Title

Methods for the Watch the Spot Trial. A Pragmatic Trial of More- versus Less-Intensive Strategies for Active Surveillance of Small Pulmonary Nodules.

Journal

Annals of the American Thoracic Society, 16(12)

ISSN

2329-6933

Authors

Gould, Michael K
Smith-Bindman, Rebecca
Kelly, Karen
et al.

Publication Date

2019-12-01

DOI

10.1513/annalsats.201903-268sd

Peer reviewed

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Methods for the Watch the Spot Trial: A Pragmatic Trial of More vs. Less Intensive Strategies for Active Surveillance of Small Pulmonary Nodules

Michael K. Gould, MD, MS[1*], Rebecca Smith-Bindman, MD[2], Karen Kelly, MD[3], Danielle E. Altman, MA[1], Igor Barjaktarevic, MD, PhD[4], Beth Creekmur, MA[1], Evan de Bie, BS[5], Debra S. Dyer, MD[6], Eduardo J. Mortani Barbosa Jr., MD[7] , Richard A. Mularski, MD[8], Lihong Qi, PhD[5], Laszlo T. Vaszar, MD[9], Sophronia Yu, MPH[2], Diana L. Miglioretti, PhD[5], on behalf of the Watch the Spot Trial Investigators

[1]Department of Research and Evaluation, Kaiser Permanente Southern California
[2]Department of Radiology and Biomedical Imaging, Department of Epidemiology and Biostatistics, and the Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco
[3]Department of Medicine, School of Medicine, University of California, Davis
[4]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles
[5]Department of Public Health Sciences, School of Medicine, University of California, Davis and Kaiser Permanente Washington Health Research Institute, Seattle, WA
[6]Department of Radiology, National Jewish Health
[7]Department of Radiology, Perelman School of Medicine, University of Pennsylvania
[8]The Center for Health Research, Kaiser Permanente Northwest
[9]Department of Medicine, Mayo Clinic Arizona
*MKG is a Deputy Editor of AnnalsATS. His participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.

Corresponding Author:
Michael K. Gould, MD, MS
Department of Research and Evaluation
100 South Los Robles Ave. Pasadena, CA 91101
michael.k.gould@kp.org
Phone: 626-564-3926

Keywords: pulmonary nodule, coin lesion, lung cancer, surveillance, screening, pragmatic trial, computed tomography

Word Count: 3,984

Abstract

Small pulmonary nodules are most often managed by surveillance imaging with chest computed tomography (CT), but the optimal frequency and duration of surveillance are unknown. The Watch the Spot Trial is a multi-center, pragmatic, comparative effectiveness trial with cluster randomization by hospital or health system that compares more vs. less intensive strategies for active surveillance of small pulmonary nodules. The study plans to enroll approximately 35,200 patients with a small pulmonary nodule that is newly detected on chest CT, either incidentally or by screening. Study protocols for more and less intensive surveillance were adapted from published guidelines. The primary outcome is the percentage of cancerous nodules that progress beyond American Joint Committee on Cancer 7th edition (AJCC 7) stage T1aN0M0. Secondary outcomes include patient-reported anxiety and emotional distress, nodule-related health care utilization, radiation exposure, and adherence with the assigned surveillance protocol. Distinctive aspects of the trial include: (1) the pragmatic integration of study procedures into existing clinical workflow; (2) the use of cluster-randomization by hospital or health system; (3) the implementation and evaluation of a system-level intervention for protocol-based care; (4) the use of highly efficient, technology-enabled methods to identify and (passively) enroll participants; (5) reliance on data collected as part of routine clinical care, including data from electronic health records and state cancer registries; (6) linkage with state cancer registries for complete ascertainment of the primary study outcome; and (7) intensive engagement with a diverse group of patient and non-patient stakeholders in the design and execution of the study.

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Pulmonary nodules are commonly identified on chest computed tomography (CT) scans, either as an incidental finding or by screening.(1, 2) While the majority of nodules are benign and harmless, up to 5% prove to be lung cancer.(3) It is important to identify cancerous nodules promptly because localized stage lung cancer can be treated and potentially cured. In the absence of suspicious features like spiculation, the standard of care for the management of most small pulmonary nodules is surveillance imaging to identify growth that is highly suggestive of malignancy, but evidence for the optimal frequency and duration of nodule surveillance is lacking. Furthermore, while professional societies have published national guidelines and other recommendations for lung nodule surveillance,(4-9) adherence to published recommendations is variable.(10-12) Ideally, management should maximize early diagnosis of individuals with cancerous nodules, while minimizing unnecessary testing of patients with nodules that are benign. The purpose of the Watch the Spot pragmatic trial is to compare the effects of more vs. less intensive surveillance imaging of small pulmonary nodules measuring ≤15 mm on a range of outcomes of importance to patients.

Methods

The study is an unblinded, cluster-randomized, pragmatic, non-inferiority, comparative effectiveness trial of more intensive vs. less intensive CT surveillance of patients found to have a small pulmonary nodule (Figure 1). The study employs cluster randomization at the hospital or health system level to assign participants, through the institution where they receive care, to a more intensive or less intensive surveillance strategy. Approximately 35,200 individuals with

small nodules will be enrolled over a 28-month period and followed for a minimum of 2 years to assess a broad range of stakeholder-prioritized outcomes that correspond to the study aims:

Aim 1. Among individuals with small pulmonary nodules identified either incidentally or by screening, compare the percentage of cancerous nodules that progress beyond American Joint Committee on Cancer 7th edition (AJCC 7) stage T1aN0M0 after more vs. less intensive surveillance imaging.(13) We hypothesize that less intensive surveillance will be non-inferior to more intensive surveillance, i.e. it will not result in a greater percentage of cancerous nodules diagnosed at a more advanced stage.

Aim 2a. Compare patient-reported outcomes of emotional distress, anxiety, general health status and satisfaction with the evaluation process.

Aim 2b. Compare provider-reported outcomes of knowledge, attitudes and beliefs about guidelines and practices for lung nodule evaluation, and provider satisfaction with the surveillance protocol and evaluation process.

Aim 3. Compare health care resource utilization and effective radiation doses received.

Aim 4. Compare patient and physician adherence to the recommended protocols for CT surveillance, and radiology department adherence to use of low radiation-dose techniques for screening and follow-up imaging.

The study is funded by the Patient-Centered Outcomes Research Institute (PCORI) through its program in Pragmatic Clinical Studies (PCS-1403012653) and registered on ClinicalTrials.gov (NCT02623712).

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Study Design and Rationale

As a pragmatic trial,(14) the study compares options for CT surveillance in diverse settings and in the context of usual clinical practice. The overarching goal of the pragmatic design is to integrate study procedures into existing clinical workflow to the greatest extent possible. As a comparative effectiveness trial, the study compares two alternatives in order to determine which one works best, for whom, and under what circumstances.(15) Designed and executed in close partnership with patient and non-patient stakeholders, the study outcomes are patient-centered and reflect the explicitly stated values and preferences of all stakeholders.(16)

The study will establish linkages to data from state cancer registries to ascertain the primary outcome, the percentage of cancerous nodules that progress beyond AJCC 7 stage T1aN0M0. In addition, by surveying participants via Internet or mail using clinically validated questionnaires, the study will compare patient-reported outcomes of emotional distress, anxiety, general health status and satisfaction with the evaluation process. Using data from electronic health records (EHR), the study will compare the two arms for resource utilization, effective radiation doses received and adherence to the recommended protocols for CT surveillance using low radiation-dose techniques.

Settings

Study participants are identified, enrolled and followed at 14 health care delivery organizations (Table 1), each of which agreed to accept randomized assignment to one of the two protocols for surveillance of incidentally detected pulmonary nodules. Seven of 14 organizations also agreed to randomized assignment for the surveillance of screening-detected nodules; the other

sites had recently implemented the Lung CT Screening Reporting & Data System (Lung-RADS™) and were not willing to accept randomized assignment to a more intensive surveillance protocol for nodules detected by screening. Willingness to be randomized for surveillance of screening-detected nodules was established prior to randomization. The multicenter design aimed to provide geographic, socioeconomic, and ethnic and racial diversity, and the participating health care organizations span the spectrum of U.S. health care delivery models. In addition, the settings include endemic areas for mycoses that are common causes of benign nodules, such as coccidioidomycosis and histoplasmosis.

Cluster Randomization

The study employs cluster randomization because the interventions could only be feasibly and consistently applied at the level of the health care system, i.e. all enrolled patients at any given hospital/health system will receive the same set of recommendations for follow-up.(17) Randomization at the level of individual patients was judged to be impractical for implementation, potentially confusing to providers and detrimental to patient care.

Random assignment by computer program to one of the two intervention groups was performed at the hospital level for the 11 medical centers at Kaiser Permanente Southern California (KPSC) and at the health system level for 13 other sites, by using matching (18, 19) and re-randomization.(20) Optimal matching divided 24 sites (clusters) into 12 pairs to minimize differences in the potential confounders within pairs before randomization; subsequently, one cluster from each pair was randomly assigned to the less intensive arm, the other to the more intensive arm. The balance of potential confounders was examined, and unbalanced

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

randomizations were discarded, followed by re-randomization. The process was continued until balance in measured characteristics was achieved. These characteristics included the annual volume of chest CT scanning; integrated vs. non-integrated setting; KPSC vs. other institutions; distribution of race/ethnicity; distribution of smoking; inclusion of patients with screening-detected nodules; timing of notification letters to participants; frequency of using positron emission tomography for nodule characterization; and distribution of insurance type. Ultimately, 24 sites were randomly assigned evenly to two groups; one non-enrolling site was replaced by an alternative in month 18 of the enrollment period. KPSC medical centers were treated as separate clusters because they are relatively large in size and sufficiently independent in their operations to enable use of different surveillance protocols at the medical center level. The larger number of clusters is important because statistical power depends partly on the total number of clusters.

Participants

The target population includes adults ≥35 years-old with small pulmonary nodules detected either incidentally or by screening and measuring ≤15 mm in widest diameter that are judged by the interpreting radiologist to require subsequent evaluation or surveillance for possible cancer. Interpreting radiologists were encouraged not to enroll patients with nodules judged likely to be benign, such as those with a benign pattern of calcification, intranodular fat or a location and morphology that are typical for an intrapulmonary lymph node.(21) In addition, radiologists were advised not to enroll most patients with associated pulmonary abnormalities such as pleural effusions, atelectasis, or lymphadenopathy (which increase the risk of lung

cancer), as they would need immediate and more aggressive evaluation, and not to enroll patients with multiple pulmonary nodules that are thought to be more consistent with infection or inflammation. However, enrollment was ultimately at the radiologist's discretion.

Exclusion criteria include: age <35 years; nodule identified on prior chest CT scan within 2 years; prior history of pulmonary or extrapulmonary cancer within the past 5 years (except for non-melanoma skin cancer); and pregnancy within 9 months before nodule identification.

Enrollment

Eligible patients are enrolled passively by the clinical radiologist at the time of image interpretation. Concurrently, the radiologist delivers the study intervention by inserting recommendations for evaluation in the dictated radiology report. Patients are flagged for possible inclusion and enrollment using methods tailored to fit each site. While some sites identify eligible patients by manually reviewing radiology transcripts, other sites are using automated methods, including insertion of unique text strings, hashtags or tracking assignments into dictated reports, or the use of a novel desktop application designed to facilitate enrollment and data collection. Sites were encouraged to customize these methods to be compatible with existing workflow.

Interventions

To compare the effectiveness of existing strategies for pulmonary nodule surveillance, the protocols for more intensive and less intensive surveillance were based on published guidelines (Tables 2 and 3).(5-7, 9) For patients with nodules detected incidentally, the study protocols

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

were based on a comparison of the (more intensive) original Fleischner Society recommendations(5, 7) with the (generally less intensive) revised Fleischner Society recommendations.(6) Ranges for follow-up times were simplified to maximize differences between study arms; for example, a recommendation for follow-up in 3 to 6 months was converted to 3 months in the more intensive arm and 6 months in the less intensive arm. For screening-detected nodules, the final protocols were based on a comparison of Lung-RADS recommendations (less intensive) with a more intensive set of recommendations based on the original Fleischner Society guidelines, mapped to Lung-RADS categories.(9) For example, in the more intensive arm, the recommendation for a Lung-RADS category 2 finding is to repeat the CT scan in 6 months (instead of 12 months), while the recommendation for a Lung-RADS category 3 finding is to repeat the CT scan in 3 months (instead of 6 months). Of note, while both Lung-RADS and the original Fleischner Society recommendations were considered by the study investigators and stakeholders to represent the *de facto* standards of care, they were judged to be based on low quality evidence, because there are no prior randomized trials or observational studies that compared two or more protocols for nodule surveillance. The newly revised and less intensive recommendations from the Fleischner Society were judged to be in need of evaluation, because they had not yet been implemented in most practice settings and had not been subjected to clinical experience.

Outcomes

Study outcomes were selected based on iterative rounds of feedback from both patient and non-patient stakeholders (Table 4), including a range of clinical, patient-centered and health

system outcomes. The primary study outcome (Aim 1) is tumor progression beyond AJCC 7th edition stage T1aN0M0 (tumor size ≤20 mm), the stage with the most favorable prognosis. This size threshold was identified as the best cut-point for discrimination of survival by the staging project of the International Association for the Study of Lung Cancer.(13) This corresponds to progression beyond stage T1bN0M0 in the newer AJCC 8th edition.(22) The primary outcome will be ascertained by linking study records with data from state cancer registries. Secondary cancer-related outcomes include time to cancer diagnosis and overall survival, both measured from the date of the index chest CT scan.

Patient-centered outcomes (Aim 2) will be ascertained by completion of web-based surveys approximately 1-2 months, 13 months and 25 months following the index chest CT scan. Outcomes of interest include: nodule-related emotional distress, measured with the 22-item Impact of Event Scale;(23) anxiety, measured using the 6-item State-Trait Anxiety Inventory;(24) and a single-item question about general health status. Patient surveys also include questions about patient satisfaction with the process of lung nodule surveillance, provider communication, preferred style of decision-making and barriers to adherence with follow-up.

Participating radiologists and ordering providers will complete novel surveys to assess knowledge, attitudes and beliefs about existing guidelines for pulmonary nodule evaluation (at baseline) and the assigned protocols for surveillance in use at their site (near the end of enrollment).

Nodule-related resource utilization (Aim 3) will be ascertained by searching structured data in the EHR for relevant Current Procedural Terminology (CPT) and International

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Classification of Disease, version 10 (ICD-10) procedure codes that appear during the surveillance period (from date of the index CT scan to the date of cancer diagnosis or 2 years of follow-up, whichever comes first). We will capture all nodule-related imaging tests (chest CT, positron emission tomography, bone scans, brain CT or magnetic resonance imaging, abdominal and pelvic CT), invasive biopsy procedures (bronchoscopy, transthoracic needle biopsy), thoracic surgical procedures, emergency department visits and hospitalizations. We will also record procedure-related complications by searching for diagnostic codes for contrast-induced nephropathy, pneumothorax, respiratory failure and major bleeding.

Outcomes for Aim 3 also include radiation exposure, as reflected by the computed tomography dose index (CTDIvol), the dose-length product (DLP), and the effective radiation dose. The CTDIvol equals the average dose emitted by the scanner within each small area imaged (often called a slice), while the DLP represents the total imparted radiation and is defined as the CTDIvol multiplied by the scan length. Effective dose is a calculated value and is a function of the DLP, the specific organs irradiated, and the sensitivity of the organs irradiated to develop cancer in the future. Effective dose will be calculated using DLP and established conversion formulas.(25)

Aim 4 will compare adherence to the assigned surveillance protocol at the level of the interpreting radiologist, the ordering provider, and the individual patient. The primary measure of adherence will be adherence with the first recommended surveillance test: was the test recommended by the radiologist, ordered by the provider, and completed by the patient? Secondary analyses will examine more granular information about adherence on a test-by-test basis and pinpoint the level of non-adherence.

Data Collection and Management

Data elements will be collected and managed locally by investigators at each site, and subsequently transferred securely to the Data Coordinating Center (DCC) at UC Davis for quality control and analysis. EHRs will be searched for information on baseline patient characteristics (e.g. demographics, smoking status, comorbid conditions) and health care utilization. Radiology reports will be searched manually and/or by using validated natural language processing algorithms to ascertain nodule size, attenuation (solid, part-solid or non-solid), location, calcification and edge characteristics. Survey data will be collected locally or centrally by the DCC, depending on the site. Patient surveys will not be distributed at one of the sites (Vanderbilt University). To ascertain the primary outcome (cancer diagnosis and stage), linkages will be made with data from state cancer registries either centrally by the DCC, or locally at selected health care systems. All sites are required to conduct monthly quality assurance by manually reviewing random samples of dictated radiology reports to ensure appropriate enrollment of eligible patients. The study is overseen by an independent Data Safety and Monitoring Board.

Statistical Power

The study was designed to demonstrate the non-inferiority of the less intensive surveillance protocol relative to the more intensive protocol. With a sample size of 960 individuals with cancerous nodules, the study will have 90% power to demonstrate non-inferiority with a margin of 5% for the primary outcome of cancer progression beyond stage T1a, using a one-sided Z test with a significance level of 0.05 and an intraclass correlation coefficient (ICC) of

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

0.012. The non-inferiority margin was selected by members of the research team in collaboration with clinical and patient stakeholders to be the narrowest possible margin that allowed for a feasible sample size. Assuming 3% of enrolled patients with nodules measuring ≤15 mm will have cancer, and allowing for a 10% loss to follow-up, the trial will require enrollment of 35,200 participants to meet the target sample size. Alternatively, if the ICC is ≤0.01, the study will still have 90% power to demonstrate non-inferiority with 888 cancerous nodules (or 32,560 participants enrolled).

Statistical Analysis

The primary analysis will evaluate whether less intensive surveillance *is non-inferior* to more intensive surveillance by examining the upper bound of the 95% confidence interval (CI) for the difference in percentage for the less-intensive surveillance vs. more-intensive surveillance. The null hypothesis is that less intensive surveillance is inferior to more intensive surveillance, i.e., the less intensive arm will result in 5 percentage-points or more of tumors progressing beyond stage T1aN0M0 than the more intensive arm. We will reject this null-hypothesis and conclude that less intensive surveillance is non-inferior to more intensive surveillance if the upper bound of the 95% CI for the difference of the percentages of patients with tumor progression beyond T1aN0M0 if the less vs. more intensive arm does not exceed the non-inferiority margin of 5%. We will model the primary outcome using hierarchical logistic regression, including random site-specific effects to account for clustering of patients within sites. Hierarchical logistic regression models will be fitted without (primary analysis) and with (secondary analysis) adjusting for potential confounders including age, gender, ethnicity/race, smoking, body mass

index, baseline nodule size, indication for CT (screening or diagnostic), and all facility-level factors balanced during the randomization. The difference in adjusted percentages of patients with tumor progression beyond T1aN0M0 will be estimated using predictive margins, averaging over the predicted values for each site and standardizing to the overall study population for models adjusting for potential confounders.(26, 27)

Primary analyses will be by intention to treat (ITT), including all patients with qualifying nodules, including those who do not undergo surveillance (i.e., non-adherent cases or patients that proceed directly to tissue diagnosis). We will perform a per protocol (PP) sensitivity analysis to evaluate outcomes by surveillance strategy received.

Pre-specified subgroup analyses will include interaction terms to evaluate whether outcomes vary by indication (lung cancer screening vs. other), smoking history, nodule density (solid, part-solid, or non-solid), health care setting (integrated vs. other), demographic characteristics (age, sex, race/ethnicity), and geographic region (endemic for mycosis vs. non-endemic). For interactions significant at the 0.20 level, we will explore the treatment effects in corresponding subgroups. We will use multiple imputation to account for missing data.

Human Subjects

The study protocol was approved by the Institutional Review Board (IRB) at each participating site. In all cases, the IRB granted a waiver of informed consent because the study is testing a system-level intervention (insertion of guideline-based recommendations for surveillance), and because the risks of participating in this comparative effectiveness study were judged to be no different than the risks commonly encountered in usual clinical practice.(28, 29) In addition, the

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

study would not be feasible or logistically possible without the waiver, because the intervention (insertion of recommendations) is delivered by the interpreting radiologist during usual clinical workflow at the time of interpretation, which typically occurs long after the patient has been discharged from the radiology department. Of note, patients, radiologists and ordering providers are permitted to deviate from the recommendations when dictated by patient preference or clinical judgment. Although the requirement for informed consent was waived, most sites decided in collaboration with their IRB to contact enrolled participants by letter or electronic mail to notify them about the study and provide an opportunity to opt-out for data collection purposes. Participants who completed surveys provided consent electronically online or by phone for this portion of the study.

Study Team and Governance

The study team includes researchers, clinicians, patients and additional stakeholders from professional societies and advocacy groups (Figure 2). All collaborative activities are guided by the PCORI engagement principles of reciprocity, co-learning, partnership, trust, transparency and honesty.(30) Both patient and non-patient stakeholders have actively participated in the design and execution of the trial and have vetted and endorsed all major decisions, including the design of the surveillance protocols, the selection of outcomes and the methods used to passively enroll and subsequently notify study participants.

Discussion

Watch the Spot is a large, unblinded, pragmatic, cluster-randomized, non-inferiority, comparative effectiveness trial that addresses an important gap in what is known about the evaluation and management of patients with small pulmonary nodules. Current guidelines for patients with nodules detected either incidentally or by screening are not based on evidence from randomized controlled trials or well-designed observational studies of comparative effectiveness. Despite this, hundreds of thousands of individuals each year undergo lung nodule follow-up that may represent either too much or too little care.(3)

By comparing existing guidelines for pulmonary nodule surveillance, the results of Watch the Spot will set the bar for the frequency and duration of nodule follow-up. If less intensive surveillance is shown to be non-inferior to more intensive care, the study will provide high quality evidence in support of using the revised Fleischner Society guidelines and the current Lung-RADS recommendations. If non-inferiority is not demonstrated, the trial will send a strong signal that the original, more intensive, Fleischner Society recommendations should be reinstated (and that Lung-RADS recommendations should be intensified). Similarly, if patient satisfaction and adherence are found to be suboptimal, this might prompt efforts to modify existing guidelines and address any barriers to adherence that we identify.

One important limitation of this comparative effectiveness trial is that the interventions to be compared were necessarily limited to existing guidelines, and we therefore were not able to include a simpler protocol for nodule surveillance. Given the pragmatic design and our focus on comparative effectiveness, it was paramount to compare strategies used in current clinical

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

practice, including one strategy thought to represent the *de facto* standard of care and another one based on newly revised yet untested recommendations from a respected professional society. In addition to ensuring equipoise between the study arms, the protocols were designed to be acceptable to practitioners and relevant to clinical and policy-level decision-making. At the same time, the two protocols were implemented in a way that made them as distinct as possible to enable us to find true differences in outcomes, if they exist.

Another limitation is that the planned ITT analysis will be biased to the null (non-inferiority) if there is poor adherence with the surveillance recommendations.(31) However, analysis by ITT is preferred because the goal of the trial is to compare the real-world effectiveness of strategies for surveillance of small pulmonary nodules, rather than efficacy under the more idealized assumptions of the PP analysis.(32) ITT preserves the benefits of cluster randomization, maintains sample size, prevents bias in analyses resulting from post-randomization exclusion, and has been widely used in non-inferiority trials.(33) In addition, results can be biased in either direction for both ITT and PP analyses.(34) In one recent review article, the authors found that the method of analysis seldom affected the results, and the ITT analysis was actually more conservative in four out of five trials.(35) Thus, we favor ITT as the primary analysis to compare the real-world effectiveness of two surveillance strategies in this pragmatic, cluster-randomized, non-inferiority trial. In contrast, the PP sensitivity analysis will address the policy-relevant question of efficacy under the assumption of perfect adherence.

A final limitation is uncertainty about the magnitude of the intraclass correlation coefficient and the prevalence of malignant nodules that could result in reduced statistical power.

Watch the Spot has several novel and distinctive features of interest to clinicians, clinical trialists and funders of research. Foremost, it is one of the first large, pragmatic clinical trials to be funded by PCORI. The overarching goal of the pragmatic design was to integrate study procedures into usual clinical care to the greatest extent possible, to maximize both the efficient use of resources and the generalizability of our findings. Second, the use of cluster-randomization and the evaluation of an intervention applied at the system-level are relatively uncommon in comparative effectiveness research, although countless other diagnostic and therapeutic protocols are potentially amenable to system-level implementation and evaluation. Third, the study protocol enables sites to customize methods for identifying and (passively) enrolling participants. Most sites employ largely automated approaches, illustrating the potential efficiency gains of technology-enabled research. Assuming the study reaches its enrollment target of approximately 35,200 participants, the cost per patient enrolled will be only $250, a small fraction of the per patient cost of a conventional randomized clinical trial. Lastly, the design and execution of the study are the product of intensive engagement with patient and non-patient stakeholders, ensuring that the study reflects the values and preferences of all concerned stakeholders, and is responsive to the information needs of patients with pulmonary nodules and the clinicians who care for them.

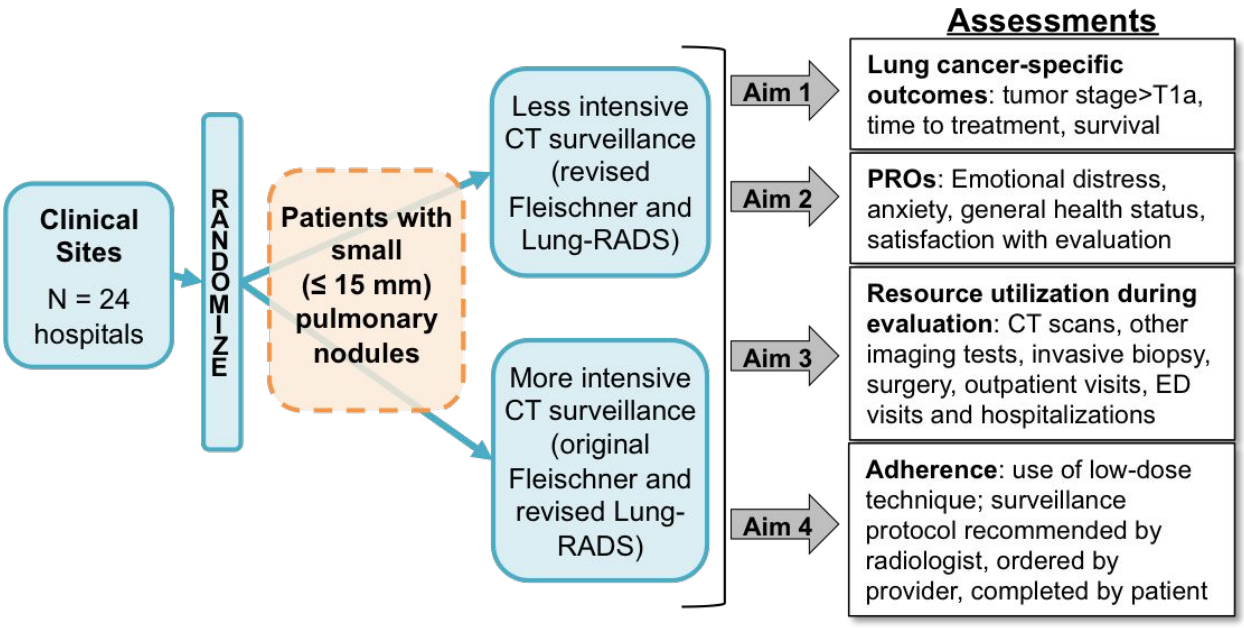
ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Figure Legends

Figure 1: Schematic representation of study design. Fleischner= Fleischner Society recommendations for pulmonary nodule evaluation. Lung-RADS= Lung Imaging Data and Reporting System.

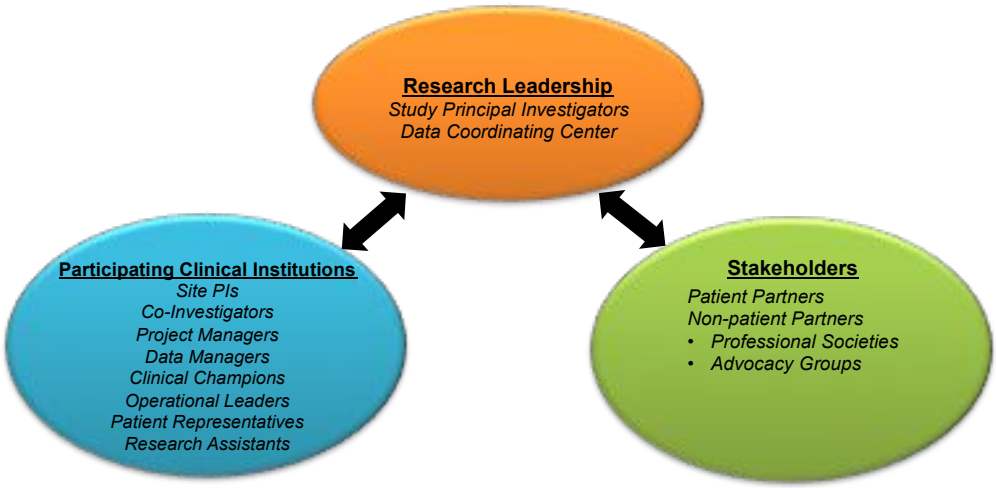
Figure 2: Governing structure. The study is led by the Principal Investigators, in collaboration with the Stakeholder Advisory Group and the Steering Committee. All decisions are made by the Executive Committee, after formal vetting and approval by the Steering Committee and Stakeholder Advisory Group. Additional Work Groups are charged with project management, data management and survey development. Local study teams at each site identify and enroll participants and have primary responsibility for secure data collection, storage and transfer to the Data Coordinating Center.

Figure 1



ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Figure 2



Committees	Composition	Meeting Frequency
Executive Committee (EC)	Research leadership (study PIs), Data Coordinating Center representatives, 1 site PM	Weekly
Stakeholder Advisory Group (SAG)	Research leadership (study PIs), all patient and non-patient stakeholder partners	Quarterly
Steering Committee (SC)	Research leadership (study PIs), all site PIs, 1 patient partner, 2 non-patient partners	Monthly
Project Manager Workgroup	All site PMs, site research support staff	Monthly
Local Study Teams (LST)	Site PI, Clinical champions, operational or administrative leader, local patient representative	Varies depending on site
Provider and Patient Survey Workgroup	Research leadership (study PIs), interested site PIs, site Co-Investigators, site PMs, and interested stakeholders	Bi-weekly, then as needed

Table 1: Description of Study Sites

Health Care Organization	Geographic Location	Presence of Endemic Mycosis	Will Enroll Patients with Screen-Detected Nodules	Type of System	Group Assign-ment (More vs. Less Intensive)
Boston Medical Center	Northeast	No	No	Safety Net	More
Cleveland Clinic	Northeast	Yes	No	Referral	Less
Health Partners, MN	Midwest	Yes	Yes	Integrated	More
Kaiser Permanente Colorado	Mountain West	No	Yes	Integrated	Less
Kaiser Permanente Northwest	Northwest	No	Yes	Integrated	Less
Kaiser Permanente Southern California	Southwest	Yes	Yes	Integrated	Both*
Medical University of South Carolina	Southeast	Yes	Yes	University	More
National Jewish Health	Mountain West	No	No	Referral	More
Portland Veterans Affairs Med Center	Northwest	No	Yes	Integrated	Less
University of California Davis	West	Yes	Yes	University	Less
University of California Los Angeles	Southwest	Yes	No	University	More
University of California San Francisco	West	No	No	University	Less
University of Pennsylvania	Northeast	No	No	University	More
Vanderbilt University	Southeast	Yes	No	University	More

*Kaiser Permanente Southern California hospitals assigned to more intensive surveillance include facilities in Downey, Fontana, Panorama City, Riverside and San Diego. Hospitals assigned to the less intensive group include those located in Baldwin Park, Los Angeles, Orange County, South Bay, West Los Angeles and Woodland Hills.

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Table 2a: Study protocol for surveillance or evaluation of solid nodules, Group A

Size (mm)	Incidental Nodule in Patient without Risk Factors (follow-up in months)	Incidental Nodule in Patient with Risk Factors (follow-up in months)	Screening-Detected Nodule (follow-up in months)
≤4	Optional at 12	12	12, 24…
>4 to ≤6	12	6, 18	6, 18, (30)
>6 to ≤8	6, 18	3, 9, 21-24	3, 15, 27…
>8	PET, biopsy or CT at 3, 9, 21-24 months		

Recommendations for incidentally detected solid nodules based on Fleischner Society guidelines (2005). Recommendations for solid nodules detected by screening adapted from the Lung CT Screening Reporting and Data System (Lung-RADS). Numbers in parentheses reflect follow-up that may occur after the study is over. Ellipsis indicates that annual screening should continue until patient no longer meets eligibility criteria.

Table 2b: Study protocol for surveillance or evaluation of sub-solid nodules, Group A

Attenuation	Size (mm)	Size of Solid Component (mm)	Patient with or without Risk Factors (follow-up in months)
Non-Solid	≤5		Incidental and solitary: None Incidental and multiple: 24, (48) Screening-detected: 12, 24…
	>5		3, 15, 27, (39)
Part-Solid	Any	<5	3, 15, 27, (39)
	Any	≥5	Repeat CT at 3 months; if persistent, biopsy or resect

Recommendations for incidentally detected sub-solid nodules based on Fleischner Society guidelines (2013). Recommendations for sub-solid nodules detected by screening adapted from the Lung CT Screening Reporting and Data System (Lung-RADS). Numbers in parentheses reflect follow-up that may occur after the study is over. Ellipsis indicates that annual screening should continue until patient no longer meets eligibility criteria.

Table 3a: Study protocol for surveillance or evaluation of solid nodules, Group B

Size (mm)	Incidental Nodule in Patient without Risk Factors (follow-up in months)	Incidental Nodule in Patient with Risk Factors (follow-up in months)	Screening-Detected Nodule (follow-up in months)
<6	None	Optional at 12	12, 24…
≥6 to ≤8	Solitary: 12, 24		12, 24…
	Multiple: 6, 18		6, 18, (30)…
>8	PET, biopsy or CT at 3, 15, (27) months		

Recommendations for incidentally detected solid nodules based on Fleischner Society guidelines (2017). Recommendations for solid nodules detected by screening based on the Lung CT Screening Reporting and Data System (Lung-RADS). Numbers in parentheses reflect follow-up that may occur after the study is over. Ellipsis indicates that annual screening should continue until patient no longer meets eligibility criteria.

Table 3b: Study protocol for surveillance or evaluation of sub-solid nodules, Group B

Attenuation	Size (mm)	Incidental Nodule in Patient with or without Risk Factors (follow-up in months)	Screening-Detected Nodule (follow-up in months)
Non-Solid	<6	Solitary: None Multiple: 6, 24, (48)	Solitary: 12, 24… Multiple: 6, 18, (30)…
	≥6	Solitary: 12, (36), (52) Multiple: 6, 24, (48)	
Part-Solid	<6	Solitary: None Multiple: 6, 24, (48)	
	≥6	6, 18, (30), (42), (54), (66); biopsy if solid component ≥6	6, 18, (30)…

Recommendations for incidentally detected sub-solid nodules based on Fleischner Society guidelines (2017). Recommendations for sub-solid nodules detected by screening based on the Lung CT Screening Reporting and Data System (Lung-RADS). Numbers in parentheses reflect follow-up that may occur after the study is over. Ellipsis indicates that annual screening should continue until patient no longer meets eligibility criteria.

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Table 4: Definitions and source information for outcomes, by specific aim

Aim	Sample	Outcome	Definition	Source
1	Participants with cancerous nodules	AJCC 7 Stage >T1aN0M0	Tumor size >20 mm at time of resection or radiotherapy, with no distant metastasis or regional lymph node involvement.	Cancer Registry, EHR
		Time to treatment	Time to surgery, radiotherapy or chemotherapy, measured from date of index CT scan to date of first treatment.	
		Survival	Measured from date of index CT scan to death or censoring.	
2	All patients with nodules and access to email	Nodule-related distress	Measured with validated Impact of Event Scale (IES-R). Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	Self-administered web survey
		Anxiety	Measured with validated State-Trait Anxiety Inventory (STAI-6). Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		General health status	Measured with 1 item from the validated Short Form Health Survey. Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Smoking history	Measured with items selected from the Cancer Care Outcomes Research and Surveillance Study patient survey. Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Health literacy	Measured with the validated Single Item Literacy Screener. Assessment performed 1-2 months after index CT scan.	
		Perceived	Measured with items adapted	

		susceptibility to cancer	from the validated Champion Health Belief Model Tool. Assessment performed 1-2 months after index CT scan.	
		Cancer worry	Measured with an item adapted from the validated Lerman Cancer Worry Scale. Assessment performed 1-2 months after index CT scan. Measured with novel items at 13 months, and at end of follow-up.	
		Patient preferences about control over decision making	Measured with an adapted version of the validated Control Preferences scale. Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Motivation to quit smoking	Measured with items adapted from Sciamanna et al.(36) Assessments performed with self-reported smokers at 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Perceived risks and benefits of lung nodule surveillance	Measured with items adapted from the validated Decisional Conflict Scale.(37) Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Concrete barriers to lung nodule surveillance	Measured with novel items, Likert-type scale. Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Provider communication about lung nodule surveillance	Measured with novel items, Likert-type scale. Assessments performed 1-2 months after index CT scan, at 13 months, and at end of follow-up.	
		Satisfaction with evaluation	Measured with novel items, Likert-type scale. Assessment performed at 13 months and at	

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

			the end of follow-up.	
	All participating radiologists, ordering providers (pulmonologists, thoracic surgeons, and PCPs)	Knowledge, attitudes, and beliefs about guidelines and practices for lung nodule evaluation; satisfaction with surveillance protocol and notification systems; organizational factors affecting adherence	Measured with novel items based on the Consolidated Framework for Implementation Research, Likert-type scale. Assessments performed within 1 – 2 months of trial launch and at 18 months after trial launch.	Self-administered web or paper-based survey
3	All patients with nodules	Nodule-related resource utilization and total radiation exposure	Includes all CT scans; PET scans; other imaging tests; invasive biopsy procedures (bronchoscopic and percutaneous); thoracic surgical procedures; all outpatient visits, ED visits and hospitalizations during the surveillance period.	EHR
4	All patients with nodules, random 10% sample for greater detail	Adherence with assigned surveillance protocol	EHR reviewed to determine whether surveillance imaging was completed per protocol; detailed review of radiology transcripts and orders to determine whether assigned protocol was recommended by radiologist and ordered by provider.	EHR, radiology transcripts

References

1. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. N Engl J Med. 2003;348(25):2535-42.
2. Rubin GD. Lung nodule and cancer detection in computed tomography screening. J Thorac Imaging. 2015;30(2):130-8.
3. Gould MK, Tang T, Liu IL, Lee J, Zheng C, Danforth KN, et al. Recent Trends in the Identification of Incidental Pulmonary Nodules. American journal of respiratory and critical care medicine. 2015;192(10):1208-14.
4. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it 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):e93S-120S.
5. MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology. 2005;237(2):395-400.
6. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology. 2017;284(1):228-43.
7. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. Radiology. 2013;266(1):304-17.
8. Network NCC. NCCN Lung Cancer Screening Guidlines Version 1.2014. National Comprehensive Cancer Network, Inc2014.
9. Radiology ACo. Lung-RADS Version 1.0 Assessment Categories 2019 Available from: www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads. Accessed June 26, 2019.
10. Eisenberg RL, Bankier AA, Boiselle PM. Compliance with Fleischner Society guidelines for management of small lung nodules: a survey of 834 radiologists. Radiology. 2010;255(1):218-24.
11. Moseson EM, Wiener RS, Golden SE, Au DH, Gorman JD, Laing AD, et al. Patient and Clinician Characteristics Associated with Adherence. A Cohort Study of Veterans with Incidental Pulmonary Nodules. Annals of the American Thoracic Society. 2016;13(5):651-9.
12. Wiener RS, Gould MK, Slatore CG, Fincke BG, Schwartz LM, Woloshin S. Resource use and guideline concordance in evaluation of pulmonary nodules for cancer: too much and too little care. JAMA internal medicine. 2014;174(6):871-80.
13. Rami-Porta R, Crowley JJ, Goldstraw P. The revised TNM staging system for lung cancer. Ann Thorac Cardiovasc Surg. 2009;15(1):4-9.

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

14. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009;62(5):464-75.

15. Carson SS, Goss CH, Patel SR, Anzueto A, Au DH, Elborn S, et al. An official American Thoracic Society research statement: comparative effectiveness research in pulmonary, critical care, and sleep medicine. American journal of respiratory and critical care medicine. 2013;188(10):1253-61.

16. Feemster LC, Saft HL, Bartlett SJ, Parthasarathy S, Barnes T, Calverley P, et al. Patient-centered Outcomes Research in Pulmonary, Critical Care, and Sleep Medicine. An Official American Thoracic Society Workshop Report. Annals of the American Thoracic Society. 2018;15(9):1005-15.

17. Cook AJ, Delong E, Murray DM, Vollmer WM, Heagerty PJ. Statistical lessons learned for designing cluster randomized pragmatic clinical trials from the NIH Health Care Systems Collaboratory Biostatistics and Design Core. Clin Trials. 2016;13(5):504-12.

18. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. Biostatistics. 2004;5(2):263-75.

19. Greevy RA, Jr., Grijalva CG, Roumie CL, Beck C, Hung AM, Murff HJ, et al. Reweighted Mahalanobis distance matching for cluster-randomized trials with missing data. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 2:148-54.

20. Morgan KL, Rubin DB. Rerandomization to improve covariate balance in experiements. Annals of Statistics. 2012;40(2):1263-82.

21. Shaham D, Vazquez M, Bogot NR, Henschke CI, Yankelevitz DF. CT features of intrapulmonary lymph nodes confirmed by cytology. Clin Imaging. 2010;34(3):185-90.

22. Rami-Porta R, Bolejack V, Crowley J, Ball D, Kim J, Lyons G, et al. The IASLC Lung Cancer Staging Project: Proposals for the Revisions of the T Descriptors in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2015;10(7):990-1003.

23. Salsman JM, Schalet BD, Andrykowski MA, Cella D. The impact of events scale: a comparison of frequency versus severity approaches to measuring cancer-specific distress. Psychooncology. 2015;24(12):1738-45.

24. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol. 1992;31 (Pt 3):301-6.

25. Christner JA, Kofler JM, McCollough CH. Estimating effective dose for CT using dose-length product compared with using organ doses: consequences of adopting International Commission on Radiological Protection publication 103 or dual-energy scanning. AJR American journal of roentgenology. 2010;194(4):881-9.

26. Chang IM, Gelman R, Pagano M. Corrected group prognostic curves and summary statistics. J Chronic Dis. 1982;35(8):669-74.

27. Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. Biometrics. 1982;38(3):613-21.

28. Faden RR, Beauchamp TL, Kass NE. Informed consent, comparative effectiveness, and learning health care. N Engl J Med. 2014;370(8):766-8.

29. Platt R, Kass NE, McGraw D. Ethics, regulation, and comparative effectiveness research: time for a change. JAMA : the journal of the American Medical Association. 2014;311(15):1497-8.

30. Sheridan S, Schrandt S, Forsythe L, Hilliard TS, Paez KA, Advisory Panel on Patient E. The PCORI Engagement Rubric: Promising Practices for Partnering in Research. Ann Fam Med. 2017;15(2):165-70.

31. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ. 1996;313(7048):36-9.

32. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. Perspect Clin Res. 2016;7(3):144-6.

33. Wiens BL, Zhao W. The role of intention to treat in analysis of noninferiority studies. Clin Trials. 2007;4(3):286-91.

34. Matsuyama Y. A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. Stat Med. 2010;29(20):2107-16.

35. Aberegg SK, Hersh AM, Samore MH. Empirical Consequences of Current Recommendations for the Design and Interpretation of Noninferiority Trials. Journal of general internal medicine. 2018;33(1):88-96.

36. Sciamanna CN, Hoch JS, Duke GC, Fogle MN, Ford DE. Comparison of five measures of motivation to quit smoking among a sample of hospitalized smokers. Journal of general internal medicine. 2000;15(1):16-23.

37. O'Connor AM. Validation of a decisional conflict scale. Medical decision making : an international journal of the Society for Medical Decision Making. 1995;15(1):25-30.

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Online Data Supplement

Methods for the Watch the Spot Trial: A Pragmatic Trial of More vs. Less Intensive Strategies for Active Surveillance of Small Pulmonary Nodules

Michael K. Gould, MD, MS, Rebecca Smith-Bindman, MD, Karen Kelly, MD, Danielle E. Altman, MA, Igor Barjaktarevic, MD, PhD, Beth Creekmur, MA, Evan de Bie, BS, Debra S. Dyer, MD, Eduardo J. Mortani Barbosa Jr., MD , Richard A. Mularski, MD, Lihong Qi, PhD[5], Laszlo T. Vaszar, MD, Sophronia Yu, MPH, Diana L. Miglioretti, PhD, on behalf of the Watch the Spot Trial Investigators

Appendix A: Watch the Spot Settings and Investigators

Kaiser Permanente Southern California: Michael Gould (PI), Brian Mittman (co-I), Danielle Altman, Beth Creekmur, Brian Huang, Chengyi Zheng, Visanee Musigdilok, Emily Rozema

Boston Medical Center: Renda Wiener (co-I), Anuradha Rebello, Hasmeena Kathuria, Karen Lasser, Linda Rosen, Vruti Virani

Cleveland Clinic: Peter Mazzone (co-I), Amy Pritchard, Ruffin Graham, Sudish Murthy, Joseph Azok, Christopher Estling

HealthPartners: Charlene McEvoy (co-I), Linda Loes, Mary T. Becker, Angela Tai

Kaiser Permanente Colorado: Debra P. Ritzwoller (co-I), Christina Clarke, Julie Steiner, Ruth Bedoy, Courtney Kraus, Caroline Joyce

Kaiser Permanente Northwest: Eric Walter (co-I), Anne Ramey, Catherine Cleveland, Jennifer Cook, Britta Torgrimson-Ojero, and Deralyn Almaguer

Medical University of South California: Gerard Silvestri (co-I), James Ravenel, Kate Taylor, Katie Kirchoff, Nichole Tanner

National Jewish Health: Debra Dyer (co-I), Elizabeth Kern, Pearlanne Zelarney

University of Pennsylvania: Anil Vachani (co-I), Eduardo Barbosa, Jennifer Steltz

University of California, Davis, School of Medicine, Department of Public Health Sciences: Diana Miglioretti (co-I), Evan de Bie, Lihong Qi, Yang Vang

University of California, Davis, School of Medicine, Department of Medicine: Karen Kelly (co-I), Friedrich Knollmann, Diem Le, Shantha Rao

University of California, Los Angeles: Denise Aberle (co-I), Chang Su, Igor Barjaktarevic

University of California, San Francisco: Rebecca Smith-Bindman (co-I), Sophronia Yu

Veterans Affairs Portland Health Care System: Christopher Slatore (co-I), Sara Golden, Danielle Apodaca, Sarah Shull, Matthew Howard

Vanderbilt University Medical Center: Kim Sandler (co-I), Emily Epstein, Karthik Ramadass

ANNALSATS Articles in Press. Published on 17-July-2019 as 10.1513/AnnalsATS.201903-268SD

Appendix B: Watch the Spot Stakeholders

Patient stakeholders

Jamie Daniel
Kathleen Fennig
Charles Florsheim
Kaitlyn Pedotti

Non-patient stakeholders and their affiliations

Jill Arnstein, American Lung Association in California
Frank Detterbeck, American College of Chest Physicians
Ella A. Kazerooni, American College of Radiology & National Lung Cancer Roundtable
Amy Moore, Bonnie J. Addario Lung Cancer Foundation
Richard Mularski, The COPD Patient Powered Research Network of the COPD Foundation
Nir Peled, International Association for the Study of Lung Cancer, Soroka Cancer Center and Ben-Gurion University, Beer Sheva, Israel
Charles Powell, American Thoracic Society
Robert Smith, American Cancer Society
Laszlo T. Vaszar, Mayo Clinic Arizona