AMERICAN THORACIC SOCIETY DOCUMENTS

Assessment of Advanced Diagnostic Bronchoscopy Outcomes for Peripheral Lung Lesions: A Delphi Consensus Definition of Diagnostic Yield and Recommendations for Patient-centered Study Designs

An Official American Thoracic Society/American College of Chest Physicians Research Statement

Anne V. Gonzalez, Gerard A. Silvestri, Daniel A. Korevaar, Yaron B. Gesthalter, Nisha D. Almeida, Alex Chen, Chris R. Gilbert, Peter B. Illei, Neal Navani, Mary M. Pasquinelli, Nicholas J. Pastis, Catherine R. Sears, Samira Shojaee, Stephen B. Solomon, Daniel P. Steinfort, Fabien Maldonado, M. Patricia Rivera, and Lonny B. Yarmus; on behalf of the American Thoracic Society Assembly on Thoracic Oncology and the American College of Chest Physicians

This Official Research Statement was approved by the American Thoracic Society December 2023 and the American College of Chest Physicians September 2023

Abstract

Background: Advanced diagnostic bronchoscopy targeting the lung periphery has developed at an accelerated pace over the last two decades, whereas evidence to support introduction of innovative technologies has been variable and deficient. A major gap relates to variable reporting of diagnostic yield, in addition to limited comparative studies.

Objectives: To develop a research framework to standardize the evaluation of advanced diagnostic bronchoscopy techniques for peripheral lung lesions. Specifically, we aimed for consensus on a robust definition of diagnostic yield, and we propose potential study designs at various stages of technology development.

Methods: Panel members were selected for their diverse expertise. Workgroup meetings were conducted in virtual or hybrid format. The cochairs subsequently developed summary statements, with voting proceeding according to a modified Delphi process. The statement was cosponsored by the American Thoracic Society and the American College of Chest Physicians.

Results: Consensus was reached on 15 statements on the definition of diagnostic outcomes and study designs. A strict definition of diagnostic yield should be used, and studies should be reported according to the STARD (Standards for Reporting Diagnostic Accuracy Studies) guidelines. Clinical or radiographic follow-up may be incorporated into the reference standard definition but should not be used to calculate diagnostic yield from the procedural encounter. Methodologically robust comparative studies, with incorporation of patient-reported outcomes, are needed to adequately assess and validate minimally invasive diagnostic technologies targeting the lung periphery.

Conclusions: This American Thoracic Society/American College of Chest Physicians statement aims to provide a research framework that allows greater standardization of device validation efforts through clearly defined diagnostic outcomes and robust study designs. High-quality studies, both industry and publicly funded, can support subsequent health economic analyses and guide implementation decisions in various healthcare settings.

Keywords: pulmonary nodule; lung cancer; advanced diagnostic bronchoscopy; diagnostic accuracy; diagnostic yield

ORCID IDs: 0000-0002-2295-2948 (A.V.G.); 0000-0002-8464-2187 (G.A.S.); 0000-0002-7979-7897 (D.A.K.); 0000-0002-2762-6404 (Y.B.G.); 0009-0002-8379-9649 (N.D.A.); 0000-0002-9351-5075 (A.C.); 0000-0003-3868-4578 (C.R.G.); 0000-0003-4440-9265 (P.B.I.); 0000-0002-6412-7516 (N.N.); 0000-0002-1015-1908 (M.M.P.); 0009-0005-0261-5986 (N.J.P.); 0000-0002-5797-3458 (C.R.S.); 0000-0001-9228-209X (S.S.); 0000-0001-5108-1021 (S.B.S.); 0000-0002-8998-2949 (D.P.S.); 0000-0002-6504-2063 (F.M.); 0000-0003-3827-5010 (M.P.R.); 0000-0002-2991-7758 (L.B.Y.).

This document was funded by the American Thoracic Society.

This document has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 209, Iss 6, pp 634–646, Mar 15, 2024

Originally Published in Press as DOI: 10.1164/rccm.202401-0192ST on February 23, 2024
Internet address: www.atsjournals.org

Contents

Summary Statements

Definition of Outcomes

1. There is considerable variation in how diagnostic yield is calculated across published studies. It is recommended that a strict definition of diagnostic yield be used. The numerator should include all patients with peripheral pulmonary nodules in whom the result of a minimally invasive diagnostic procedure establishes a specific benign or malignant diagnosis that is sufficient to inform patient care, whereas the denominator should include all patients in whom the minimally invasive procedure was attempted or performed.
2. Typical diagnostic accuracy measures (sensitivity, specificity, positive predictive value, and negative predictive value) refer to the ability of a test to identify a single target condition (e.g., presence or absence of lung cancer). In contrast, peripheral pulmonary nodules may be caused by a range of benign and malignant conditions. Therefore, when evaluating the performance of a test for peripheral pulmonary nodules, diagnostic yield is a more informative and recommended primary outcome, because it refers to the ability of a test to establish a specific pathological diagnosis that is sufficient to inform patient care. All series reporting diagnostic accuracy measures should clearly report the prevalence of malignancy to enhance ease of comparison across studies.
3. Studies on the diagnostic performance of advanced bronchoscopy techniques should be reported according to the STARD (Standards for Reporting Diagnostic Accuracy Studies) guidelines, with inclusion of a detailed flow diagram. A STARD-inspired flowchart, adapted for studies of minimally invasive diagnostic procedures for peripheral pulmonary nodules, is proposed.
4. Diagnostic outcome reporting of bronchoscopic studies targeting the lung periphery should be standardized to allow comparison and/or pooling of results across studies. Reporting of results should be detailed, transparent, and allow calculation of diagnostic yield using the strict definition.
5. Minimally invasive diagnostic procedures targeting the lung periphery include advanced diagnostic bronchoscopy procedures and computed tomography (CT)-guided transthoracic biopsies. Diagnostic outcome reporting of CT-guided transthoracic biopsies targeting the lung periphery should be similarly standardized to allow comparison and/or pooling of results across studies. Reporting of results should be detailed, transparent, and allow calculation of diagnostic yield using the strict definition.
6. For determination of sensitivity for malignancy, clinical or radiographic follow-up that occurs after the bronchoscopy or CT-guided biopsy is completed and pathology/microbiology results are available may be incorporated into the definition of the reference standard to determine whether the peripheral pulmonary nodule targeted for biopsy was indeed benign or malignant. This clinical follow-up should span a minimum of 12 months. However, clinical and radiographic follow-up should not be used to calculate diagnostic yield from the procedural encounter, as per the strict definition of diagnostic yield.
7. Nonspecific pathological findings suggestive but not diagnostic of malignancy (e.g., atypia, suspicious cells) are considered nondiagnostic in the calculation of diagnostic yield.
   a. There is a need to better understand patient outcomes after nonspecific pathological findings suggestive but not diagnostic of malignancy.
   b. We advocate for data through multicenter collaborative efforts to better understand the impact and downstream events (e.g., additional procedures, length and extent of surveillance, ultimate diagnosis, and/or stage shift in those diagnosed with cancer) for patients whose initial procedure yielded nonspecific pathological findings suggestive but not diagnostic of malignancy.
8. Nonspecific pathological findings suggestive but not diagnostic of specific benign etiologies (e.g., nonspecific inflammation) are considered nondiagnostic in the calculation of diagnostic yield.
   a. There is a need to better understand patient outcomes after nonspecific pathological findings suggestive but not diagnostic of a specific benign etiology.
   b. We advocate for data through multicenter collaborative efforts to better understand the impact and downstream events (e.g., additional/salvage procedures, extent of surveillance, ultimate diagnosis) for patients whose initial procedure yielded nonspecific pathological findings suggestive but not diagnostic of a benign etiology.

Study Designs

9. Different study designs may be most appropriate at various stages of innovation of new technologies focused on bronchoscopic peripheral lung nodule sampling. We propose adoption of the IDEAL (idea, development, exploration, assessment, long-term study) collaboration framework for evaluation of novel technology targeting the lung periphery.
10. Methodologically robust comparative studies, preferentially randomized controlled trials, are needed to adequately assess and validate minimally invasive diagnostic technologies targeting the lung periphery.
11. Aside from clear diagnostic performance and safety outcomes as primary endpoints, additional methodological factors should be clearly established in the study protocols. Examples include setting of recruitment, patient selection, target characteristics, and other relevant factors related to the index test. These factors should be

recorded and reported according to STARD guidelines.
12. Patient-centered and patient-reported outcomes should be studied and incorporated into the evaluation of novel bronchoscopy techniques, including the following:
    a. Better characterization of downstream patient events (e.g., need for salvage procedures, complications and/or hospitalizations, delays in diagnosis)
    b. Consideration of patient preferences for differing diagnostic pathways
13. Inclusion of academic and community centers serving diverse patient populations, including underrepresented and underserved patients, is strongly encouraged.
14. Informative and inclusive study designs can support subsequent health economic analyses, guide implementation decisions in various healthcare settings, and help address disparities such as access to technological innovations.
15. There is a need for funding mechanisms for high-quality studies of novel technology for sampling of the lung periphery. These include the following:
    a. Pathways for improved partnership with industry
    b. Wider commitment for industry validation and postmarket reporting
    c. Societal partnerships

Introduction

The evaluation of peripheral pulmonary nodules is an increasingly common clinical problem that remains a challenge for all chest physicians. The increasingly routine use of computed tomography (CT) in clinical practice has resulted in over 1.5 million adult Americans having an incidental pulmonary nodule identified each year (1). Lung cancer screening with low-dose CT is also associated with a high detection rate of small lung nodules that warrant further evaluation (2) and may necessitate invasive diagnostic testing. In addition, new software techniques using deep learning are expected to further increase the number of pulmonary nodules detected on routine imaging (3).

Conventional flexible bronchoscopy with reliance on fluoroscopy alone has a limited role in the evaluation of peripheral pulmonary nodules (4). Over the last two decades, advanced diagnostic bronchoscopy has developed at an accelerated pace, and pulmonologists are faced with a wide array of techniques and tools to reach and sample peripheral lesions. Tools to access the periphery include thin and ultrathin conventional bronchoscopes, radial endobronchial ultrasound, virtual bronchoscopy, transparenchymal nodule access, electromagnetic navigation bronchoscopy, and robotic-assisted bronchoscopy. In addition, there have been advances in near-time image guidance technologies such as tomosynthesis, augmented fluoroscopy, and cone-beam CT guidance (5, 6).

Certain advances in diagnostic bronchoscopy, such as endobronchial ultrasound-guided transbronchial needle aspiration, have been widely adopted and are first-line procedures for diagnosing central lesions and staging lung cancer (7). Translation of early published experience into successful clinical use has been more variable for advanced bronchoscopy targeting the lung periphery. The evidence base to support introduction of innovative bronchoscopy techniques into clinical practice has been limited, mostly consisting of single-center experience at high-volume expert centers (8, 9) with variable definitions of what constitutes a diagnostic procedure (10, 11). Comparative studies of different technologies or approaches are rare (6, 12, 13). As a result, the evidence base for peripheral bronchoscopy is fraught with uncertain outcomes and limited generalizability (12, 14, 15).

A major issue in the diagnostic bronchoscopy literature relates to variable reporting of diagnostic yield. Differing definitions of what constitutes a diagnostic procedure render comparisons between technologies nearly impossible. The reported diagnostic yield of peripheral bronchoscopy varies widely, ranging from 39% to 88% (12, 14, 16–19). In a recent study of a hypothetical cohort of 1,000 patients generated from a prospective study of robot-assisted bronchoscopy, diagnostic yield was calculated using three definitions (strict, intermediate, and liberal), and reported yields varied by 20% on this basis alone (20). An updated systematic review of 126 studies on guided bronchoscopy in over 16,000 patients with lung nodules compiled diagnostic yield as reported by authors; when risk of bias was examined according to QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) criteria, absolute yield was 5% lower in low versus high risk of bias studies (21). In addition, there was marked heterogeneity of studies in which multiple guidance tools were used in a single procedural encounter, making pooling of results at times difficult to interpret. A prospective study by Thiboutot and colleagues investigating bronchoscopy with electromagnetic guidance for peripheral lung lesions reported a diagnostic yield of 49%, using a strict definition (22). This outcome is far inferior to the frequently quoted yield estimates from studies using more liberal definitions. Efforts to report diagnostic outcomes accurately and consistently are an important first step toward generating comparable data and require a robust definition of diagnostic yield (23).

It has been difficult to establish the true efficacy of various devices, and this may, in part, be because of how technology is cleared for commercialization and introduced into the healthcare setting. The vast majority of medical devices in the United States are cleared by the U.S. Food and Drug Administration through the 510(k) pathway, designed to allow commercialization of medical devices deemed "substantially equivalent" in safety and effectiveness to already commercialized, so-called predicate devices (24–26). The more stringent premarket approval pathway, which allows formal approval (as opposed to "clearance") of new medical devices, is similar to the regulatory pathway for pharmaceuticals and requires demonstration of clinical safety and effectiveness. This pathway, however, only pertains to class 3 devices, defined as devices that "usually sustain or support life, are implanted, or present potential unreasonable risk of illness or injury," such as implantable cardiac defibrillators. Thus, diagnostic bronchoscopy platforms have not traditionally required clinical studies or evidence of improved patient outcomes for commercialization, shifting the responsibility for providing useful clinical data from medical device manufacturers to clinicians (27).

Lack of evidence for improved patient outcomes is a barrier to the introduction of novel technology in both "pluralistic" and single-payer systems. The process of introduction of new technology may vary, depending on the country and healthcare system. In general, however, the process typically takes several steps, informed by

knowledge derived from clinical and health economics research. New medical technologies should undergo rigorous clinical testing and evaluation to establish their safety and effectiveness. Regulatory authorities must then review evidence to ensure the technology meets the necessary standards for medical use. Regulatory approval is typically followed by a health technology assessment (28), which considers the best available evidence on the clinical effectiveness, safety, and cost-effectiveness of the novel technology compared with existing options. Poor-quality research can undermine decision makers' confidence in the safety, effectiveness, and cost-effectiveness of new technologies, making it more difficult to introduce these technologies in single-payer systems.

This statement proposes a research framework to standardize the evaluation of advanced diagnostic bronchoscopy techniques for peripheral lung lesions and ultimately provide the level of evidence required for successful and rational integration of novel technologies into clinical practice. Specifically, we propose a robust consensus definition of diagnostic yield and discuss potential study designs for evaluation of novel bronchoscopy technology at various stages of introduction. Cost-effectiveness analyses are beyond the scope of this project, which aims to ensure that better effectiveness data are available before broad uptake of novel bronchoscopy techniques. Additional considerations include transparent reporting of relevant patient, operator, and procedural factors and the incorporation of patient-reported outcomes into the evaluation of technological innovations.

Methods

The present statement received approval as an American Thoracic Society Thoracic Oncology Assembly project in January 2022, with subsequent cosponsorship by the American College of Chest Physicians in April 2022. During the initial phase of the project, the cochairs (A.V.G., G.A.S., and L.B.Y.) reviewed the current literature on diagnostic outcome measures and optimal study designs as relevant to advanced diagnostic bronchoscopy. An informal discussion between cochairs and panel member attendees took place during the 2022 American Thoracic Society international conference.

Panel members were selected for their complementary expertise while ensuring sex and geographic diversity. All panelists were vetted for potential conflicts of interest. A total of 19 panelists were invited to participate, and 18 accepted. The following specialties are represented: pulmonology/interventional pulmonology, 11; thoracic oncology, 4; epidemiology and health technology assessment, 1; thoracic pathology, 1; interventional radiology, 1. One-third of panelists were women, and one-third were from outside the United States.

All meetings were conducted in virtual (www.zoom.com) or hybrid format. An initial meeting was held in July 2022 to introduce the project and assign subgroups and tasks. Between September and November 2022, three workgroup meetings were held to tackle the challenges of technology evaluation and introduction in various healthcare settings, improved reporting of diagnostic accuracy studies, and optimal study designs. After these discussions, the cochairs developed summary statements (definition of outcomes and study designs), which were circulated for comments, and revisions were made on the basis of panel suggestions before voting.

Voting proceeded according to a modified Delphi process. Participants anonymously voted virtually (or by e-mail) to state their agreement or disagreement, using a 5-point Likert scale (strongly disagree, somewhat disagree, neither agree nor disagree, somewhat agree, strongly agree). Responses were scored agree (strongly agree, somewhat agree), neutral (I neither agree or disagree), or disagree (strongly disagree, somewhat disagree). If consensus had not been reached during the first round, the plan was to review feedback and re-present the statements. All data were collected prospectively using the zoom.com software (virtual voting), which was exported into Microsoft Excel (Microsoft) for analysis after adding e-mail vote results. Consensus was defined *a priori* as a greater than or equal to 80% agreement response to a question.

Results

Consensus was reached on all statements after the first round of voting. The 15 statements and associated vote results are detailed in Table E1 in the online supplement.

Definition of Outcomes

Statement 1. There is considerable variation in how diagnostic yield is calculated across published studies. It is recommended that a strict definition of diagnostic yield be used. The numerator should include all patients with peripheral pulmonary nodules in whom the result of a minimally invasive diagnostic procedure establishes a specific benign or malignant diagnosis that is sufficient to inform patient care, and the denominator should include all patients in whom the minimally invasive procedure was attempted or performed.

The investigation of peripheral lung nodules seeks to achieve early detection of lung cancer or other malignancies but may also yield alternative noncancer diagnoses (e.g., granulomatous inflammation or other benign lesions), which must be clearly distinguished from nonspecific biopsy results. The use of variable definitions of diagnostic yield has plagued the advanced diagnostic bronchoscopy literature (10, 14), making valid comparisons of innovative technologies across studies difficult if not impossible. In the numerator, some studies only include procedures that resulted in a specific malignant or benign pathological diagnosis (i.e., "strict" definition of diagnostic yield), whereas others also include procedures with nonspecific benign diagnoses that turn out to be nonmalignant at follow-up (i.e., "intermediate" definition) or even nondiagnostic procedures (e.g., normal lung tissue) that turn out to be nonmalignant at follow-up (i.e., "liberal" definition). It is particularly problematic that the so-called intermediate and liberal definitions of diagnostic yield incorporate clinical follow-up into the index bronchoscopy result.

A recent systematic review on the diagnostic performance of navigation bronchoscopy in peripheral lung nodules included 95 studies, of which only 26% used a strict definition of diagnostic yield, 32% an intermediate definition, and 26% a liberal definition. In 16% of studies, the definition used was unclear (29). The impact of variable definitions of diagnostic yield was demonstrated in the simulation study by Vachani and colleagues, in which different definitions used in the same hypothetical dataset resulted in estimates that differed by more than 20% (20).

The AQuIRE (American College of Chest Physicians Quality Improvement Registry, Evaluation, and Education) registry considered

a bronchoscopy procedure diagnostic if a specific malignant or specific benign diagnosis was established with the biopsy results. In contrast, the NAVIGATE (Clinical Evaluation of superDimension Navigation System for Electromagnetic Navigation Bronchoscopy) study (10) reported a 12-month diagnostic yield, incorporating the results of a 12-month CT follow-up into the interpretation of biopsy results. This is akin to "revisionist history" in the presence of nonspecific inflammatory or nondiagnostic findings at the index bronchoscopy (23, 30). Importantly, this "12-month" diagnostic yield erroneously incorporates an element of the reference standard (CT follow-up) into the definition of the index test result, a problem also associated with the so-called intermediate and liberal definitions of diagnostic yield. Table 1 summarizes the proposed definitions for diagnostic outcome measures. We recommend against using so-called intermediate or liberal definitions of diagnostic yield, because a level of uncertainty remains in patients without a specific diagnosis at the time the procedure is performed (i.e., some patients with a "nonspecific diagnosis: benign" or "nondiagnostic" result may be malignant after follow-up, and these procedures should not be considered as diagnostic).

With the present recommendation comes a significant shift in perspective and recommendation on the length of follow-up needed for bronchoscopic studies. Historically a follow-up time of 12–24 months has been incorporated into the "intermediate" and "liberal" definitions of diagnostic yield discussed above. By mandating a strict definition of diagnostic yield, the procedure results are finalized at the time of pathology/microbiology result reporting, potentially allowing prompt reporting of preliminary investigative procedural results.

Sufficient follow-up, however, is still required as part of the definition of the refence standard and the assessment of device- and procedure-related complications. Adequate follow-up will also inform efforts to better understand the implications of nondiagnostic bronchoscopy procedures but should no longer be incorporated into the factors needed for diagnostic yield reporting.

Statement 2. Typical diagnostic accuracy measures (sensitivity, specificity, positive predictive value, and negative predictive value) refer to the ability of a test to identify a single target condition (e.g., presence or absence of lung cancer). In contrast, peripheral pulmonary nodules may be caused by a range of benign and malignant conditions. Therefore, when evaluating the performance of a test for peripheral pulmonary nodules, diagnostic yield is a more informative and recommended primary outcome, because it refers to the ability of a test to establish a specific pathological diagnosis that is sufficient to inform patient care. All series reporting diagnostic accuracy measures should clearly report the prevalence of malignancy to enhance ease of comparison across studies.

Diagnostic yield will vary significantly according to the underlying prevalence of malignant disease, because establishing a pathological diagnosis of malignancy is generally more straightforward than establishing a specific, benign pathological diagnosis. Therefore, all series should clearly report the prevalence of malignancy to enhance ease of comparison across studies.

Typical diagnostic accuracy measures such as sensitivity and negative predictive value are important in estimating the

Table 1. Definitions of Diagnostic Outcome Measures

Diagnostic Outcome Measure	Definition
Diagnostic yield	The proportion of all individuals undergoing the diagnostic procedure under evaluation in whom a specific malignant or benign diagnosis is established The numerator should include all patients with peripheral pulmonary nodules in whom the result of a minimally invasive diagnostic procedure establishes a specific benign or malignant diagnosis that is sufficient to inform patient care, and the denominator should include all patients in whom the minimally invasive procedure was attempted or performed
Diagnostic accuracy measures*	
Sensitivity	The proportion of individuals with the disease of interest (e.g., cancer), as determined by the reference standard, who have a positive test result (i.e., malignancy based on bronchoscopy samples)
Specificity†	The proportion of individuals without the disease of interest (e.g., cancer), as determined by the reference standard, who have a negative test result (i.e., no evidence of malignancy based on bronchoscopy samples)
Positive predictive value†	The likelihood that a patient has the target disease (e.g., cancer), given a positive test result (i.e., malignancy based on bronchoscopy samples)
Negative predictive value	The likelihood that a patient does not have the target disease (e.g., cancer), given a negative test result (i.e., no evidence of lung malignancy based on bronchoscopy samples)
Diagnostic accuracy‡	The proportion of all test results that are correct for the condition of interest. For example, all correctly identified patients with or without cancer, divided by all patients who underwent advanced diagnostic bronchoscopy

*Diagnostic accuracy measures include sensitivity, specificity, positive predictive value, and negative predictive value. Please refer to epidemiology textbooks for calculation of these measures using classical 2×2 tables (59, 60).
†In studies of diagnostic bronchoscopy procedures, specificity and positive predictive value are often considered 100% (i.e., no false-positive malignancy results), and a reference standard is generally not systematically performed in patients with positive test results, making the reporting of these outcomes somewhat redundant.
‡Diagnostic accuracy has limited clinical usefulness, because it is an aggregate measure that varies with disease prevalence and within which key information on sensitivity and predictive value is buried (61).

likelihood of having missed cancer in the presence of a nonmalignant bronchoscopy result. Addressing this question requires a clearly established reference standard, which may include clinical follow-up of appropriate duration.

Statement 3. Studies on the diagnostic performance of advanced bronchoscopy techniques should be reported according to the STARD (Standards for Reporting Diagnostic Accuracy Studies) guidelines, with inclusion of a detailed flow diagram. A STARD-inspired flowchart, adapted for studies of minimally invasive diagnostic procedures for peripheral pulmonary nodules, is proposed.

Statement 4. Diagnostic outcome reporting of bronchoscopic studies targeting the lung periphery should be standardized to allow comparison and/or pooling of results across studies. Reporting of results should be detailed, transparent, and allow calculation of diagnostic yield using the strict definition.

Incomplete reporting of studies is considered a major source of avoidable research waste (31). When crucial information is missing from a study report, the study may not be reproducible, and the identification of potential sources of bias is hampered (32, 33). Unfortunately, reporting is often suboptimal for diagnostic accuracy studies (34, 35). STARD (36–38) is a reporting guideline for diagnostic accuracy studies providing a checklist of essential items to ensure fully informative study reports. STARD was first published in 2003 and updated in 2015 and has been endorsed by hundreds of biomedical journals, indicating that they require or recommend adherence of diagnostic accuracy studies to STARD.

Studies of advanced bronchoscopy targeting the lung periphery should adhere to STARD to allow comparisons between studies and technologies (36, 37). This includes clear reporting of which patients were included in the numerator and denominator of the diagnostic yield calculation (STARD 2015, item 14), details on the reference standard used (item 10b), and what was done in case of indeterminate or missing bronchoscopy or reference standard results (items 15 and 16). The authors should specify the flow of participants through the study (item 19) through inclusion of a detailed flowchart. A STARD-inspired flowchart, adapted for studies of minimally invasive diagnostic procedures for peripheral pulmonary nodules, is provided in Figure 1 (36, 37). On the basis of a complete flowchart, readers can determine at a glance how often bronchoscopy failed to localize and/or sample the nodule or was diagnostic (i.e., specific malignant or benign diagnosis). Readers can also assess how often a reference standard was performed and what the reference standard results were (i.e., malignant, benign, or indeterminate). This information will allow readers to (re)calculate the remaining risk of malignancy in patients whose bronchoscopy yielded a nonspecific or nondiagnostic result.

Statement 5. Minimally invasive diagnostic procedures targeting the lung periphery include advanced diagnostic bronchoscopy procedures and CT-guided transthoracic biopsies. Diagnostic outcome reporting of CT-guided transthoracic biopsies targeting the lung periphery should be similarly standardized to allow comparison and/or pooling of results across studies. Reporting of results should be detailed, transparent, and allow calculation of diagnostic yield using the strict definition.

The use of variable definitions of diagnostic yield with insufficient clarity in reporting is not limited to studies of advanced diagnostic bronchoscopy for peripheral pulmonary lesions.

A recent meta-analysis pooled 24,668 transthoracic needle biopsies (TTNBs) in 106 studies and reported an incidence of nondiagnostic biopsies of 6.8% (95% confidence interval, 6–7.6%) (39). The findings of malignancy, atypical cells, specific benign results (6.9%), and nonspecific benign results (14.2%) were considered separate categories. The authors considered the following specific benign diagnoses: benign lung tumors, infectious pneumonia, pulmonary tuberculosis, silicosis, vasculitis, or "others." Nonspecific benign disease included acute or chronic nonspecific inflammation, granuloma, focal fibrosis, or a specimen without evidence of malignancy. The proportion of nodules proven malignant at follow-up ranged from 1.5% for specific benign results to 20.6% for nonspecific benign results and up to 59.3% of nondiagnostic biopsies. Simple exclusion of nondiagnostic results from diagnostic accuracy measure calculations is inappropriate, because it leads to an overestimation of diagnostic accuracy; including nondiagnostic results decreased sensitivity by 4.5% and specificity by 10.7%, on average (39).

Future studies of transthoracic needle aspiration and/or TTNB should be reported according to clear definitions of diagnostic outcomes and with sufficient granularity to allow accurate comparisons of diagnostic yield.

Statement 6. For determination of sensitivity for malignancy, clinical or radiographic follow-up that occurs after the bronchoscopy or CT-guided biopsy is completed and pathology/microbiology results are available may be incorporated into the definition of the reference standard to determine whether the peripheral pulmonary nodule targeted for biopsy was indeed benign or malignant. This clinical follow-up should span a minimum of 12 months. However, clinical and radiographic follow-up should not be used to calculate diagnostic yield from the procedural encounter, as per the strict definition of diagnostic yield.

The most common question after a negative bronchoscopy is, "Is it cancer?" Subsequent clinical decision making is based on the answer to this question. The question can be addressed only through adequate clinical follow-up and transparent reporting, with calculation of the remaining risk of malignancy in those with specific benign, nonspecific benign diagnoses, and nondiagnostic bronchoscopy procedures. Using the proposed flowchart (Figure 1), the remaining risk of malignancy in those without a specific diagnosis is calculated as the sum of procedures with a "final diagnosis: malignant" in those with a "nonspecific diagnosis: benign" or "nondiagnostic" result (numerator), divided by the sum of all patients with a "reference standard available" in these two subgroups (denominator).

Importantly, nonspecific bronchoscopy must be considered nondiagnostic, regardless of 12–24-month follow-up for diagnostic yield calculation. Capturing the proportion of patients who undergo treatment (e.g., radiotherapy) after a nondiagnostic biopsy is imperative. However, nondefinitive bronchoscopy results (including atypia or suspicious cells) that lead to empiric treatment for presumed cancer must be considered nondiagnostic.

Statement 7. Nonspecific pathological findings suggestive but not diagnostic of malignancy (e.g., atypia, suspicious cells) are considered nondiagnostic in the calculation of diagnostic yield.

a. There is a need to better understand patient outcomes after nonspecific pathological findings suggestive but not diagnostic of malignancy.

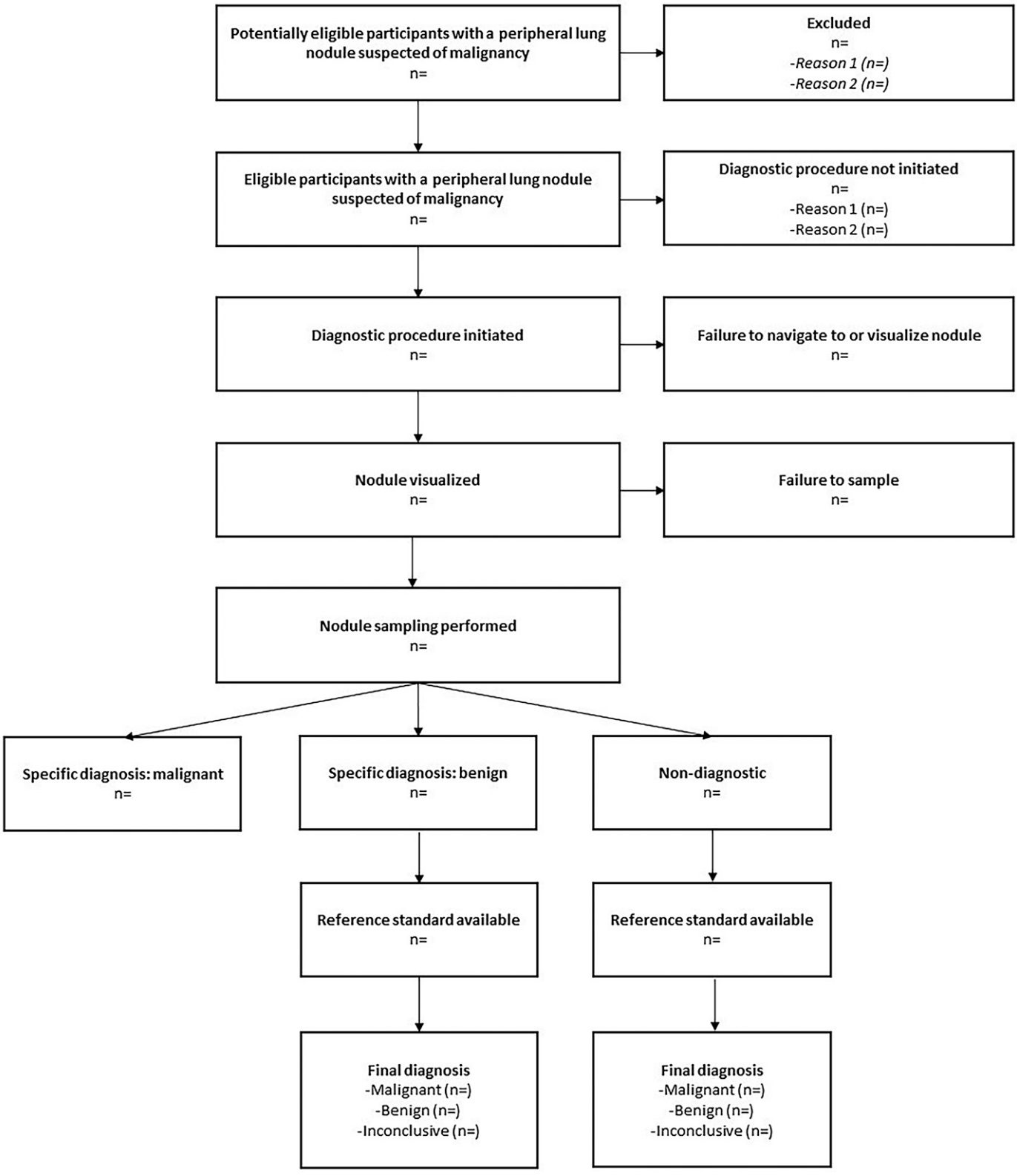

Figure 1. Standards for Reporting Diagnostic Accuracy Studies-inspired flow diagram, adapted for studies of minimally invasive diagnostic procedures for peripheral pulmonary nodules (36, 37). Diagnostic yield is calculated as the sum of "specific diagnosis: malignant" and "specific diagnosis: benign" (in the numerator), divided by "procedures performed" (in the denominator). The remaining risk of malignancy in those without a specific diagnosis is subsequently calculated as the sum of procedures with a "final diagnosis: malignant" in those with a "nondiagnostic" result (in the numerator), divided by the sum of patients with a "reference standard available" in the "nondiagnostic" result group (in the denominator). The reference standard in many cases is clinicoradiological follow-up.

b. We advocate for data through multicenter collaborative efforts to better understand the impact and downstream events (e.g., additional procedures, length and extent of surveillance, ultimate diagnosis, and/or stage shift in those diagnosed with cancer) for patients whose initial procedure yielded nonspecific pathological findings suggestive but not diagnostic.

Statement 8. Nonspecific pathological findings suggestive but not diagnostic of specific benign etiologies (e.g., nonspecific inflammation) are considered nondiagnostic in the calculation of diagnostic yield.

a. There is a need to better understand patient outcomes after nonspecific pathological findings suggestive but not diagnostic of a specific benign etiology.
b. We advocate for data through multicenter collaborative efforts to better understand the impact and downstream events (e.g., additional/salvage procedures, extent of surveillance, ultimate diagnosis) for patients whose initial procedure yielded nonspecific pathological findings suggestive but not diagnostic of a benign etiology.

Panel members extensively discussed the question of what constitutes a diagnostic bronchoscopy procedure, particularly as relates to specific benign diagnoses. There was agreement that less common but specific benign pathological diagnoses, such as benign lung tumors (e.g., hamartoma, lipoma), vasculitis, or amyloidosis, constituted specific benign diagnoses. Similarly, cytopathological findings suggestive of a mycotic infection or tuberculosis constitute specific benign diagnoses.

Cytopathological findings suggestive of bacterial infection in the presence of congruent culture results may also be considered a specific benign diagnosis in the appropriate clinical context. However, a positive bacterial culture finding on BAL fluid may occur in the presence of malignant cytopathology findings, and this finding in isolation would not typically constitute a specific benign diagnosis (e.g., BAL findings designated post-procedure as a contaminant and not treated with appropriate antimicrobial treatment). Ultimately, transparent reporting is essential to ensure fair comparisons and pooling of results.

The finding of granulomatous inflammation, whether necrotizing or nonnecrotizing, is also believed to constitute a specific benign diagnosis. In the systematic review by Chae and colleagues, granulomatous inflammation was considered a nonspecific benign result, together with acute or chronic nonspecific inflammation and focal fibrosis (39). Kim and colleagues, however, reviewed nonspecific pathological results in 226 TTNB samples and reported that the finding of granulomatous inflammation was a robust indicator of a true negative biopsy result, with no diagnoses of cancer among the 81 patients with granulomatous inflammation (40). Although granulomatous inflammation has been associated with certain types of malignancy (Hodgkin's disease, testicular cancer) (41), the chance of missing malignancy in the presence of clear granulomatous inflammation on a lung cytopathological specimen is low.

Table 2 summarizes what constitutes a specific malignant or benign diagnosis versus a nondiagnostic procedure result. Nondiagnostic results refer to specimens showing only blood, necrosis, normal lung parenchyma, or insufficient tissue to make any diagnosis (39). The findings of atypia and nonspecific pathological findings, including nonspecific inflammation, also constitute nondiagnostic results. These definitions may evolve through thoughtful data collection and continued advances in diagnostic testing of small biopsy samples. We recognize that bronchoscopy results may not all fit neatly into the proposed boxes, which is why clinical context should be considered and transparently reported. Figure 2 illustrates the classification of biopsy results into diagnostic and nondiagnostic categories.

Clinical decision making downstream of the index diagnostic bronchoscopy may provide additional clarity in adjudicating bronchoscopy results as diagnostic versus nondiagnostic. Consider a patient with nonspecific inflammation in bronchoscopy samples referred for surgical resection due to a high suspicion of malignancy. Even if the surgical pathology reveals a benign nodule etiology, the bronchoscopy result cannot be considered diagnostic, because no patient would be referred to surgery had the bronchoscopy convincingly confirmed a benign lung nodule etiology. A bronchoscopy that does not provide a specific benign or malignant result sufficient to inform patient care simply cannot be considered a diagnostic procedure.

Study Designs

Statement 9. Different study designs may be most appropriate at various stages of innovation of new technologies focused on bronchoscopic peripheral lung nodule sampling. We propose adoption of the IDEAL (idea, development, exploration, assessment, long-term study) collaboration framework for evaluation of novel technology targeting the lung periphery.

The evidence to support using current advanced diagnostic bronchoscopy techniques clinically remains weak overall, consisting mainly of single-center, retrospective series. The inherent challenges to robust evaluation of clinical innovations are not limited to interventional pulmonary

Table 2. Examples of Specific Malignant and Benign Diagnoses versus Nondiagnostic Test Results

Specific Malignant Diagnoses	Specific Benign Diagnoses	Nondiagnostic Test Results
Non–small cell lung cancer	Benign lung tumors (e.g., hamartoma, lipoma, papilloma)	Atypia/atypical cells
Adenocarcinoma	Organizing pneumonia	Acute or chronic inflammation
Squamous cell carcinoma	Granulomatosis with polyangiitis	Normal lung parenchyma
Large cell carcinoma	Fungal (e.g., *Aspergillus*, *Histoplasma*)	Insufficient tissue for diagnosis
Small cell carcinoma	*Mycobacterium* (e.g., *M. tuberculosis*)	Blood
Carcinoma metastatic to lung (e.g., melanoma, breast cancer)	Bacterial (e.g., *Streptococcus pneumoniae*)	Necrosis
Typical/atypical carcinoid tumor	Granuloma/granulomatous inflammation	—

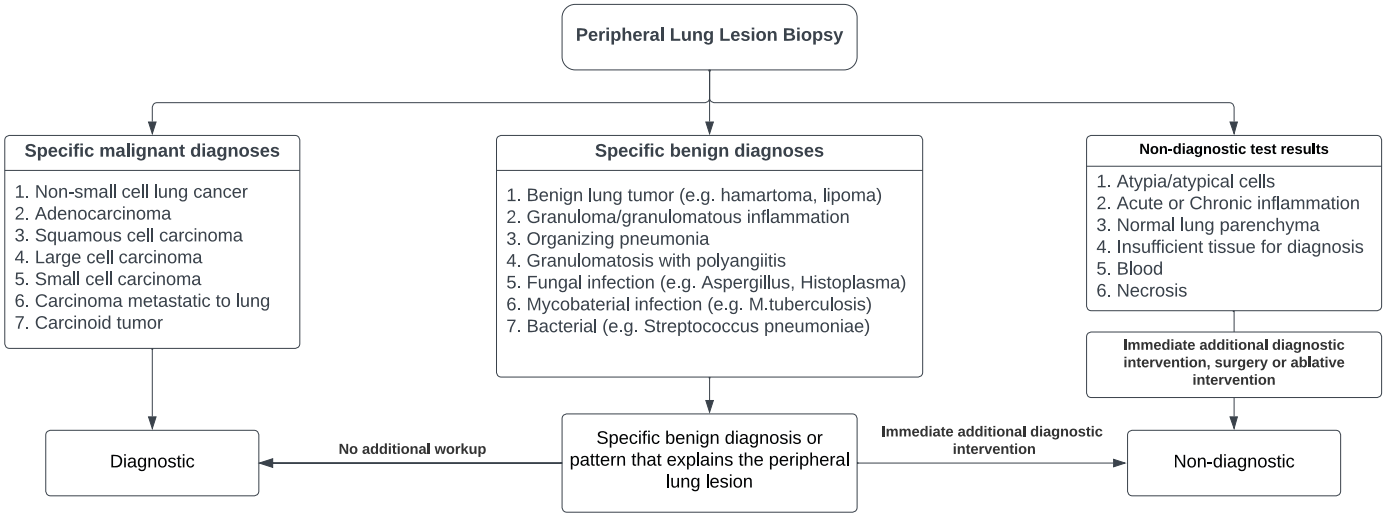

Figure 2. Algorithmic approach to classifying peripheral lung biopsy results.

medicine and extend to surgery and other interventional fields (42).

The IDEAL Collaboration stems from a series of exchanges between surgeons and methodologists at Balliol College, Oxford, that examined the difficulties of conducting high-quality trials in surgery and potential solutions (43–45). The IDEAL framework identifies the various stages of (surgical) innovation (idea, development, exploration, assessment, and long-term study) and proposes optimal study designs for conducting evaluation at each stage. Table 3 outlines key aspects of the IDEAL framework, with examples drawn from the advanced diagnostic bronchoscopy literature.

The idea stage (1) focuses on demonstrating whether the device can safely achieve its intended goal. Mandatory reporting of all "first-in-human" procedures is recommended, and anonymous reporting should be permitted to avoid repeated failures by others (46). The IDEAL-D framework, adapted for medical devices, adds a preclinical phase 0 for device development, which may include cadaver and/or animal studies (47). In the development (2a) stage, modifications to the procedure or patient selection criteria are captured through sequential, detailed reporting beyond safety and efficacy outcomes. The exploration stage (2b) is frequently the "tipping point" of innovation, where broader dissemination occurs. Prospective multicenter observational studies should report key patient characteristics (case mix), technical variables (including operator learning curve), and predefined standardized outcomes (48). The goal is to prepare for definitive evaluation in the assessment stage (3), ideally through randomized controlled trials (RCTs) or comprehensive protocol-driven observational studies when an RCT is not feasible (e.g., quasiexperimental study designs such as interrupted time series or propensity score–matching analyses). The long-term stage (4) focuses on surveillance of rare adverse events and effectiveness. Disease-based registries may be best suited to provide an understanding of variations in patient selection for various procedures (49). The IDEAL framework thus provides important practical guidance for evaluation of innovation throughout the total product life cycle (50).

Statement 10. Methodologically robust comparative studies, preferentially RCTs, are needed to adequately assess and validate minimally invasive diagnostic technologies targeting the lung periphery.

There is broad recognition within the interventional pulmonology community of the need for higher-quality studies, particularly comparative effectiveness studies of different techniques. Robust studies in the assessment phase of evaluation (IDEAL stage 3) are lacking from the bronchoscopy evidence landscape. The preferred study design for definitive evaluation is an RCT, although the window of opportunity for conducting one may sometimes seem narrow. The smaller the incremental impact of novel technology, the greater the need for a randomized trial (49, 51). The major benefit of an RCT is minimization of bias, because random allocation ensures balance in both known and unknown confounders affecting the outcome(s) of interest. Comparison needs not be limited to alternative advanced diagnostic bronchoscopy techniques but should consider the broader standard of care for investigation of peripheral pulmonary lesions. An example would be comparison of novel (robotic) bronchoscopy platforms with CT-guided TTNBs, which are associated with higher diagnostic yields for peripheral pulmonary lesions (without the need for general anesthesia) but higher complication rates (52).

If an RCT is not feasible, alternative quasiexperimental study designs may be considered. Quasiexperimental study designs attempt to replicate RCTs by rendering treatment groups as similar as possible on all known confounders (53). Examples of quasiexperimental designs include interrupted time series with temporal controls, propensity score matching, regression discontinuity, and instrumental variables (48). Such prospective observational studies require careful design and planning, with robust measures to address potential biases, including selection bias. A nonrandomized controlled trial would compare a cohort of patients evaluated with a novel advanced bronchoscopy platform with a concurrent group of patients undergoing a standard diagnostic intervention. Standardized data collection with control for all known confounders is required to address selection bias. Propensity score matching, which balances study groups on the likelihood of

Table 3. Bronchoscopy-adapted Study Designs According to IDEAL (45–49) Stage of Innovation

	1 Idea*	2a Development	2b Exploration	3 Assessment	4 Long-Term Study
Question	Can the procedure or device achieve a specific physical or diagnostic goal?	What is the optimal technique, and for which patient does it work best?	What are the diagnostic outcomes of more widespread use?	How well does the procedure work compared with current standards of care?	What are the long-term outcomes of the procedure?
Aim	Proof of concept	Safety, efficacy	Efficacy	Comparative effectiveness	Surveillance
Patients	Single to few (<10)	10 s	100 s	100 s+	100 s+
Optimal study design(s)	First-in-human study; structured case reports	Prospective development studies	Prospective collaborative observational studies, or feasibility randomized controlled trial	Randomized controlled trial Nonrandomized controlled trials or interrupted time series are potential alternatives to randomized controlled trials	Comprehensive disease-based registry or database
Key aspects	Confidential reporting of all first-in-human procedures	Protocol Sequential reporting of cases and modifications of technique	Protocol Standardized outcome definitions Learning curve evaluation	Quantitative comparison with standards of care	Monitoring for long-term safety (rare adverse events) and effectiveness
Examples at this stage†	First-in-Human Use of a Hybrid Real-Time Ultrasound-guided Fine-Needle Acquisition System for Peripheral Pulmonary Lesions: A Multicenter Pilot Study (62)	Robotic Bronchoscopy for Peripheral Pulmonary Lesions: A Multicenter Pilot and Feasibility Study (BENEFIT) (63)	A Multicenter Prospective Trial Assessing the Diagnostic Yield of Electromagnetic Bronchoscopic and Transthoracic Navigation for Peripheral Pulmonary Nodules (22)	Standard Bronchoscopy with Fluoroscopy vs Thin Bronchoscopy and Radial Endobronchial Ultrasound for Biopsy of Pulmonary Lesions: A Multicenter, Prospective, Randomized Trial (12)	Diagnostic Yield and Complications of Bronchoscopy for Peripheral Lung Lesions. Results of the AQuIRE Registry (14)

*Addition of a preclinical phase 0 for device development (47).
†Examples drawn from the advanced diagnostic bronchoscopy literature.

receiving an intervention to reduce confounding by indication, may be considered (54, 55). Interrupted time series is an alternative quasiexperimental study design (56) whereby key outcomes are measured before and after the introduction of a novel intervention (i.e., the interruption). Interrupted time series with a control group for time are preferred to avoid biases due to temporal trends. Here again, analysis requires careful adjustment for key patient and/or peripheral pulmonary lesion characteristics.

Statement 11. Aside from clear diagnostic performance and safety outcomes as primary endpoints, additional methodological factors should be clearly established in the study protocols. Examples include setting of recruitment, patient selection, target characteristics, and other relevant factors related to the index test. These factors should be recorded and reported according to STARD guidelines.

Beyond clearly defined and transparently reported diagnostic outcome measures, relevant patient, operator, and procedural factors should be delineated at the study outset and reported. The STARD guidelines provide guidance for reporting of these important factors, with a complete checklist available at equator-network.org/reporting-guidelines/stard/. These include patient characteristics, with reporting of both individual pretest probability of malignancy and prevalence of malignancy, underlying lung disease, and lung nodule characteristics known to be associated with diagnostic yield (size, anatomic location, presence of a bronchus sign, etc.). Operator factors include learning curve and diagnostic yield before formal evaluation of the new technology, as well as procedural volume of operators/centers reporting their experience. Relevant procedural factors include the use of moderate sedation versus general anesthesia, availability of rapid on-site evaluation by a cytopathologist, and use of imaging adjuncts (e.g., radial endobronchial ultrasound, cone-beam CT). Procedure duration and personnel and equipment requirements (including disposables) are additional relevant factors that relate to feasibility and cost of the intervention and warrant reporting.

Statement 12. Patient-centered and patient-reported outcomes should be studied and incorporated into the evaluation of novel bronchoscopy techniques, including the following:

a. Better understanding of downstream patient events (e.g., need for salvage procedures, complications and/or hospitalizations, delays in diagnosis)
b. Consideration of patient preferences for differing diagnostic pathways

The adoption of technology into clinical practice is multidimensional and not guided

by diagnostic outcomes alone. Patient-centered outcomes (e.g., impact of bronchoscopic procedure on patient trajectory, including the need for additional invasive procedures) and patient-reported outcomes (e.g., patient's willingness to have a repeat procedure) should be incorporated into studies of advanced diagnostic bronchoscopy. The type and severity of complications also guide decision making. Complication definitions should be outlined *a priori*. When available, standardized definitions of complications facilitate uniform reporting and meaningful pooling of data (57).

There is limited data on patient-reported outcomes and/or patient preferences in advanced diagnostic bronchoscopy and investigation of pulmonary nodules. Older data relate to patient satisfaction with bronchoscopy and willingness to repeat the procedure (58). Very little data is available on how patients weigh current diagnostic pathway alternatives. Considerations may include the need for general anesthesia versus moderate sedation, the potential risks of complications (e.g., risk of pneumothorax with transthoracic needle aspiration vs. bronchoscopy), the need for additional prebronchoscopic imaging, the uncertainty associated with a nondiagnostic procedure, or the need for repeat invasive procedures (including time off work, discomfort, or complications).

Statement 13. Inclusion of academic and community centers serving diverse patient populations, including underrepresented and underserved patients, is strongly encouraged.

Statement 14. Informative and inclusive study designs can support subsequent health economic analyses, can guide implementation decisions in various healthcare settings, and may help address disparities such as access to technological innovations.

Evidence-based integration of novel bronchoscopy technologies into clinical practice should be grounded in outcome data representative of the diverse patient population referred for investigation of peripheral pulmonary lesions. Deliberate, collaborative efforts are needed to ensure prospective multicenter studies are conducted in a broad range of clinical settings.

Cost-effectiveness analyses guide the rational introduction of novel technology into a range of clinical settings. Health economic analyses are beyond the scope of this research statement. However, robust comparative effectiveness studies with transparent reporting of all personnel and equipment requirements can support accurate cost evaluations.

Statement 15. There is a need for funding mechanisms for high-quality studies of novel technology for sampling of the lung periphery. These include the following:

a. Pathways for improved partnership with industry
b. Wider commitment for industry validation and postmarket reporting
c. Societal partnerships

There is an urgent need for private and public investments in robust evaluative research at various stages of introduction of innovative bronchoscopy technology. Regulators should demand rigorous clinical evidence, which typically includes "pivotal" clinical trials, before approval of medical devices. The new European Medical Device Regulation modified the level of evidence required for CE (European conformity) certification (50). The evidence level varies according to the device's stage of development, from approval to surveillance, reflecting the "total product life cycle."

Prospective multicenter registries are central to postmarket surveillance of innovative technology, including monitoring of real-life outcomes and rarer complications, but are costly to maintain. Partnerships are needed for long-term study.

Conclusions

The introduction of innovations into clinical practice is often based on the promise and excitement of early adoption rather than solid proof. There is a trade-off between the benefit of early access to promising devices and the risk of harm from incompletely validated devices. Most medical devices currently enter clinical practice through the U.S. Food and Drug Administration's 510(k) pathway, a less stringent approval process with inherent limitations. Lack of high-quality research with firm outcome definitions constitutes a barrier to introduction of novel technology in both pluralistic and single-payer systems. This American Thoracic Society/American College of Chest Physicians statement aims to provide a research framework that allows greater standardization of device validation efforts through clearly defined diagnostic outcomes and robust study designs. High-quality studies, both industry and publicly funded, can support subsequent health economic analyses and guide implementation decisions in various healthcare settings. ■

This research statement was prepared by an *ad hoc* subcommittee of the ATS Assembly on Thoracic Oncology and the American College of Chest Physicians.

Members of the subcommittee are as follows:
Anne V. Gonzalez, M.D., M.Sc. (*Co-Chair*)[1]
Gerard A. Silvestri, M.D., M.S. (*Co-Chair*)[2]
Lonny B. Yarmus, D.O., M.B.A. (*Co-Chair*)[3]
Nisha D. Almeida, Ph.D.[4]
Alex Chen, M.D.[5]
Yaron B. Gesthalter, M.D.[6]
Chris R. Gilbert, D.O., M.S.[2]
Peter B. Illei, M.D.[3]
Daniel A. Korevaar, M.D., Ph.D.[7]
Fabien Maldonado, M.D., M.Sc.[8]
Neal Navani, M.A, M.Sc., Ph.D.[9]
Mary M. Pasquinelli, D.N.P.[10]
Nicholas J. Pastis, M.D.[11]
M. Patricia Rivera, M.D.[12]
Catherine R. Sears, M.D.[13]
Samira Shojaee, M.D., M.P.H.[8]
Stephen B. Solomon, M.D.[14]
Daniel P. Steinfort, M.B.B.S., Ph.D.[15]

[1]Division of Respiratory Medicine, McGill University Health Centre, Montreal, Quebec, Canada; [2]Medical University of South Carolina, Charleston, South Carolina; [3]Johns Hopkins University School of Medicine, Baltimore, Maryland; [4]Division of Clinical Epidemiology, McGill University, and Health Technology Assessment Unit, McGill University Health Centre, Montreal, Quebec, Canada; [5]Washington University School of Medicine, St. Louis, Missouri; [6]University of California, San Francisco, San Francisco, California; [7]Department of Respiratory Medicine, Amsterdam University Medical Centers, Amsterdam, the Netherlands; [8]Vanderbilt University Medical Center, Nashville, Tennessee; [9]UCL Respiratory, University College London, London, United Kingdom; [10]University of Illinois Chicago, Chicago, Illinois; [11]The Ohio State University, Colombus, Ohio; [12]University of Rochester Medical Center, Rochester, New York; [13]Indiana University School of Medicine and Richard L. Roudebush Veterans Affairs Medical Center, Indianapolis, Indiana; [14]Interventional Radiology Service, Memorial Sloan Kettering Cancer Center, New York, New York; and [15]Melbourne Medical School, University of Melbourne, Parkville, Australia

Subcommittee disclosures: A.V.G. served on data safety and monitoring board for Laurent Pharmaceuticals; received research support from Biodesix, Kinova, Lung Ambition, Lung Cancer Canada, Nucleix. G.A.S. served as consultant for Biodesix, Johnson & Johnson, Olympus; received research support from Biodesix, Delfi, Exact Science, Johnson & Johnson – Auris, Nucleix, Olympus, Prognomiq. Y.G. served on advisory board for Patient Power; served as consultant for Guidepoint, Intuitive Surgical, Lead Optik, Noah Medical, Remedy; provided expert testimony for Guidepoint; received honoraria from Spiration; holds patent SF2023-062. A.C. served as consultant for Auris, Biodesix, Boston Scientific, Intuitive, Olympus; received research support from Auris, Boston Scientific, Johnson & Johnson, Olympus. C.R.G. served as consultant for AstraZeneca and Verathon. P.B.I. served on advisory board for AstraZeneca, Bristol Myers Squibb, Janssen, Roche; served as consultant for Veran Medical; received research support from AbbVie, AstraZeneca, Bristol Myers Squibb, Eli Lilly, Erbe GmBH, Genentech, Janssen, Merus, Roche AG, Sanofi; served as speaker for Eli Lilly and Genentech. N.N. served on advisory board for Bristol Myers Squibb, Janssen, Lilly, MSD, Olympus; received research support from Amgen, AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, EQRx, FujiFilm, Guardant Health, Intuitive, Janssen, Lilly, Merck Sharp & Dohme, Olympus, OncLive, PeerVoice; served as speaker for Amgen and AstraZeneca. N.J.P. served as consultant for Olympus America. C.R.S. employee of U.S. Department of Veterans Affairs; served on advisory board for bioAffinity Technologies; served as consultant for Bristol Myers Squibb; received research support from American Cancer Society, Biodesix, NIH, U.S. Department of Veterans Affairs; served as speaker for American College of Chest Physicians, CHEST, National Comprehensive Cancer Network; received travel support from American Cancer Society, Florida Society of Clinical Oncology, Go2 Foundation. S.S. received research support from Biodesix. S.B.S. served as consultant for XACT Robotics; served on data safety and monitoring board for Candel Therapeutics; received research support from Candel Therapeutics, GE Healthcare, Impact Biotech, Johnson & Johnson, MedX, Merck, Varian. D.P.S. received research support from Gala Therapeutics, Uptake Medical, Zidan Medical. F.M. served as consultant for Intuitive, Johnson & Johnson, Medtronic; received research support from Medtronic. M.P.R. employee of University of Rochester Medical Center; served as consultant for American Board of Internal Medicine and American Cancer Society. L.B.Y. served as consultant for Elucent Medical, Erbe, FujiFilm, Olympus; has financial stake in Pleural Dynamics; received research support from AstraZeneca, Erbe, Intuitive, Olympus; served as speaker for Olympus. D.A.K., N.D.A., M.M.P. reported no commercial or relevant non-commercial interests from ineligible companies.

References

1. Gould MK, Tang T, Liu IL, Lee J, Zheng C, Danforth KN, *et al.* Recent trends in the identification of incidental pulmonary nodules. *Am J Respir Crit Care Med* 2015;192:1208–1214.
2. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, *et al.*; National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
3. Kim H, Lee KH, Han K, Lee JW, Kim JY, Im DJ, *et al.* Development and validation of a deep learning-based synthetic bone-suppressed model for pulmonary nodule detection in chest radiographs. *JAMA Netw Open* 2023;6:e2253820.
4. Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143(5 Suppl):e142S–e165S.
5. Chen AC, Pastis NJ, Machuzak MS, Gildea TR, Simoff MJ, Gillespie CT, *et al.* Accuracy of a robotic endoscopic system in cadaver models with simulated tumor targets: ACCESS study. *Respiration* 2020;99:56–61.
6. Yarmus L, Akulian J, Wahidi M, Chen A, Steltz JP, Solomon SL, *et al.*; Interventional Pulmonary Outcomes Group (IPOG). A prospective randomized comparative study of three guided bronchoscopic approaches for investigating pulmonary nodules: the PRECISION-1 study. *Chest* 2020;157:694–701.
7. Silvestri GA, Gonzalez AV, Jantz MA, Margolis ML, Gould MK, Tanoue LT, *et al.* Methods for staging non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(5 Suppl):e211S–e250S.
8. Eberhardt R, Anantham D, Ernst A, Feller-Kopman D, Herth F. Multimodality bronchoscopic diagnosis of peripheral lung lesions: a randomized controlled trial. *Am J Respir Crit Care Med* 2007; 176:36–41.
9. Chen CH, Cheng WC, Wu BR, Chen CY, Chen WC, Hsia TC, *et al.* Improved diagnostic yield of bronchoscopy in peripheral pulmonary lesions: combination of radial probe endobronchial ultrasound and rapid on-site evaluation. *J Thorac Dis* 2015;7:S418–S425.
10. Folch EE, Pritchett MA, Nead MA, Bowling MR, Murgu SD, Krimsky WS, *et al.*; NAVIGATE Study Investigators. Electromagnetic navigation bronchoscopy for peripheral pulmonary lesions: one-year results of the prospective, multicenter NAVIGATE study. *J Thorac Oncol* 2019;14: 445–458.
11. Chen A, Chenna P, Loiselle A, Massoni J, Mayse M, Misselhorn D. Radial probe endobronchial ultrasound for peripheral pulmonary lesions. a 5-year institutional experience. *Ann Am Thorac Soc* 2014;11: 578–582.
12. Tanner NT, Yarmus L, Chen A, Wang Memoli J, Mehta HJ, Pastis NJ, *et al.* Standard bronchoscopy with fluoroscopy vs thin bronchoscopy and radial endobronchial ultrasound for biopsy of pulmonary lesions: a multicenter, prospective, randomized trial. *Chest* 2018;154:1035–1043.
13. Vakil E, Tremblay A. Flying blind despite all our instruments. *Chest* 2020; 158:1312–1313.
14. Ost DE, Ernst A, Lei X, Kovitz KL, Benzaquen S, Diaz-Mendoza J, *et al.*; AQuIRE Bronchoscopy Registry. Diagnostic yield and complications of bronchoscopy for peripheral lung lesions. results of the AQuIRE Registry. *Am J Respir Crit Care Med* 2016;193:68–77.
15. Durakovic A, Andersen H, Christiansen A, Hammen I. Retrospective analysis of radial EBUS outcome for the diagnosis of peripheral pulmonary lesion: sensitivity and complications. *Eur Clin Respir J* 2015;2:28947.
16. Eberhardt R, Anantham D, Herth F, Feller-Kopman D, Ernst A. Electromagnetic navigation diagnostic bronchoscopy in peripheral lung lesions. *Chest* 2007;131:1800–1805.
17. Kuo CH, Lin SM, Lee KY, Chung FT, Lo YL, Hsiung TC, *et al.* Endobronchial ultrasound-guided transbronchial biopsy and brushing: a comparative evaluation for the diagnosis of peripheral pulmonary lesions. *Eur J Cardiothorac Surg* 2014;45:894–898.
18. Herth FJ, Mayer M, Thiboutot J, Kapp CM, Sun J, Zhang X, *et al.* Safety and performance of transbronchial cryobiopsy for parenchymal lung lesions. *Chest* 2021;160:1512–1519.
19. Wang Memoli JS, Nietert PJ, Silvestri GA. Meta-analysis of guided bronchoscopy for the evaluation of the pulmonary nodule. *Chest* 2012; 142:385–393.
20. Vachani A, Maldonado F, Laxmanan B, Kalsekar I, Murgu S. The impact of alternative approaches to diagnostic yield calculation in studies of bronchoscopy. *Chest* 2022;161:1426–1428.
21. Nadig TR, Thomas N, Nietert PJ, Lozier J, Tanner NT, Wang Memoli JS, *et al.* Guided bronchoscopy for the evaluation of pulmonary lesions: an updated meta-analysis. *Chest* 2023;163:1589–1598.
22. Thiboutot J, Pastis NJ, Akulian J, Silvestri GA, Chen A, Wahidi MM, *et al.* A multicenter, single-arm, prospective trial assessing the diagnostic yield of electromagnetic bronchoscopic and transthoracic navigation for peripheral pulmonary nodules. *Am J Respir Crit Care Med* 2023;208: 837–845.
23. Thiboutot J, Yarmus LB, Lee HJ, Rivera MP, Ost DE, Feller-Kopman D. Real-world application of the NAVIGATE trial. *J Thorac Oncol* 2019;14: e146–e147.
24. Maldonado F, Eberl J. Bearing the burden of "innovation": the ontological implications of substantial equivalence and the FDA 510(k) pathway. *Chest* 2023;163:1225–1227.

25. Rathi VK, Ross JS. Modernizing the FDA's 510(k) pathway. *N Engl J Med* 2019;381:1891–1893.
26. Salazar JW, Redberg RF. Leading the call for reform of medical device safety surveillance. *JAMA Intern Med* 2020;180:179–180.
27. Landeen KC, Langerman A, Maldonado F. An oversight in oversight: reconciling the medical device industry, clinicians, and regulation. *Chest* 2022;161:300–302.
28. Drummond MF, Augustovski F, Bhattacharyya D, Campbell J, Chaiyakanapruk N, Chen Y, *et al.*; ISPOR HTA Council Working Group on HTA in Pluralistic Healthcare Systems collaborators. Challenges of health technology assessment in pluralistic healthcare systems: an ISPOR Council report. *Value Health* 2022;25:1257–1267.
29. Kops SE, Heus P, Korevaar DA, Damen JAA, Idema DL, Verhoeven RLJ, *et al.* Diagnostic yield and safety of navigation bronchoscopy: a systematic review and meta-analysis. *Lung Cancer* 2023;180:107196.
30. Gonzalez AV. Stating the obvious … or leading the way toward more judicious use of diagnostic bronchoscopy. *Chest* 2020;157: 1409–1410.
31. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, *et al.* Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267–276.
32. Reitsma JB, Rutjes A, Whiting P, Yang B, Leeflang MM, Bossuyt PM, Deeks JJ. Chapter 8: Assessing risk of bias and applicability. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editors. Cochrane handbook for systematic reviews of diagnostic test accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from https://training.cochrane.org/handbook-diagnostic-test-accuracy/current
33. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340:c869.
34. Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med* 2014;19:47–54.
35. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N, *et al.* Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology* 2015;274:781–789.
36. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, *et al.*; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; 351:h5527.
37. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
38. Korevaar DA, Cohen JF, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, *et al.* Updating standards for reporting diagnostic accuracy: the development of STARD 2015. *Res Integr Peer Rev* 2016;1:7.
39. Chae KJ, Hong H, Yoon SH, Hahn S, Jin GY, Park CM, *et al.* Non-diagnostic results of percutaneous transthoracic needle biopsy: a meta-analysis. *Sci Rep* 2019;9:12428.
40. Kim JI, Park CM, Kim H, Lee JH, Goo JM. Non-specific benign pathological results on transthoracic core-needle biopsy: how to differentiate false-negatives? *Eur Radiol* 2017;27:3888–3895.
41. Murthi M, Yoshioka K, Cho JH, Arias S, Danna E, Zaw M, *et al.* Presence of concurrent sarcoid-like granulomas indicates better survival in cancer patients: a retrospective cohort study. *ERJ Open Res* 2020;6: 00061-2020.
42. Stefanidis D, Fanelli RD, Price R, Richardson W. Guidelines for the introduction of new technology and techniques. 2014 [accessed 2023 Jul 13]. Available from: https://www.sages.org/publications/guidelines/guidelines-introduction-new-technology-techniques/.
43. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Altman DG, *et al.*; Balliol Collaboration. Evaluation and stages of surgical innovations. *Lancet* 2009;374:1089–1096.
44. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, *et al.*; Balliol Collaboration. Challenges in evaluating surgical innovation. *Lancet* 2009;374:1097–1104.
45. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, *et al.*; Balliol Collaboration. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374:1105–1112.
46. McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK; IDEAL Group. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013;346:f3012.
47. Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ* 2016;353:i2372.
48. Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG; IDEAL Group. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ* 2013;346:f3011.
49. Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A; IDEAL Group. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ* 2013;346:f2820.
50. Fleetcroft C, McCulloch P, Campbell B. IDEAL as a guide to designing clinical device studies consistent with the new European Medical Device Regulation. *BMJ Surg Interv Health Technol* 2021;3:e000066.
51. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;334:349–351.
52. Wiener RS, Schwartz LM, Woloshin S, Welch HG. Population-based risk for complications after transthoracic needle lung biopsy of a pulmonary nodule: an analysis of discharge records. *Ann Intern Med* 2011;155: 137–144.
53. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin; 2002.
54. Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J* 2011;32:1704–1708.
55. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–2281.
56. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002;27:299–309.
57. U.S. Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. 2017 [accessed 2023 Jul 13]. Available from: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf.
58. Lechtzin N, Rubin HR, White P Jr, Jenckes M, Diette GB. Patient satisfaction with bronchoscopy. *Am J Respir Crit Care Med* 2002;166: 1326–1331.
59. Fletcher RH, Fletcher SW, Fletcher GS. Diagnosis. In: Clinical epidemiology: the essentials. 5th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2014. pp. 108–131.
60. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, *et al.* Interpreting diagnostic information. In: Decision making in health and medicine: integrating evidence and values. 2nd ed. Cambridge, UK: Cambridge University Press; 2014. pp. 118–144.
61. Gonzalez AV, Ost DE, Shojaee S. Diagnostic accuracy of bronchoscopy procedures: definitions, pearls, and pitfalls. *J Bronchology Interv Pulmonol* 2022;29:290–299.
62. Yarmus LB, Mallow C, Pastis N, Thiboutot J, Lee H, Feller-Kopman D, *et al.*; Interventional Pulmonary Outcomes Group (IPOG). First-in-human use of a hybrid real-time ultrasound-guided fine-needle acquisition system for peripheral pulmonary lesions: a multicenter pilot study. *Respiration* 2019;98:527–533.
63. Chen AC, Pastis NJ Jr, Mahajan AK, Khandhar SJ, Simoff MJ, Machuzak MS, *et al.* Robotic bronchoscopy for peripheral pulmonary lesions: a multicenter pilot and feasibility study (BENEFIT). *Chest* 2021;159: 845–852.