

Evaluating Molecular Biomarkers for the Early Detection of Lung Cancer: When Is a Biomarker Ready for Clinical Use? An Official American Thoracic Society Policy Statement

Peter J. Mazzone, Catherine Rufatto Sears, Doug A. Arenberg, Mina Gaga, Michael K. Gould, Pierre P. Massion, Vish S. Nair, Charles A. Powell, Gerard A. Silvestri, Anil Vachani, and Renda Soylemez Wiener; on behalf of the ATS Assembly on Thoracic Oncology

THIS OFFICIAL POLICY STATEMENT OF THE AMERICAN THORACIC SOCIETY WAS APPROVED JULY 2017

Background: Molecular biomarkers have the potential to improve the current state of early lung cancer detection. The goal of this project was to develop a policy statement that provides guidance about the level of evidence required to determine that a molecular biomarker, used to support early lung cancer detection, is appropriate for clinical use.

Methods: An *ad hoc* project steering committee was formed, to include individuals with expertise in the early detection of lung cancer and molecular biomarker development, from inside and outside of the Assembly on Thoracic Oncology. Key questions, generated from the results of a survey of the project steering committee, were discussed at an in-person meeting. Results of the discussion were summarized in a policy statement that was circulated to the steering committee and revised multiple times to achieve consensus.

Results: With a focus on the clinical applications of lung cancer screening and lung nodule evaluation, the policy statement outlines categories of results that should be reported in the early phases of

molecular biomarker development, discusses the level of evidence that would support study of the clinical utility, describes the outcomes that should be proven to consider a molecular biomarker clinically useful, and suggests study designs capable of assessing these outcomes.

Conclusions: The application of molecular biomarkers to assist with the early detection of lung cancer has the potential to substantially improve our ability to select patients for lung cancer screening, and to assist with the characterization of indeterminate lung nodules. We have described relevant considerations and have suggested standards to apply when determining whether a molecular biomarker for the early detection of lung cancer is ready for clinical use.

Keywords: clinical utility; outcomes; study design; lung cancer screening; lung nodules

Contents

Overview

Introduction

Methods

Background

Clinical Utility Phase of
Biomarker Development
Cost-Effectiveness

Results

Discussion: Clinical Applications of
Molecular Biomarkers for the Early
Detection of Lung Cancer
Lung Cancer Screening
Lung Nodule Evaluation

Conclusions

Appendix 1: Definitions
Appendix 2: Additional References
Used during the Project

Overview

Molecular biomarkers, developed to assist with the early detection of lung cancer, may be applied in the settings of lung cancer screening and lung nodule evaluation. True-positive and -negative results may benefit patients, whereas false-positive and -negative results may lead to harm. To have clinical utility, the molecular biomarker must affect clinical management decisions

An Executive Summary of this document is available at <http://www.atsjournals.org/doi/suppl/10.1164/rccm.201708-1678ST>

Correspondence and requests for reprints should be addressed to Peter J. Mazzone, M.D., M.P.H., Respiratory Institute, Cleveland Clinic, 9500 Euclid Avenue, A90, Cleveland, OH 44195. E-mail: mazonp@ccf.org

Am J Respir Crit Care Med Vol 196, Iss 7, pp e15–e29, Oct 1, 2017

Copyright © 2017 by the American Thoracic Society

DOI: 10.1164/rccm.201708-1678ST

Internet address: www.atsjournals.org

in a manner that improves clinical outcomes. This policy statement describes items to consider when determining whether the evidence supports clinical use of an early lung cancer detection molecular biomarker. Key points made in the statement include the following:

- Key results from studies in all phases of biomarker development may influence the interpretation of clinical utility. A list of results that should be reported is provided.
- Calculations are available to help determine the minimal accuracy of a molecular biomarker that could lead to a favorable clinical impact, and thus justify investment in a clinical utility study. An example is provided.
- A clinically useful molecular biomarker applied as the initial test for lung cancer screening may improve the balance of benefit to harm by identifying those most likely to benefit from screening while minimizing exposure to the harms among those least likely to benefit.
- To be considered clinically useful, a molecular biomarker used to identify patients eligible for lung cancer screening must lead to:
 - Fewer lung cancer deaths in the population tested compared with the current standard of care for that population, without substantially increasing harms and expense, or
 - A similar number of lung cancer deaths in the population tested compared with the current standard of care for that population, with fewer harms or less expense.
- A clinically useful molecular biomarker applied to the evaluation of lung nodules may lead to expedited therapy for early lung cancer and/or fewer aggressive interventions in patients with benign lung nodules.
- To be considered clinically useful, a molecular biomarker used to assist with lung nodule management must lead to:
 - Earlier diagnosis of malignant nodules without substantially increasing the number of procedures performed on patients with benign nodules, or
 - Fewer procedures for patients with benign nodules without substantially delaying the diagnosis of cancer in patients with malignant nodules.
- Biomarker-stratified, enrichment, and biomarker strategy study designs may provide the evidence required to assess

the outcomes of interest when determining whether a lung cancer screening or lung nodule management molecular biomarker is clinically useful.

Introduction

Molecular biomarkers have the potential to improve the current state of early lung cancer detection. Biomarkers capable of identifying the presence of presymptomatic lung cancer may help to optimize patient selection for lung cancer screening. Biomarkers capable of characterizing pulmonary nodules may help to expedite therapy of early-stage lung cancers while minimizing the harms of evaluating patients with benign disease.

A new lung cancer biomarker will be clinically useful if it fulfills an unmet clinical need or provides an advantage over standard practice (e.g., more accurate, simpler to use, provides results more rapidly, lowers costs). The ultimate measure of lung cancer biomarker performance is whether and how the result affects clinical management decisions and clinical outcomes (1). Even biomarkers that are sensitive and specific enough to be considered accurate by most clinicians may not impact clinical care, or may adversely impact clinical care (e.g., Table 1).

The goal of this project was to develop a policy statement that provides guidance about the evidence required to determine whether a molecular biomarker for the early detection of lung cancer is appropriate for clinical use.

Methods

During the first stage of the project, the project co-chairs (P.J.M. and C.R.S.) developed an overview of biomarker

development principles and the current state of lung cancer risk prediction, early detection, and diagnosis, as well as a survey with questions related to each of the phases of biomarker development. A conference call was held to introduce the project to the steering committee. The documents were then circulated to the steering committee for review. The survey was completed and returned by 11 of the 12 steering committee members, and the results were collated. The survey responses were used to generate discussion questions.

The project *ad hoc* steering committee was selected to include individuals with an interest in lung cancer biomarker development and expertise in the various phases of biomarker development. Representatives of American Thoracic Society international partner societies (Chinese Thoracic Society, European Respiratory Society, and Japanese Respiratory Society) and the U.S. Food and Drug Administration were invited to participate on the steering committee. Potential conflicts of interest were disclosed and managed in accordance with the policies and procedures of the American Thoracic Society. The committee had a face-to-face meeting on May 13, 2016 at the annual international conference of the American Thoracic Society. The meeting included a presentation from a patient advocate, and presentations related to the phases of biomarker development. The formal presentations were followed by discussions guided by questions generated from the survey responses (*see below*).

Separately, the steering committee reviewed definitions, reporting considerations, target conditions, target populations, reference standards, and the potential impact of true-positive/-negative and false-positive/-negative results within each potential clinical application. A draft of the current document was developed by

Table 1. Potential for Harm from an Accurate Biomarker

	Nodule Malignant	Nodule Benign	Total
Test result positive	9	99	108
Test result negative	1	891	892
Total	10	990	1,000

If a molecular biomarker is 90% sensitive and 90% specific for the detection of a malignant lung nodule, it would generally be considered an accurate test. If applied to a population of patients with lung nodules with a 1% probability of malignancy (e.g., solid nodules 4–8 mm in diameter), 92% of all positive test results would be false positive, potentially leading to more aggressive evaluation of many patients without lung cancer, with physical, social, and behavioral consequences.

the project co-chairs and then circulated for review. The draft was modified several times, based on written feedback from the project steering committee and on feedback received during a phone conference. Relevant definitions and additional references are presented in Appendices 1 and 2.

Background

Clinical Utility Phase of Biomarker Development

The clinical utility phase of biomarker development follows successful completion of biomarker discovery, analytical validation of the biomarker assay, and clinical validation of the accuracy of the biomarker (Figure 1). A biomarker should be used in clinical practice only if it reliably adds to a clinician’s judgment, resulting in a more favorable clinical outcome for the target population. Biomarker accuracy is not enough to imply clinical utility. Clinical utility is dependent on how the test result affects subsequent clinical decisions and outcomes. Both the benefit of clinical decisions influenced by true-positive and true-negative results, as well as the harms of clinical decisions influenced by false-positive and false-negative results, must be considered. Clinical decisions based on misleading biomarker results may expose the patient to adverse consequences and increase the cost of care.

Outcomes of relevance in clinical utility testing include the frequency and manner in which the biomarker impacts clinical decisions (use of other testing, treatment decisions) and the consequences of those decisions (complications from testing and treatment, patient quality of life, timeliness of accurate diagnosis, survival). There are many challenges to evaluating the clinical utility of a biomarker in randomized controlled trials, as is typically required for determination of the therapeutic efficacy of a drug or intervention.

Research designs capable of determining the clinical utility of a biomarker have been described (2–5) (Figures 2 and 4):

1. *Biomarker stratified*: All patients are randomly assigned to one of two or more management options regardless of biomarker status. This design is the most efficient way to determine the best management option for each biomarker subgroup when there is no evidence that one option is preferred. Biomarker-stratified designs include the following:
 - a. Prospective controlled clinical trials where the test result is used to stratify patients who are then randomized to the management options being studied. The biomarker is not used for patient management, but the primary objective of the clinical trial is to assess the clinical utility of the test for its intended use. Limitations of this pathway include the potential for bias from randomly missing test results (can be minimized in the statistical analysis), limited statistical power to detect small effects, and an inability to test strategies with a large number of management options.
 - b. Prospective-retrospective studies using archived specimens from previously conducted controlled clinical trials that address the intended use of the test. This pathway is the least time and resource intensive. It is appropriate for biomarkers that can be evaluated retrospectively in a reliable way. Limitations of this pathway include the difficulty in finding adequate archived specimens (quality and number) and the potential for bias from randomly missing test results. Because missing specimens may not be equally distributed between the randomized management arms, the

- study design must ensure that an adequate number of patient samples is available for each biomarker subgroup.
2. *Enrichment*: Patients in one biomarker-defined subgroup (e.g., biomarker positive) are randomly assigned to one of two or more management options, and the other biomarker-defined subgroup (e.g., biomarker negative) is managed on the basis of standard of care for the population. This design is applicable in settings where there is evidence to suggest that the benefit of a management option would be limited to one biomarker-defined subgroup. The enrichment study design can only address questions about the best management strategy within the biomarker-defined subgroup that is randomized to the studied management options. It is an efficient study design when the assumption that benefit will only be seen in one biomarker-defined subgroup is correct, and when the incidence of the condition in question is low. This design cannot completely address clinical utility as it does not assess the potential utility of the management strategies in the other biomarker-defined subgroup.
3. *Biomarker strategy*: Patients are randomly assigned to a biomarker-directed arm where the management strategy is based on the biomarker result, or to a control arm. Patients in the control arm may be assigned to standard of care management, or be randomly assigned to one of two or more management strategies, independent of the biomarker result. This design determines whether biomarker-directed management is better than standard of care management. This design most directly assesses the clinical utility of a biomarker but is the most resource and time intensive and carries the highest risk to the study participants. This design may not completely assess clinical utility as it does not determine whether the management strategies assessed in the biomarker-directed arm are better than standard of care regardless of the biomarker result.

Evidence of clinical utility may not result in widespread adoption of a biomarker, particularly if measurement of the biomarker is costly, requires nonroutine

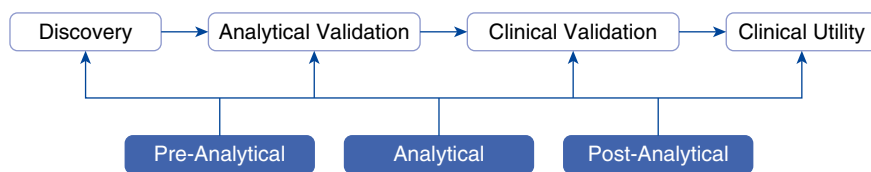


Figure 1. Phases of biomarker development. (Top row) Phases of biomarker development. (Bottom row) Process considerations. This applies to the development of all categories of biomarkers: risk prediction, cancer detection, and diagnosis.

sample collection, is technically difficult, or requires a change in the culture of clinical practice.

Cost-Effectiveness

Cost-effectiveness analysis (CEA) is an important tool in determining the impact of the use of a biomarker. CEA measures may be used to guide the development of recommendations for clinical use of a biomarker from professional societies, government, and industry payer sources. Some regulatory authorities (e.g., the U.S. Food and Drug Administration) do not ordinarily consider monetary costs (e.g., medical bills, societal or insurer costs), but have considered health costs (e.g., mortality or morbidity) (6). This practice is likely to evolve, with monetary costs more routinely considered by payers.

The measurement and interpretation of cost-effectiveness is complex. It requires accurate estimates of the net cost of implementing a biomarker for a given outcome. CEA is only relevant if clinical utility has been proven (e.g., reduced mortality in a cancer screening trial). CEA uses data from published studies to determine measures such as sensitivity, specificity, and disease prevalence. It is imperative that CEA be measured in the correct clinical context across diverse ethnic and social groups, and be compared with accepted clinical practices (incremental cost-effectiveness ratio). It is difficult to compare the cost-effectiveness of one biomarker or intervention with another if different measures are used. For that reason, uniform measurement and reporting of cost-effectiveness using quality-adjusted life-years (QALYs) or similar measures are recommended. For example, sensitivity analyses should be performed and cost-effectiveness acceptability curves published to account for variability in the costs and outcomes by location, population, and clinical context (7–11).

Results

The following questions were discussed, with group consensus summarized after each question:

1. **Should we organize our comments by biomarker category (risk assessment, cancer detection, and diagnosis) or by potential clinical use (risk mitigation,**

screening, and symptom and lung nodule evaluation)?

The project steering committee recognized that biomarkers within a given biomarker category may have more than one clinical use (Table 2). As this statement is intended to be a guide to help determine when a molecular biomarker is ready for clinical use for the early detection of lung cancer, the committee determined that it was best to organize our comments by the most relevant clinical applications (lung cancer screening and lung nodule evaluation).

2. **What level of evidence would support the clinical use of a validated lung cancer biomarker? Does the level of evidence differ based on the intended clinical use?**

Large, well-designed and conducted studies, capable of determining the impact of testing on clinical decisions and outcomes, are required to confirm the clinical utility of a molecular biomarker. The clinical application, intended use population, consequences of true and false results, and current state of clinical care in the field can influence the required level of evidence. The steering committee has provided guidance about the level of evidence (study outcomes and study design) that would support the clinical use of a biomarker for lung cancer screening and lung nodule evaluation (fit for purpose) in the clinical application sections of the DISCUSSION portion of this document.

3. **a. What amount of time before a diagnosis of lung cancer would be considered adequate to separate a risk prediction biomarker from a cancer detection biomarker?**
 - b. **Should a biomarker of risk of “being diagnosed” with lung cancer or of “dying of” lung cancer be defined?**
 - c. **Is “being diagnosed” with lung cancer or “developing” lung cancer a better way to define a risk prediction biomarker?**

The steering committee recognized that it is difficult to distinguish between a biomarker that predicts the risk of lung cancer developing over a period of time and a cancer detection biomarker. A biomarker whose intended use is to determine the risk

of developing lung cancer, may identify lung cancer that is present but cannot be detected by currently available means. To be defined as a biomarker of risk, the reference standard available at the time the biomarker sample is collected should be used to exclude the presence of lung cancer. A time interval between when the biomarker sample was collected and lung cancer was identified may be established to add further reassurance that the cancer was not present at the time of sample collection, or the accuracy of the biomarker at various time intervals from collection to diagnosis can be assessed.

There are different arguments for and against defining a risk prediction biomarker as one that would predict the risk of developing, being diagnosed with, or dying of lung cancer. The accuracy of the biomarker would be influenced by the definition. Not all lung cancers will be diagnosed, making it impossible to confirm how many cancers develop (although statistical models can be used to estimate the number). A definition that requires death from lung cancer would exclude overdiagnosed cancers but would not be able to address the influence of treatment or competing risks of death on the accuracy of the biomarker. The steering committee concluded that it is more important that the definition of a risk prediction biomarker be clearly described when it is used or studied than it is to mandate a single standardized definition. A general definition is provided in Table 2.

4. **a. How do we define early detection of lung cancer?**
 - b. **Should we add language about potential lethality of the cancer? Should biomarkers applied for the early detection of lung cancer be considered only for asymptomatic individuals or include a group with symptoms undergoing evaluation?**

The World Health Organization describes two components of early detection of cancer (12). The first is early diagnosis through prompt action when symptoms or signs of cancer are present. The second is through screening someone at risk for having cancer but who is free of symptoms or signs. The steering committee recognized these and other acceptable definitions, such as diagnosing lung cancer at a stage that is more amenable to successful treatment (e.g., stage I, localized or locoregional) or

Table 2. Biomarker Categories, Definitions, and Clinical Applications: Clinical Validation

Biomarker Category	Biomarker Definition	Study Definitions	Clinical Applications
Risk prediction	Cancer is not present, based on current standards; biomarker assesses the probability of cancer developing and being diagnosed over time	<i>Target condition:</i> The diagnosis of lung cancer after a defined period of time <i>Target population:</i> Individuals without symptoms, signs, or current imaging evidence of the presence of lung cancer <i>Reference standard:</i> Biopsy-confirmed lung cancer after a defined period of time	Risk mitigation Screening
Cancer detection	Cancer is present but has not been detected. The patient may or may not have symptoms; biomarker identifies the potential presence of cancer	<i>Target condition:</i> Undetected lung cancer <i>Target population:</i> Individuals with or without symptoms or signs of the presence of lung cancer <i>Reference standard:</i> Biopsy-confirmed lung cancer	Screening Symptom evaluation
Diagnosis	A nodule, mass, or other imaging finding is known to be present, but its etiology has not been determined; biomarker helps to determine the probability that the finding is malignant	<i>Target condition:</i> Indeterminate lung nodule, mass, or other imaging finding <i>Target population:</i> Individuals with imaging showing an indeterminate lung nodule, mass, or other imaging finding <i>Reference standard:</i> Biopsy confirmation or surveillance imaging without growth for a period of time, in keeping with current guidelines	Evaluation of a lung nodule, mass, or other imaging finding

diagnosing lung cancer earlier than it would otherwise have been diagnosed. There was substantial debate about whether the definition of early detection of lung cancer should include lung cancer diagnosis when symptoms or signs of cancer are present, or only include lung cancer diagnosis in asymptomatic individuals. A consensus definition of early detection of lung cancer was not developed. We concluded that it is more important that the definition of early detection be clearly described within studies assessing the individual biomarkers, with distinction between biomarker-detected lung cancer in patients with and without symptoms.

To be clinically useful, an early detection biomarker should detect lung cancer that could be lethal if not detected and treated early. This should be considered when interpreting studies of early detection biomarkers. For the reasons described in question 3 above, the steering committee did not believe it was practical to mandate that lethality of the cancer be included in the definition of an early detection biomarker.

5. Should we include guidance about how accurate a clinically validated biomarker should be to consider evaluating it for clinical utility for each

potential clinical use of the biomarker? Do we compare potentially useful accuracies with current standard practice, prediction tools, and other biomarkers currently in use (e.g., positron emission tomography [PET] for lung nodule management)?

To justify the investment required to complete the clinical utility phase of biomarker development, the steering committee members agreed that it would be helpful to provide guidance about the minimal accuracy, as assessed in the clinical validation phase, that could lead to a positive clinical impact. This minimal accuracy would vary by clinical application, and be guided by an understanding of the consequences of true and false results (13). Several methods have been described to assist with the estimate of minimal accuracy, such as calculation of the optimal slope of the receiver operating characteristic curve (14). Here we describe a formula to help with this estimate (15).

Biomarkers are frequently optimized for sensitivity or specificity based on their intended clinical application. For tests where a positive biomarker result leads to an action whereas a negative biomarker result is associated with standard of care for the population, the formula states:

$\text{sensitivity}/(1 - \text{specificity}) \geq [(1 - \text{prevalence})/\text{prevalence}] \times \text{harm/benefit}$, where harm/benefit is the ratio of the net harm of a falsely positive test result to the net benefit of a true-positive test result.

For tests where a negative result leads to an action other than standard of care for the population, the formula states: $\text{specificity}/(1 - \text{sensitivity}) \geq [\text{prevalence}/(1 - \text{prevalence})] \times \text{harm/benefit}$, where harm/benefit is the ratio of the net harm of a falsely negative test result to the net benefit of a true-negative test result.

$\text{Sensitivity}/(1 - \text{specificity})$ is known as the *positive likelihood ratio*, and $\text{specificity}/(1 - \text{sensitivity})$ is 1 divided by the *negative likelihood ratio*. *Prevalence* refers to the percentage of cases in the intended use population. The harm/benefit ratio in the formulas can be articulated in one of two ways:

- a. $1/N$, where in the first scenario N is the maximum number of control subjects testing positive that is tolerated to benefit one case subject testing positive (or in the second formula, the maximum number of case subjects testing negative that is worth the benefit of one control subject testing negative). For example, if the biomarker is used as an upfront lung

Table 3. Results That Should Be Reported in Various Phases of Biomarker Evaluation

Clinical validation

- The sensitivity and specificity of a technically validated biomarker, with fixed interpretation of the biomarker result, applied to the intended use population, as compared with the reference standard for the clinical application
- The clinical features of the cancer and control groups in clinical validation studies of the biomarker compared with the intended use population
- Biomarker results for relevant clinical subgroups
- Biomarker performance compared with and combined with clinical calculators, standard practice, and/or clinician judgment

Clinical utility

- The frequency with which the biomarker result impacts a clinical decision
- The impact (benefit and harm) of patient management decisions on patient outcomes when the biomarker is used

Cost-effectiveness

- The cost-effectiveness of the biomarker compared with the currently accepted standards for the clinical application

cancer screening test, N could represent the largest acceptable number of patients without lung cancer who have a positive test result and therefore undergo computed tomography (CT) screening for every one patient with lung cancer who tests positive.

- b. $R/(1 - R)$, where R is the risk threshold above which procedures consequent to positive testing, and below which avoidance of procedures consequent to negative testing, seem worthwhile. For example, when evaluating a lung nodule with a biomarker, the risk threshold R could be the probability or risk that a lung nodule is malignant at which one is indifferent to choosing surveillance imaging versus active investigation.

It is important to compare and combine biomarker accuracies with testing available in current practice. The accuracy required to impact clinical care is dependent on the needs within a clinical application and the potential consequences of the results. This will vary by clinical application, highlighting the need to judge the utility of the biomarker on a fit-for-purpose basis.

6. a. **Should we include guidance about cost-effectiveness or leave this to society and regulators to decide?**
 b. **Is affordability an adequate outcome at the expense of accuracy?**

It is important that cost-effectiveness analysis be performed and reported within

each clinical application. Ideally, a third-party independent analysis would be performed to avoid the potential for bias. The steering committee members decided that defining a threshold of cost-effectiveness that is acceptable to society is beyond the scope of this statement.

7. **Should we include guidance about the components of study design and details of study results that should be reported for all phases of biomarker development?**

As the focus of this project is on clinical utility of early detection biomarkers, the steering committee agreed that we should suggest components of the study design and details of the study results that should be reported for the phases of biomarker development that most directly influence the interpretation of clinical utility. This includes clinical validation, clinical utility, and cost-effectiveness (Table 3).

Discussion: Clinical Applications of Molecular Biomarkers for the Early Detection of Lung Cancer

The goal of this project is to provide guidance about the evidence required to determine that a molecular biomarker for early lung cancer detection is ready for clinical use. The most relevant clinical applications of a biomarker for the early detection of lung cancer are the selection of individuals for further lung cancer

Table 4. Models of the Risk of Developing Lung Cancer

	Bach <i>et al.</i> (19)	Spitz <i>et al.</i> (20)	Cassidy <i>et al.</i> (21)	Tammemägi <i>et al.</i> (22)	Hoggart <i>et al.</i> (23)	Katki <i>et al.</i> (24)
Source	CARET	MDA	LLP	PLCO	EPIC	PLCO
Subjects	18,172 10–60 cpd for 25–55 yr	3,852 never- and ever-smokers	1,736 never- and ever-smokers	80,375 ever-smokers	169,035 ever-smokers	105,556 ever-smokers
Age, yr	50–75	20–80	20–80	55–74	35–65	55–74
Variables	Age Asbestos Sex Smoking	Age Dust Emphysema Family history Sex Smoking	Age Asbestos Family history Pneumonia Prior cancer Sex Smoking	Age BMI Chest X-ray COPD Education Family history Smoking	Age Smoking	Age BMI Education Emphysema Family history Race Sex Smoking

Definition of abbreviations: BMI = body mass index; CARET = Carotene and Retinol Efficacy Trial; COPD = chronic obstructive pulmonary disease; cpd = cigarettes per day; EPIC = European Prospective Investigation into Cancer and Nutrition; LLP = Liverpool Lung Project; MDA = M. D. Anderson; PLCO = Prostate, Lung, Colorectal, and Ovarian Screening Trial.
 Reprinted by permission from Reference 39.

screening, and assistance with the characterization and management of lung nodules. For these two clinical applications, we discuss currently accepted practice, and then outline (1) the potential clinical utility and category of the biomarker, (2) the potential impact of applying the biomarker, (3) the level of evidence and accuracy that could support assessment of clinical utility, and (4) the level of evidence required to confirm the clinical utility of the biomarker.

Lung Cancer Screening

Current state. The National Lung Screening Trial (NLST) randomized more than 53,000 people at high risk of developing lung cancer (ages, 55–74 yr; active or former smokers of at least 30 pack-years who had smoked within the past 15 yr) to receive a baseline and two annual low-radiation-dose chest CT scans or a baseline and two annual chest radiographs. Fewer people in the chest CT arm died of lung cancer (16).

Several potential harms from lung cancer screening have been described. For example, lung nodules are frequently identified. Although usually benign, their identification leads to patient distress, additional imaging, and nonsurgical and surgical biopsies, all with potential complications. Radiation exposure during chest imaging and the evaluation and treatment of overdiagnosed lung cancers are other harms that have been reported (17, 18).

Multiple models exist to help estimate the risk of developing lung cancer (Table 4) (19–24). One model, PLCO (Prostate, Lung, Colorectal, and Ovarian Screening Trial, 2012), was evaluated in comparison with the NLST criteria, showing marginally improved sensitivity with similar specificity for identifying patients with lung cancer (25). At this time, it is not clear whether having a risk of developing lung cancer equal to that of the cohort obtained using NLST criteria, based on factors included in a risk model, will result in a similar balance of benefit to harm from lung cancer screening.

The cost-effectiveness of lung cancer screening by low-dose CT, based on the NLST data, was estimated to be \$81,000 per QALY gained, which is within the range typically considered cost-effective. Sensitivity analysis showed a range of \$32,000–\$615,000. Factors associated with this variability included the patient’s lung

Table 5. Trade-Offs with Biomarker Use before Computed Tomography Screening among People Who Meet the National Lung Screening Trial Screening Criteria

Sensitivity (%)	Specificity (%)	PPV (%)	NNTS	NPV (%)	Cancers Found (%)
90	90	7.0	14	99.91	90
80	80	3.2	31	99.79	80
80	40	1.1	91	99.58	80

Definition of abbreviations: NNTS = number needed to screen; NPV = negative predictive value; PPV = positive predictive value. The NNTS to diagnose lung cancer is lower than screening all currently eligible patients (prevalence, 0.83%; NNTS of 120 during the years of active screening in the National Lung Screening Trial), whereas the percentage of cancers found is lower than what would be diagnosed according to current eligibility standards. This assumes that all lung cancers in patients with true-positive biomarker results are identified by subsequent computed tomography imaging.

cancer risk, sex, age, and smoking status at the time of screening (26).

Potential clinical utility of a molecular biomarker applied as an initial test in a screening context. A clinically useful molecular biomarker applied as the initial test for lung cancer screening may improve the balance of benefit to harm of lung cancer screening by identifying those most likely to benefit from screening while minimizing exposure to harm among those least likely to benefit.

Category of biomarker. A molecular biomarker applied as an initial test in a screening context would be either a risk prediction or cancer detection biomarker.

Potential impact of applying a molecular biomarker as the initial test in a screening context.

- True-positive results—more individuals with lung cancer could be identified at curable stages.
- True-negative results—individuals without lung cancer (or at low risk of developing lung cancer) could avoid the harms associated with low-dose CT screening.
- False-positive results—individuals without lung cancer (or at low risk of developing lung cancer) could be enrolled in a low-dose CT screening program and be exposed to the associated harms.
- False-negative results—individuals with (or who will develop) lung cancer may not be enrolled in, and thus not have an opportunity to benefit from, a low-dose CT-based lung cancer screening program.

Level of evidence suggested to determine whether a biomarker applied to lung cancer screening justifies an assessment of clinical utility.

- The biomarker should be more accurate at identifying patients with (or who will develop) potentially curable lung cancer

than current eligibility criteria and available clinical risk prediction calculators, alone or in combination.

- If the biomarker is applied to the population that is currently *eligible* for screening, a judgment of an acceptable harm/benefit ratio of using the biomarker would have to be made. An example of trade-offs based on test accuracies is provided in Table 5.
- If the biomarker is applied to the population that is currently *not eligible* for lung cancer screening, an estimate of the required biomarker accuracy can be calculated by applying the accepted harm/benefit balance within the currently eligible group to the intended use population. For example, the population currently eligible for lung cancer screening was based on the NLST results, where the incidence of lung cancer was approximately 1 out of 120 participants (0.83%) during the active screening years (27). This can represent an accepted risk threshold level. If we set a goal of expanding the eligible population for screening to include a lower risk population, such as those with

Table 6. Representative Biomarker Accuracies That Would Satisfy the Calculation Yielding a Sensitivity/(1 – Specificity) ≥ 4.18

Sensitivity (%)	Specificity (%)
40	90.4
50	88.0
60	85.6
70	83.4
80	80.9

a clinical risk of 1 out of 500 participants (0.2%) having lung cancer, the formula described in question 5 of RESULTS shows the minimum positive likelihood ratio as $\text{sensitivity}/(1 - \text{specificity}) \geq [(1 - 0.002)/0.002] \times 0.0083/(1 - 0.0083) = 4.18$. Representative biomarker accuracies that would meet this standard are shown in Table 6.

Evidence required for a molecular biomarker to be considered clinically useful in the context of lung cancer screening. To be considered clinically useful, a molecular biomarker used to identify patients eligible for lung cancer screening must lead to:

- Fewer lung cancer deaths in the population tested compared with the current standard of care for that population, without substantially increasing harms and expense, or
- A similar number of lung cancer deaths in the population tested compared with the current standard of care for that population, with fewer harms or less expense.

It can be difficult, time consuming, and resource intensive to perform studies capable of proving clinical utility. Examples of study designs that could be used to obtain the above outcomes include the following (Figure 2):

BIOMARKER STRATIFIED.

- A molecular biomarker is measured retrospectively from archived samples obtained prospectively during a controlled trial of lung cancer screening.
- A molecular biomarker is measured prospectively during a controlled trial of lung cancer screening in which the biomarker result is not used to guide which lung cancer screening intervention the patient is assigned to.
- A molecular biomarker is measured prospectively, and the result is used to stratify patients into two arms (positive result and negative result). Patients in each arm are then randomized to one of two screening interventions (one of which may be standard of care for the population).

Note: These designs can be incorporated into any controlled trial of lung cancer screening.

ENRICHMENT.

- A molecular biomarker is measured prospectively, and the result is used to stratify patients into two arms (positive result and negative result). Patients in the biomarker-positive arm are then

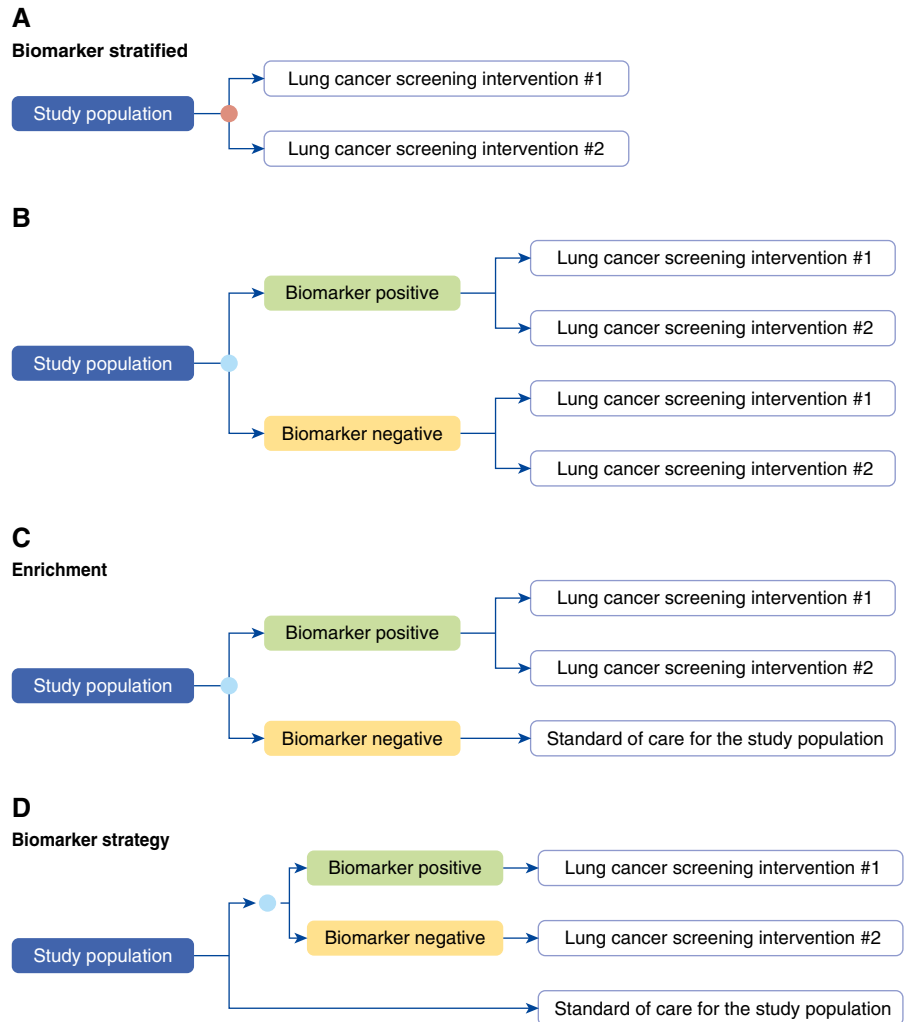


Figure 2. Examples of study designs capable of assessing the study outcomes of interest in a trial assessing the clinical utility of a lung cancer screening molecular biomarker. “Lung cancer screening intervention” may represent low-dose computed tomography screening, no screening, standard of care for the population, or another screening test. Biomarker stratified: (A) Biomarker (orange circle) is measured retrospectively from prospectively collected archived samples, or is measured prospectively, but the results are not used to determine the study arm. (B) Biomarker (blue circle) is measured prospectively, and the result is used to stratify patients. Ideally, both biomarker-positive and biomarker-negative patients are randomized to one of the screening interventions. Enrichment: (C) In populations at very low risk of having lung cancer it may not be practical to randomize those in the biomarker-negative arm to a screening intervention other than standard of care for the studied population. Biomarker strategy: (D) The study population is randomized to a biomarker (blue circle)-strategy arm, where the receipt of the screening intervention is based on the biomarker results, and an arm where all receive standard of care for the study population.

randomized to one of two screening interventions (one of which may be standard of care for the population). Patients in the biomarker-negative arm receive standard of care for the study population.

Note: This design may be preferred in study cohorts at very low risk of having lung cancer as it may not be practical to randomize those with a negative biomarker

result to a screening intervention other than standard of care.

BIOMARKER STRATEGY.

- A prospective controlled trial in which study subjects are randomized to a study arm where the receipt of the screening intervention is based on the result of a molecular biomarker, and a second arm where subjects receive standard of care for the study population.

Note: This design can be used in currently eligible or ineligible populations. It is less efficient than biomarker-stratified designs and cannot address whether the screening intervention is effective regardless of biomarker status.

Cost-effectiveness considerations for a lung cancer screening biomarker.

- It is important to consider whether the additional costs of the test would result in enough benefit to prove cost-effective (e.g., based on QALYs or incremental cost-effectiveness ratio) in comparison with currently practiced screening strategies alone within the intended use population.
- To be cost-effective, a test applied to a large population with a relatively low incidence of lung cancer (such as in screening) would need to be relatively inexpensive.

Lung Nodule Evaluation

Current state. Lung nodule management algorithms, based on the probability of malignancy, are available for solid subcentimeter nodules, solid larger (1- to 3-cm) nodules, and for subsolid nodules (28–31).

Solid subcentimeter lung nodules have a low probability of being malignant and are difficult to characterize by additional imaging or nonsurgical biopsies. Thus, surveillance imaging is the most appropriate management strategy. The interval and duration of surveillance are based on the size of the nodule.

Solid nodules larger than 1 cm have a higher probability of malignancy. Lung nodule risk calculators have been developed for this group (32, 33). Additional imaging and nonsurgical biopsies are more helpful for characterizing these nodules as benign or malignant. Very low-risk nodules enter a surveillance strategy, low- to moderate-risk nodules can be further characterized by PET imaging and/or nonsurgical biopsy, whereas high-risk nodules may proceed directly to surgical resection. PET imaging has a sensitivity for malignancy near 90%, whereas the specificity is lower and more variable (61–77%) (34). Nonsurgical biopsies have a yield of 60–80% and carry risks of bleeding and pneumothorax (35, 36). Approximately one out of four surgical biopsies is performed for a benign nodule (37).

Subsolid nodules have a higher baseline risk of malignancy than solid nodules of equal size, but are generally more indolent in their behavior when malignant. The higher probability of malignancy and less aggressive behavior inform the management algorithm

Table 7. Pretest Probabilities Required for the Stated Test Accuracies to Impact a Clinical Decision if the Test Result Is Positive or Negative

Sensitivity (%)	Specificity (%)	PrTP to Give PoTP > 65%	PrTP to Give PoTP < 10%
90	90	>17%	<50%
80	80	>32%	<31%
80	40	>58%	<18%

Definition of abbreviations: PoTP = posttest probability; PrTP = pretest probability. Assuming for a given patient that an aggressive approach will be taken if the posttest probability is greater than 65% and a surveillance approach if the posttest probability is less than 10%, the values listed in the table represent the pretest probabilities required for the stated test accuracies to impact a clinical decision if the test is positive (third column) or negative (fourth column).

for subsolid nodules. Growth in the total size of a subsolid nodule, or growth of the solid component, strongly suggests the nodule is malignant. The low metabolic activity leads to a low yield from PET imaging. Nonsurgical biopsies also have a relatively low yield (38).

Potential clinical utility of a molecular biomarker applied to assist with the characterization of a lung nodule. A clinically useful molecular biomarker applied to the evaluation of lung nodules may lead to expedited therapy for early lung cancer and/or

fewer aggressive interventions in patients with benign lung nodules.

Category of biomarker. A molecular biomarker applied to assist with the characterization of a lung nodule would be in the diagnosis biomarker category.

Potential impact of applying a molecular biomarker for lung nodule management.

- True-positive results—individuals with malignant lung nodules could be identified sooner or with fewer costly and/or invasive interventions.

Sensitivity	Specificity	Probability	Expedite Rx	Avoid Rx	Overaggressive Evaluation	Delay Rx
90	90	1	9	891	99	1
		40	360	540	60	40
		90	810	90	10	90
80	80	1	8	792	198	2
		40	320	480	120	80
		90	720	80	20	180
80	40	1	8	396	594	2
		40	320	240	360	80
		90	720	40	60	180

Figure 3. Trade-offs associated with lung nodule management biomarker accuracies. For the three test accuracies listed, the test is applied to a theoretical cohort of 1,000 patients with a probability of malignancy of 1, 40, and 90%. The numbers represent the impact of clinical decisions based on an interpretation of the test result as positive or negative, where positive leads to more aggressive management and negative to less aggressive management. It is assumed that for those with a probability of malignancy of 1% surveillance would have been recommended without the test, and for those with a probability of malignancy of 90% an aggressive management approach would have been advised. *Black* numbers suggest no change in management based on the test result. *Green* numbers suggest benefit, whereas *red* numbers suggest potentially avoidable harm. Rx = treatment.

- True-negative results—individuals with benign lung nodules could avoid costly and/or invasive testing.
- False-positive results—individuals with benign lung nodules could undergo costly and/or invasive testing including surgical resection.
- False-negative results—individuals with malignant lung nodules may not receive an aggressive diagnostic evaluation, potentially delaying treatment.

Level of evidence suggested to determine whether a biomarker applied to lung nodule management justifies an assessment of clinical utility.

- A lung nodule molecular biomarker should improve on the accuracy of clinician judgment, lung nodule risk calculators, and fludeoxyglucose F 18–PET imaging, alone or in combination.
- The accuracy of the biomarker should be high enough to suggest that it could move the pretest probability of malignancy beyond a clinical decision threshold (Table 7). For example, one can categorize the probability that the nodule is malignant based on the interventions recommended in each category—surveillance imaging, additional testing (imaging or nonsurgical biopsy), and surgical resection. The biomarker should be able to shift a patient from one category of intervention to another. The result of this shift should lead to improved clinical outcomes in the intended use population.
- On balance, the consequences of applying the biomarker, if the result was interpreted as a dichotomous positive or negative, are likely to lead to improved clinical outcomes. The values of the benefits and harms of clinical decisions that are impacted by test results are not equivalent, and thus a judgment about acceptable trade-offs must be considered (Figure 3).
- Calculations can be performed on the basis of an estimate of the relative benefit of a true-positive (or -negative) result to that of a false-positive (or -negative) result to estimate the accuracy required for a biomarker to have clinical benefit. The estimate of required accuracy will differ, based on whether the test will be used to rule in or rule out cancer, as the relative benefit of true and false results for each use is not the same.
 - For example, for an intermediate-risk lung nodule with 40% probability of

malignancy, the benefit of a true-positive result would be to expedite treatment of a localized lung cancer, whereas the harm of a false-positive result may be aggressive management of a benign nodule (surgical resection) that would otherwise have been

evaluated as an intermediate-risk nodule (imaging or nonsurgical biopsy). If the cost–benefit ratio of aggressive management versus management as an intermediate-risk nodule is valued at 5:1 (one false-positive control to justify five true-positive cases), the minimal

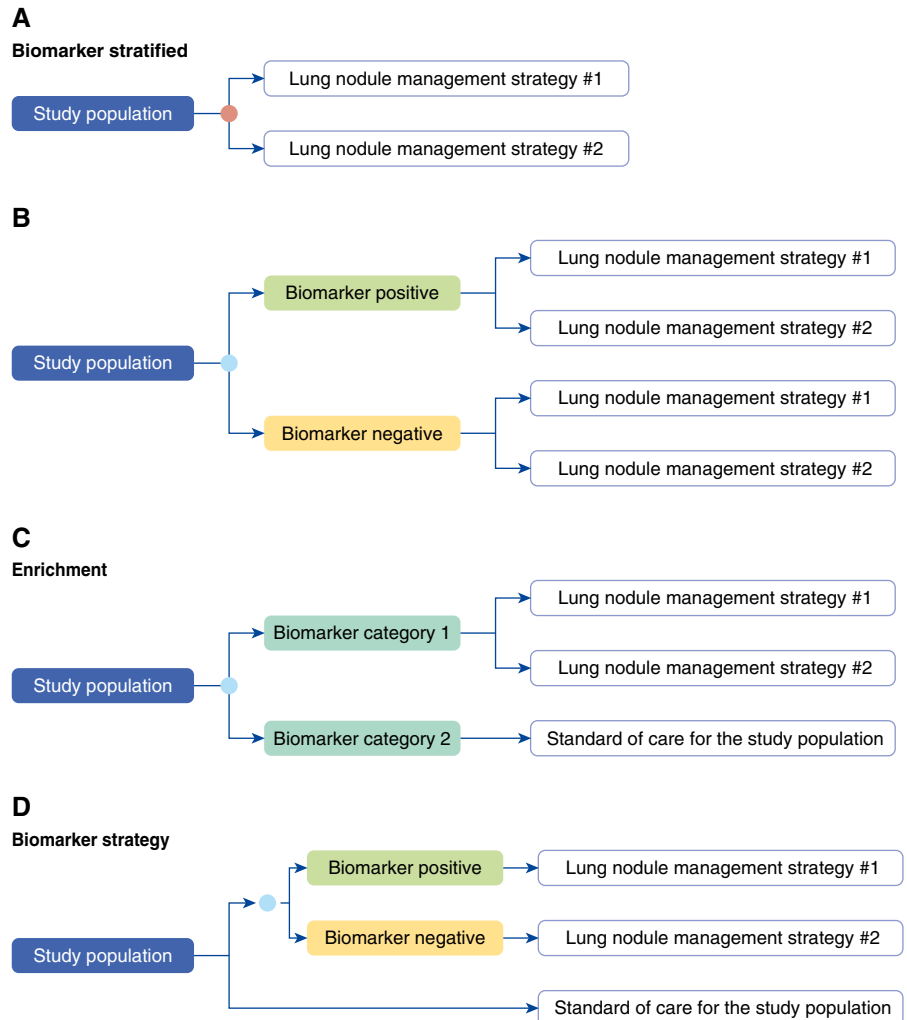


Figure 4. Examples of study designs capable of assessing the study outcomes of interest in a trial assessing the clinical utility of a lung nodule management molecular biomarker. “Lung nodule management strategy” could be surveillance imaging at a set interval, [¹⁸F]fluorodeoxyglucose–positron emission tomography imaging, nonsurgical biopsy, surgical resection, or clinician decision. Biomarker stratified: (A) Biomarker (orange circle) is measured retrospectively from prospectively collected archived samples, or is measured prospectively, but the results are not used to determine the study arm. (B) Biomarker (blue circle) is measured prospectively and the result is used to stratify patients. Both biomarker-positive and biomarker-negative patients are randomized to one of the lung nodule management strategies. Enrichment: (C) In populations with very low-risk or very high-risk lung nodules it may not be practical to randomize those in the biomarker-negative arm (for those at very low risk) or biomarker-positive arm (for those at very high risk) to a screening intervention other than standard of care for the studied population. Biomarker strategy: (D) The study population is randomized to a biomarker (blue circle)–strategy arm, where the lung nodule management strategy is based on the biomarker results, and an arm where all receive standard of care for the study population. Those in the second arm could also be randomized to different lung nodule management strategies. Alternatively, the biomarker could be measured for the entire study population but only used to determine the management strategy for a defined portion of the study population.

positive likelihood ratio, calculated using the above formula, would be as follows: $\text{sensitivity}/(1 - \text{specificity}) \geq [(1 - 0.4)/0.4] \times 5 = 7.5$. For example, a sensitivity of 75% and specificity of 90% ($1 - \text{specificity} = 10\%$) would equal 7.5.

- Conversely, the benefit of a true-negative result may be to avoid standard evaluation of an intermediate-risk nodule in favor of surveillance, and the harm of a false-negative result may be to delay the diagnosis of a localized lung cancer. If the cost-benefit ratio of surveillance versus evaluation of an intermediate-risk nodule is valued at 3:1 (one false-negative case to justify three true-negative controls), a calculation using the above formula would show a requirement that the $\text{specificity}/(1 - \text{sensitivity}) \geq [0.4/(1 - 0.4)] \times 3 = 2$. For example, a specificity of 40% and sensitivity of 80% ($1 - \text{sensitivity} = 20\%$) would equal 2.

Level of evidence required for a molecular biomarker to be considered clinically useful in the context of lung nodule management. To be considered clinically useful, a molecular biomarker used to assist with lung nodule management must lead to:

- Earlier diagnosis of malignant nodules without substantially increasing the number of procedures performed on patients with benign nodules, or
- Fewer procedures for patients with benign nodules without substantially delaying the diagnosis of cancer in patients with malignant nodules.

Examples of study designs that could be used to obtain the above outcomes include the following (Figure 4):

BIOMARKER STRATIFIED.

- A molecular biomarker is measured retrospectively from archived samples obtained prospectively during a controlled trial of lung nodule management strategies.

- A molecular biomarker is measured prospectively during a controlled trial of lung nodule management strategies in which the biomarker result is not used to guide which lung nodule management strategy the patient is assigned to.
- A molecular biomarker is measured prospectively, and the result is used to stratify patients into two arms (positive result and negative result). Patients in each arm are then randomized to one of two lung nodule management strategies.

Note: These designs can be incorporated into any controlled trial of lung nodule management strategies.

ENRICHMENT.

- A molecular biomarker is measured prospectively, and the result is used to stratify patients into two arms (positive result and negative result). Patients in one of the arms (biomarker positive or biomarker negative) are then randomized to one of two lung nodule management strategies. Patients in the other biomarker arm receive standard of care for the study population.

Note: This design may be preferred in study cohorts with very low- or very high-risk lung nodules as it may not be practical to randomize those with a negative biomarker result (in a very low-risk nodule cohort) or positive biomarker result (in a very high-risk cohort) to a nodule management strategy other than standard of care.

BIOMARKER STRATEGY.

- A prospective controlled trial in which study subjects are randomized to one of two study arms. In the first, the nodule management strategy is based on the result of a molecular biomarker. In the second, subjects could either receive standard of care for the study population or be randomized to nodule management strategies without use of the biomarker. Alternatively, the biomarker could be measured for the entire study population but only used to determine the

management strategy for a defined portion of the study population.

Note: This design can be used to address whether a biomarker-driven strategy is better than a standard management strategy. It is less efficient than biomarker-stratified designs and cannot address whether the nodule management strategy is effective regardless of biomarker status.

Cost-effectiveness considerations for a biomarker used for the evaluation of lung nodules.

- The costs related to nodule evaluation may vary, based on differences in practice location, resources, clinical judgment, and patient populations.
- Cost-effectiveness analyses should include a description of the study setting, costs, benefits, and how they were calculated. A sensitivity analysis is particularly important.

Conclusions

The application of molecular biomarkers to assist with the early detection of lung cancer has the potential to substantially improve our ability to select patients for lung cancer screening, and to assist with the characterization of indeterminate lung nodules. To support the application of molecular biomarkers in these clinical settings there must be evidence that the molecular biomarker leads to clinical decisions whose benefits outweigh their harms. Although it is tempting to apply novel testing based on promising discovery or validation level studies, the lung cancer community should insist on additional evidence of clinical utility before changing practice. We have described relevant considerations and have suggested standards to apply when determining whether a molecular biomarker for the early detection of lung cancer is ready for clinical use. ■

This official statement was prepared by an *ad hoc* subcommittee of the Assembly on Thoracic Oncology.

Members of the subcommittee are as follows:

PETER J. MAZZONE, M.D., M.P.H. (*Co-Chair*)
 CATHERINE RUFATTO SEARS, M.D. (*Co-Chair*)
 DOUG A. ARENBERG, M.D.
 MINA GAGA, M.D., Ph.D.

MICHAEL K. GOULD, M.D., M.S.
 PIERRE P. MASSION, M.D.
 VISH S. NAIR, M.D., M.S.
 CHARLES A. POWELL, M.D.
 GERARD A. SILVESTRI, M.D., M.S.
 ANIL VACHANI, M.D.

RENDA SOYLEMEZ WIENER, M.D., M.P.H.

Author disclosures: P.J.M. received research support from InDi, 20/20 Genesystems, Metabolomx, and Oncimmune; and served on an advisory committee for Genentech USA, Grail, InDi, Nucleix, and Oncimmune. D.A.A. served on an advisory committee for Nucleix

and VisionGate. M.G. served on an advisory committee for AstraZeneca, Chiesi, GlaxoSmithKline, Menarini, Novartis, and Pharmaten; served as a consultant for Boehringer Ingelheim; served as a speaker for Teva; and received research support from AstraZeneca, Boehringer Ingelheim, Chiesi, Elpen, GlaxoSmithKline, Novartis, and Teva. M.K.G. received author royalties from UpToDate and received research support from Medial. C.A.P. served on the advisory

committee for Genentech USA; and served as a consultant for Genentech USA, Nucleix, Siemen, Chugai Pharmaceutical, and Takeda Pharmaceutical. G.A.S. received research support from Integrated Diagnostics, Oncocyte, and Exact Sciences. A.V. served on an advisory committee for Allegro Diagnostics and Veracety; received research support from Allegro Diagnostics, Integrated Diagnostics, Janssen Research and Development, MagArray, and Viomics; and

provided expert testimony for Ford Motor Company and Honeywell International. C.R.S., P.P.M., V.S.N., and R.S.W. reported no relationships with relevant commercial interests.

Acknowledgment: The writing committee is grateful for the guidance of Gene Pennello and Stuart Baker and the inspiration provided by Dennis O'Brien.

References

- Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88–94.
- Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst* 2010;102:152–160.
- Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per Med* 2010;7:33–47.
- Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009;101:1446–1452.
- Micheel CM, Nass SJ, Omenn GS, editors. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; Institute of Medicine. Chapter 4: Evaluation of omics-based tests for clinical utility and use. In: Evolution of translational omics: lessons learned and the path forward. Washington, DC: National Academies Press; 2012 [accessed 2015 Dec 15]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK202169/>
- McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975;293:211–215.
- Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC; Panel on Cost-Effectiveness in Health and Medicine. The role of cost-effectiveness analysis in health and medicine. *JAMA* 1996;276:1172–1177.
- Siegel JE, Weinstein MC, Russell LB, Gold MR; Panel on Cost-Effectiveness in Health and Medicine. Recommendations for reporting cost-effectiveness analyses. *JAMA* 1996;276:1339–1341.
- Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA* 1996;276:1253–1258.
- Edejer TTT, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans DB, Murray CJL. Part 1: Methods for generalized cost-effectiveness analysis. In: Making choices in health: WHO guide to cost-effectiveness analysis. Geneva, Switzerland: World Health Organization; 2003 [accessed 2015 Dec 15]. pp. 3–97. Available from: http://www.who.int/choice/publications/p_2003_generalised_cea.pdf
- Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, Kuntz KM, Meltzer DO, Owens DK, Prosser LA, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA* 2016;316:1093–1103.
- World Health Organization. Cancer: early detection of cancer. 2017 [accessed 2015 Dec 15]. Available from: <http://www.who.int/cancer/detection/en/>
- Ost DE, Gould MK. Decision making in patients with pulmonary nodules. *Am J Respir Crit Care Med* 2012;185:363–372.
- Baker SG. Improving the biomarker pipeline to develop and evaluate cancer screening tests. *J Natl Cancer Inst* 2009;101:1116–1119.
- Pepe MS, Janes H, Li CI, Bossuyt PM, Feng Z, Hilden J. Early-phase studies of biomarkers: what target sensitivity and specificity values might confer clinical utility? *Clin Chem* 2016;62:737–742.
- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD; National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- Agency for Healthcare Research and Quality. Benefits and harms of computed tomography lung cancer screening programs for high-risk populations. AHRQ Publication No. 13-05196-EF-2. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
- Harris RP, Sheridan SL, Lewis CL, Barclay C, Vu MB, Kistler CE, Golin CE, DeFrank JT, Brewer NT. The harms of screening: a proposed taxonomy and application to lung cancer screening. *JAMA Intern Med* 2014;174:281–285.
- Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, Hsieh LJ, Begg CB. Variations in lung cancer risk among smokers. *J Natl Cancer Inst* 2003;95:470–478.
- Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, Shete S, Etzel CJ. A risk model for prediction of lung cancer. *J Natl Cancer Inst* 2007;99:715–726.
- Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, Field JK. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* 2008;98:270–276.
- Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, Chaturvedi AK, Silvestri GA, Riley TL, Commins J, et al. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368:728–736.
- Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Østergaard JN, Kaaks R, Canzian F, Boeing H, Steffen A, et al. A risk model for lung cancer incidence. *Cancer Prev Res (Phila)* 2012;5:834–846.
- Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and validation of risk models to select ever-smokers for CT lung cancer screening. *JAMA* 2016;315:2300–2311.
- Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, Commins J, Berg CD. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. *PLoS Med* 2014;11:e1001764.
- Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, Naeim A, Church TR, Silvestri GA, Gorelick J, et al.; National Lung Screening Trial Research Team. Cost-effectiveness of CT screening in the National Lung Screening Trial. *N Engl J Med* 2014;371:1793–1802.
- Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, Clingan KL, Duan F, Fagerstrom RM, Gareen IF, et al.; National Lung Screening Trial Research Team. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med* 2013;369:920–931.
- Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, Wiener RS. Evaluation of individuals with pulmonary nodules: when is it lung cancer? In: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143(5 Suppl):e93S–e120S.
- MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, Patz EF Jr, Swensen SJ; Fleischner Society. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005;237:395–400.
- Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, Macchiarini P, Crapo JD, Herold CJ, Austin JH, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013;266:304–317.
- Callister MEJ, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, Franks K, Gleeson F, Graham R, Malhotra P, et al.; British Thoracic Society Pulmonary Nodule Guideline Development Group; British Thoracic Society Standards of Care Committee. British Thoracic

- Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;70:ii1–ii54.
32. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules. *Arch Intern Med* 1997; 157:849–855.
 33. Herder GJ, van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, Hoekstra OS. Clinical prediction model to characterize pulmonary nodules: validation and added value of ^{18}F -fluorodeoxyglucose positron emission tomography. *Chest* 2005;128:2490–2496.
 34. Deppen SA, Blume JD, Kensinger CD, Morgan AM, Aldrich MC, Massion PP, Walker RC, McPheeters ML, Putnam JB Jr, Grogan EL. Accuracy of FDG-PET to diagnose lung cancer in areas with infectious lung disease: a meta-analysis. *JAMA* 2014;312: 1227–1236.
 35. Wang Memoli JS, Nietert PJ, Silvestri GA. Meta-analysis of guided bronchoscopy for the evaluation of the pulmonary nodule. *Chest* 2012;142:385–393.
 36. Ost DE, Ernst A, Lei X, Kovitz KL, Benzaquen S, Diaz-Mendoza J, Greenhill S, Toth J, Feller-Kopman D, Puchalski 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.
 37. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, Byers T, Colditz GA, Gould MK, Jett JR, *et al.* Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* 2012;307:2418–2429.
 38. Detterbeck FC, Homer RJ. Approach to the ground-glass nodule. *Clin Chest Med* 2011;32:799–810.
 39. Mazzone PJ. Obstacles to and solutions for a successful lung cancer screening program. *Semin Respir Crit Care Med* 2016;37:659–669.

Appendix 1: Definitions

Analytical performance: The ability of a biomarker assay to measure the underlying biomarker quantity under a variety of conditions.

Analytical specificity: The ability of a biomarker assay to measure solely the biomarker of interest without interference by other substances or cross-reactivity with other analytes.

Analytical validity: Acceptable performance in the measurement or detection of characteristics of a biomarker; how well the test measures or identifies molecular changes in a person.

Bias: A systematic difference between the compared groups that impacts results in a manner that does not reflect an underlying reality. The systematic erroneous association of a characteristic with a group in a way that distorts a comparison with another group.

Biomarker: A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention; a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease.

Clinical performance: The ability of the biomarker to inform about a clinical condition of interest.

Clinical reference standard: The best available method for establishing the presence or absence of the target condition.

Clinical utility: A biomarker's ability to improve clinical outcomes when measured and used as directed for its intended use.

Clinical validity: The demonstrated association of a test result with the presence or absence of the target condition.

Companion diagnostic: A biomarker that is essential for the safe and effective use of a therapy.

Confidence intervals: An interval about a point estimate that quantifies the statistical uncertainty in the true value being estimated (e.g., an accuracy metric) due to variability in the subject/sample selection process. A $1 - \alpha$ level confidence interval contains the true value in $100(1 - \alpha)\%$ of applications (but in any given application either contains it or does not).

Context of use: What the test measures, why, and in what population(s) it should be used.

Cost-effectiveness analysis: An evaluation that compares the net cost of an intervention with the benefits gained by that intervention.

Diagnostic accuracy: The extent of agreement between the outcome of the new test and the reference standard.

Diagnostic marker: A test used in people with signs or symptoms to aid in assessing whether they have a condition.

External validity: The generalizability of the comparison results to persons outside of the study.

Fundamental comparison: The process of arranging and analyzing groups of subjects and specimens to learn whether a difference observed in the compared groups is related to a particular condition.

Independent validation: The assessment of analytical and clinical performance on a set of subjects that is independent of the data set used in the development of the test.

Index test: The test under evaluation.

Intended use of the test: The population, condition, and question for which the biomarker is being developed (e.g., diagnosis, staging, screening, surveillance, prediction, prognosis).

Intention to diagnose analysis: An analysis that includes all study subjects, whether or not all of their test results are available.

Intermediate precision: Precision when some conditions vary and others are held constant.

Internal validity: The strength or fairness of the comparison of groups within the study.

Lead time: The length of time the diagnosis is advanced by testing with the biomarker.

Limit of detection: The lowest assay level at which the presence of the analyte is detected with reliability in repeated measurement.

Linearity: The ability to provide results directly proportional to the amount of analyte in the test sample within a given measuring range.

Measurement accuracy: The closeness of agreement between a measurement result and an accepted reference value; an aggregate of trueness and precision.

Measurement trueness: The closeness of agreement between the average of an infinite number of replicate test results and the reference value. Trueness is usually expressed numerically by the statistical measure bias.

Medical tests: Results of a clinical, imaging, or laboratory-based assay that are used alone or with other information to help assess a subject's health condition of interest, or target condition.

Negative predictive value (NPV): The predictive value of a negative result; the proportion of subjects with a negative test result who do not have the target condition.

Performance around the cutoff: The measurement accuracy of an assay at biomarker levels near the threshold chosen to distinguish a positive and negative result for the intended use of the test.

Positive predictive value (PPV): The predictive value of a positive result; the proportion of subjects with a positive test result who have the target condition.

Precision: The closeness of agreement of replicate test results under stipulated conditions.

Predictive marker: A biomarker that assesses the safety or efficacy of a specific therapy.

Prognostic marker: A biomarker used in subjects diagnosed with a condition to predict subsequent outcomes, such as disease recurrence or progression.

Qualitative result: A biomarker result consisting of a set number of possible responses (often two).

Quantitative result: A biomarker result that is numerical in amount or level of a physical quantity.

Repeatability: Precision when replicate measurements are taken under the same conditions (within a run).

Reproducibility: Precision when one of the conditions being varied is the laboratory for an *in vitro* diagnostic measurement.

Role of the test: The position of the index test relative to other tests for the same condition.

Screening marker: A biomarker used in asymptomatic people to detect a disease or condition at an early stage.

Semiquantitative: Results of a test that fall into an approximate range of values.

Sensitivity: The proportion of subjects with the target condition in whom the test result is positive.

Signature: Multiple variables combined to provide a single result.

Specificity: The proportion of subjects without the target condition in whom the test result is negative.

Target condition: The disease or condition that the index test is expected to detect.

Tumor marker: A qualitative or quantitative alteration or deviation from normal of a molecule, substance, or process that can be detected by some type of assay; surrogate indicators that increase or decrease the clinician's suspicion that future clinically important events, such as cancer onset, recurrence, progression or patient death, will or will not happen, and/or that a specific treatment will decrease the risk of such events.

Appendix 2: Additional References Used during the Project

- Baker SG. Biomarker evaluation in randomized trials: addressing different research questions. *Stat Med* 2014;33:4139–4140.
- Baker SG, Kramer BS, Peirce, Youden, and receiver operating characteristic curves. *Am Stat* 2007;61:343–346.
- Baker SG, Kramer BS, Srivastava S. Markers for early detection of cancer: statistical guidelines for nested case–control studies. *BMC Med Res Methodol* 2002;2:4.
- Beaver JA, Tzou A, Blumenthal GM, McKee AE, Kim G, Pazdur R, Phillip R. An FDA perspective on the regulatory implications of complex signatures to predict response to targeted therapies. *Clin Cancer Res* 2017;23:1368–1372.
- Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; 69:89–95.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HCW, et al.; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- Center for Devices and Radiological Health, Food and Drug Administration, U.S. Department of Health and Human Services. Guidance for industry and FDA staff: statistical guidance on reporting results from studies evaluating diagnostic tests. 2007 March 3 [accessed 2015 Dec 15]. Available from: <https://www.fda.gov/RegulatoryInformation/Guidances/ucm071148.htm>
- Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;13:1.
- Duffy MJ, Sturgeon CM, Sölétormos G, Barak V, Molina R, Hayes DF, Diamandis EP, Bossuyt PM. Validation of new cancer biomarkers: a position statement from the European Group on Tumor Markers. *Clin Chem* 2015;61:809–820.
- Gluud C, Gluud LL. Evidence based diagnostics. *BMJ* 2005;330:724–726.

- Hayes DF, Bast RC, Desch CE, Fritsche H Jr, Kemeny NE, Jessup JM, Locker GY, Macdonald JS, Mennel RG, Norton L, *et al*. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456–1466.
- Henry NL, Hayes DF. Cancer biomarkers. *Mol Oncol* 2012;6:140–146.
- Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M, Silvestri GA, Chaturvedi AK, Katki HA. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med* 2013;369:245–254.
- Mazzone P, Powell CA, Arenberg D, Bach P, Detterbeck F, Gould MK, Jaklitsch MT, Jett J, Naidich D, Vachani A, *et al*. Components necessary for high-quality lung cancer screening: American College of Chest Physicians and American Thoracic Society policy statement. *Chest* 2015;147:295–303.
- McMahon PM, Kong CY, Bouzan C, Weinstein MC, Cipriano LE, Tramontano AC, Johnson BE, Weeks JC, Gazelle GS. Cost-effectiveness of computed tomography screening for lung cancer in the United States. *J Thorac Oncol* 2011;6:1841–1848.
- Moore HM, Kelly A, Jewell SD, McShane LM, Clark DP, Greenspan R, Hainaut P, Hayes DF, Kim P, Mansfield E, *et al*. Biospecimen reporting for improved study quality. *Biopreserv Biobank* 2011;9:57–70.
- Moyer VA; U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330–338.
- Office of Public Health Strategy and Analysis, Office of the Commissioner, U.S. Food and Drug Administration. The public health evidence for FDA oversight of laboratory developed tests: 20 case studies. 2015 Nov 16 [accessed 2015 Dec 15]. Available from: <https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM472777.pdf>
- Oken MM, Hocking WG, Kvale PA, Andriole GL, Buys SS, Church TR, Crawford ED, Fouad MN, Isaacs C, Reding DJ, *et al*.; PLCO Project Team. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. *JAMA* 2011;306:1865–1873.
- Patakya R, Gulati R, Etzioni R, Black P, Chi KN, Coldman AJ, Pickles T, Tyldesley S, Peacock S. Is prostate cancer screening cost-effective? A microsimulation model of prostate-specific antigen-based screening for British Columbia, Canada. *Int J Cancer* 2014;135:939–947.
- Pennello GA. Analytical and clinical evaluation of biomarkers assays: when are biomarkers ready for prime time? *Clin Trials* 2013;10:666–676.
- Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001;93:1054–1061.
- Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–1438.
- Pepe MS, Fan J, Seymour CW, Li C, Huang Y, Feng Z. Biases introduced by choosing controls to match risk factors of cases in biomarker research. *Clin Chem* 2012;58:1242–1251.
- Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. *J Clin Oncol* 2010;28:698–704.
- Romanus D, Cardarella S, Cutler D, Landrum MB, Lindeman NI, Gazelle GS. Cost-effectiveness of multiplexed predictive biomarker screening in non-small-cell lung cancer. *J Thorac Oncol* 2015;10:586–594.
- Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ* 2002;324:539–541.
- Schully SD, Carrick DM, Mechanic LE, Srivastava S, Anderson GL, Baron JA, Berg CD, Cullen J, Diamandis EP, Doria-Rose VP, *et al*. Leveraging biospecimen resources for discovery or validation of markers for early cancer detection. *J Natl Cancer Inst* 2015;107:djv012.
- Sharaf RN, Ladabaum U. Comparative effectiveness and cost-effectiveness of screening colonoscopy vs. sigmoidoscopy and alternative strategies. *Am J Gastroenterol* 2013;108:120–132.
- Sturgeon CM, Hoffman BR, Chan DW, Ch'ng SO, Hammond E, Hayes DF, Liotta LA, Petricoin EF, Schmitt M, Semmes OJ, *et al*.; National Academy of Clinical Biochemistry. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in clinical practice: quality requirement. *Clin Chem* 2008;54:e1–e10.
- Villanti AC, Jiang Y, Abrams DB, Pyenson BS. A cost-utility analysis of lung cancer screening and the additional benefits of incorporating smoking cessation interventions. *PLoS One* 2013;8:e71379.
- Wu D, Rosner GL, Broemeling LD. Bayesian inference for the lead time in periodic cancer screening. *Biometrics* 2007;63:873–880.
- Zauber AG. Cost-effectiveness of colonoscopy. *Gastrointest Endosc Clin N Am* 2010;20:751–770.