergoing image-guided core needle biopsy of the lung were studied retrospectively. Informed consent was waived because of the retrospective nature of the study. The cohort included all individuals in whom a diagnostic CT or positron emission tomography (PET)–CT was acquired and interpreted at our institution within 1 month of percutaneous lung biopsy. We limited the analysis to parenchymal lung lesions to constrain the diversity of thoracic pathology to lesions potentially related to lung carcinoma. The study cohort was established by querying our institutional radiology information system for procedures coded as CT-guided lung biopsy during the calendar year 2009 (Fig 1). In 2010, the thoracic radiology section at our institution began experimenting with various standardized templates for radiology reporting. However, as the need for, and composition of, such templates is debated, they are used by only a subset of radiologists who complete them to varying degrees. Therefore, to establish the degree of discordance during the most recent time period in which all radiologists were reporting in their most "natural" state, our study was limited to the year 2009. The query returned 299 lesions in 284 patients. Of these, 93 patients were excluded for the following reasons: the biopsy was of the tissue other than the lung (n = 34), such as the pleural or chest wall lesions (n = 34); histopathology was known at the time of diagnostic interpretation (n = 7); diagnostic radiology reports were not generated from our institution (n = 38); a diagnostic radiology report was not obtained before biopsy or did not reference the lesion of interest (n = 10); or the diagnostic imaging modality was neither CT nor PET-CT (n = 4). Five randomly selected cases were used to train readers in the classification procedure and were not included in the results. All statistical analyses were performed using R, version 3.0.1 (18).

Inter-reader Agreement on Interpretation Content

Text reports of diagnostic CT or PET-CT examinations rendered by institutional radiologists before biopsy were retrieved and deidentified. Radiologic diagnoses were independently classified as benign or malignant by five readers representing different levels of medical experience to measure the degree of agreement between readers on report content. Given the straightforward nature of the task, nonradiologists were included as readers. The group was composed of a biomedical informatician, a general internist, a pathologist, a radiologist, and a medical student, none of whom had previously reviewed the radiology reports. All readers were blinded to the corresponding pathology results and received standardized instructions on how to classify reports (Supplementary Appendix 1); five training cases were used to ensure an understanding of the classification task. Table 1 lists sample reader scoring. Responses were dichotomous (1 = yes or 0 = no) for each of the four independent determinations: malignant (not otherwise specified), primary lung cancer, metastatic disease, and benign disease. Interobserver agreement on the information content of the report among the five readers was determined using Fleiss kappa statistic, which adjusts the percent agreement for the level of agreement that would be expected entirely due to chance